HI-C Contacts Encode Heterogeneity in Sub-diffusive Motion of E. coli Chromosomal Loci

Palash Bera, Abdul Wasim, and Jagannath Mondal
Tata Institute of Fundamental Research, Hyderabad 500046, India

Underneath its apparently simple architecture, the circular chromosome of E. coli is known for displaying complex dynamics in its cytoplasm. Recent experiments have hinted at an inherently heterogeneous dynamics of chromosomal loci, the origin of which has largely been elusive. In this regard, here we investigate the loci dynamics of E. coli chromosome in a minimally growing condition at 30°C by integrating the experimentally derived Hi-C interaction matrix within a computer model.

Our quantitative analysis demonstrates that, while the dynamics of the chromosome is sub-diffusive in a viscoelastic media in general, the diffusion constants and the diffusive exponents are strongly dependent on the spatial coordinates of chromosomal loci. In particular, the loci in Ter Macro-domain display slower mobility compared to the others. The result is found to be robust even in the presence of active noise. Interestingly, a series of control investigations reveal that the absence of Hi-C interactions in the model would have abolished the heterogeneity in loci diffusion, indicating that the observed coordinate-dependent chromosome dynamics is heavily dictated via Hi-C-guided long-range inter-loci communications. Overall, the study underscores the key role of Hi-C interactions in guiding the inter-loci encounter and in modulating the underlying heterogeneity of the loci diffusion.

I. INTRODUCTION

The chromosome of the prototypical bacteria E. coli consists of 1.6 μm super-coiled circular DNA of 4.64 Mb. Unlike eukaryotes, it has no nucleus and it is confined within a (2 – 4)μm long spherocylinder [1, 2]. However, the perception that bacterial chromosome is a randomly packed DNA, is fast changing. Precedent investigations have provided numerous evidences that the ‘nucleoid’ is just not a complex blob of genomic DNA, RNA and associated proteins. Rather it is being replaced by a picture of self-organized architecture with distinctly segregated macro-domains (MDs) and non structured (NS) regions [3, 4]. In addition to traditional molecular biology based experiments [3-9], the emergence of Chromosome conformation capture techniques in bacteria [10–12] are providing a microscopic view of the genome-level organization of its chromosome. In particular, a high resolution (5kb), Hi-C contact map for E. coli has recently been reported by Lioy et al [13]. The subsequent integration of this Hi-C interaction map in a polymer-based computational model [14], had bestowed a clear chromosomal compartmentalization, a fundamental building block of E. coli chromosomal architecture. In light of this, the center of our current investigation is: how does the self-organized nature of the E. coli chromosome impact the dynamics of its individual loci?

The dynamics of chromosomal loci is generally quantified by mean square displacements (MSD). Generally, for a diffusive particle $MSD \sim t^\alpha$, where $t$ is lag time and $\alpha$ is called MSD exponent. $\alpha = 1$ indicates normal diffusion [15, 16], $\alpha > 1$ indicates super-diffusion [17–20] and $\alpha < 1$ implies sub-diffusion. Previous interesting experimental and theoretical investigations by Weber et al. [21–24] and Javer et al. [25] mainly focused on the spatiotemporal dynamics of E. coli chromosomal loci. Weber et al. explored the dynamics of chromosomal loci between (100) s time interval in LB medium (wt37LB). Their investigations showed that the movement of individual loci is sub-diffusive with a universal diffusive exponent $\alpha = 0.39$ and this value was reported by them to be robust for both E. coli and Caulobacter crescentus with different drug treatments. A key conclusion of the investigations by Weber et al [21, 22] was that this universal value can not be explained by typical Rouse-like polymer model [26, 27]. Their measurement of velocity auto-correlation function with a negative peak indicated a viscoelastic nature of cytoplasm. A subsequent study by the same group had shown that ATP dependent biological activities do not change these robustness of MSD exponent but mobility of loci changes significantly [24]. However, in a later crucial investigation, Javer et al [25] had investigated the dynamics of a large set of chromosomal loci across the genome within very short time-scale (=(0.1 – 10) second) in minimal medium, using high resolution tracking method. While this investigation also found that the MSD exponents of all loci indicate a sub-diffusive motion even in short time, there is a large variation of mobility across the chromosome loci. More importantly, the time scale and diffusivity of Ter macro domain were found to be very low compared to other macro domains. Together, these investigations bring out the complexity in the dynamics of the E. coli chromosome and allude to a heterogeneous loci dependence. However, a microscopic picture of the origin of such observations are yet to emerge, especially at individual chromosome loci level.

Based on these precedent investigations and particularly the observation of the heterogeneous diffusivity across the loci, we surmised that the inter-loci cross-talk and specific identity of loci pair might play an important role. Towards this end, we planned on employing our recent model of E. coli chromosome [14], which had integrated the high-resolution (5 kb) Hi-C interaction map [13] in an excluded-volume polymer-based framework, for investigating the dynamics of the chromosomal loci. Specifically we use this Hi-C-encoded model to explore the mobility of each of 30 loci (same set of loci as investigated by Javer et al’s tracking experiments) via Brownian Dynamics (BD) simulations. Our analysis of the simulations trajectories reveal that both the diffusion constant and the exponents are loci-coordinate dependent and indicate a heterogeneous distribution. The result is found to be robust in presence of active noise. Interestingly, our control simulations via turning off the Hi-C interactions in the model abolishes this heterogeneity, suggesting that the loci-specific inter-genome interactions, as encoded in the Hi-C map, hold the key to the complex dynamics of the bacterial chromosome.
II. RESULTS AND DISCUSSION

A. An integrative Computer model Recapitulates E. coli’s Hi-C Contacts

The chromosome of E. coli grown in minimal media at 30°C (wt30MM) forms the basis of our investigation of the dynamics of its loci. Majority of the relevant experiments involving dynamics has also been performed on this growth condition. The cytoplasm of E. coli at this particular growth condition, contains a single chromosome, is not complicated by the fast replication process and hence is devoid of replication forks. As a result, the chromosome serves as a prototypical system which can be investigated for its short-time dynamics (20 min in real physical unit), without worrying about the chromosome segregation. Towards this end, we implemented a computer model of the E. coli chromosome by integrating beads-spring polymer topology with recently reported Hi-C interactions matrix of E. coli [13] chromosome at this particular condition (wt30MM). For this purpose, we employ a recently proposed protocol by our group [14]. As detailed in the Simulation Model and Methods section, the excluded volume interaction, Hi-C contacts and a spherocylindrical confinement form the key interactions of the model. We generated an ensemble of dynamical trajectories of the time-evolution of the chromosome model via BD simulations under over-damped condition (see Model and Methods). Figure 1 (a) renders a representative snapshot of the chromosome with color-coded encircled beads referring to various loci for different macro domains and various non-structured regions, as obtained in the simulations. The segregation of chromosome into six key macro domains, a signature feature of E. coli, is clearly evident. To assess the precision of the simulated model, we first compare the ensemble-averaged inter-bead contact probability with the experimental Hi-C matrix. Figure 1 (b) and 1(c) depict the heat maps of experimental and simulated contact probability matrices. The occurrence of an intense diagonal in both experimentally derived and simulated matrices represents that there is a higher contact probability between the chromosomal region for lower distances. Figure 1 (d) and 1 (e) demonstrate the heat map difference between experimental and simulated contact probability matrices and a histogram of this difference value respectively. A Pearson correlation coefficient value of 0.88 between experimental and simulated contact probability matrices and an absolute difference of the mean value of 0.067 between them imply reasonably good concurrence with in-vivo and in-silico chromosomal interactions.

B. Loci Mobility and Diffusive Exponents are Chromosome Coordinate dependent

Figure 2 (a) shows the genomic position of different loci in the circular chromosome [3, 4, 25]. Each of the color segments in Figure 2 (a) represents an individual macro domain and non structured region (NSR, NSL). Every macro domain and non structured region contains its own corresponding loci position, giving rise to a total number of 30 loci. To study the dynamics of chromosomal loci, here we have used BD simulations [28] in overdamped condition with a time step δt = 1 x 10^-4 τ_{BD} (see Model and Methods section). We simulate a time-range of (0.1 – 100)τ_{BD}, which corresponds to (1.2s – 20min) (significantly shorter than the doubling time of the bacteria at 30°C in MM). First we have calculated the time averaged MSD for each trajectory (total 200 initial configurations) and subsequently have performed an average over all the trajectories, which we define as ‘ensemble average’ (⟨MSTD⟩). The time profiles of MSD and the MSD exponents enable us to characterize the type of diffusion i.e how fast the particles are spreading in space. The time-average MSD for a particular loci (n) defined as ⟨MSD⟩_n(τ) = ⟨|r_n(t + τ) − r_n(t)|^2⟩, = τ_{tot} ∑_τ | r_n(t + τ) − r_n(t) |^2, where τ and T_{tot} are lag time and total simulation time respectively. For a more appropriate power-law fitting of the ensemble average MSD, we have used (⟨MSTD⟩ = Dτ^α, D is constant). Generally (⟨MSTD⟩ = 6D_{app}τ^α (in 3d). D_{app} is the apparent diffusion coefficient, so we denote D = 6D_{app} and α is the MSD exponent. We have divided the lag time into two regions: (0.1 – 10)τ_{BD} and (10 – 100)τ_{BD}. We have also cross-checked that instead of the whole time lag, power
law fitting is better in these two divided regions.

Figure 2 (b) demonstrates the time averaged MSD, from a single trajectory as a function of lag time for different loci. The figure shows that there is a large spread in the time profiles of MSD across all loci. We first focus our attention on mobilities of two particular loci (Ori2, Ter3). Specifically, figure 2 (c) and 2 (d) compare the ensemble averaged MSD for two loci (Ori2 and Ter3) with time lag \((0.1 - 10)\tau_{BD}\). Scatter points (red for Ori2 and orange for Ter3) are simulated data and dotted lines (gray fit for Ori2 and green fit for Ter3) are fitted data. Ori2 and Ter3 have sub-diffusive exponents 0.49 and 0.54 respectively. For very short time lag \((0.1 - 10)^{-1}\), Ori2 and Ter3 have MSD exponent 0.40 and 0.54 respectively, less than as mentioned in (c). For all the plots MSD and time are in units of \(\sigma^2\) and \(t\) respectively.

Together these figures depict that Ori2 and Ter3 differ significantly from a single trajectory as a function of lag time for different loci. More importantly, the individual power law fitting of the MSD profiles of these two loci reveal that the diffusive exponent is significantly lower than 1, thereby confirming their sub-diffusive behavior. More specifically, consistent with the loci-dependent trend in MSD, the power-law fitting indicates a significant difference in the exponent values between Ori2 and Ter3. For very short time lag \((0.1 - 10)^{-1}\) MSD exponents are given by: \(\alpha = 0.49\) for Ori2 and \(\alpha = 0.55\) for Ter3, which are slightly higher compared to the long time lag \((10 - 100)^{-1}\) MSD exponents \(\alpha = 0.40\) for Ori2 and \(\alpha = 0.54\) for Ter3. Experiment reports [25] that at shorter time scales, the lower and upper limit of these exponent as 0.4 and 0.5 respectively, which is in line to what we also observe from our simulations.

Together, the aforementioned observations of considerable difference in Ori2 and Ter3 movements reveal that both the loci mobility (characterized by MSD values) and the respective values of MSD exponents are significantly dependent on the relative coordinates of these two loci. These results prompted us for a comprehensive investigation of the mobility and the diffusive exponents for thirty chromosomal loci (see Figure 2 (a)) and characterize their trend. Accordingly, we calculated the MSD exponents and apparent diffusion constants from each time averaged simulation trajectories \((30 \times 200 = 6000\) by fitting two different lag time regions \((0.1 - 10)^{-1}\tau_{BD}\) and \((10 - 100)^{-1}\tau_{BD}\) respectively. From these two figures, it is clear that there are significant differences in diffusion coefficients that vary across chromosomal genomic positions in both time scales (short and long). In particular, it is also evident that among all the loci, Ori1 and Ori2 have substantially high diffusion coefficients and most of the Ter loci show lower diffusion coefficients. Figure 3 (c) and 3 (d) illustrate the box plot of diffusion exponents \(\alpha\) of individual loci for short and long lag times respectively. For short lag time, the median values of the

FIG. 2. (a) Different loci position in a circular chromosome according to their genomic coordinates. Different color chunks represent the macro domains and non structured regions (NSR, NSL). (b) Time averaged MSD from single trajectory as a function of lag time for different loci. (c) Power law fitting of ensemble average MSD for two loci (Ori2 and Ter3) with time lag \((0.1 - 10)^{-1}\). Scatter points (red for Ori2 and orange for Ter3) are simulated data and dotted lines (gray fit for Ori2 and green fit for Ter3) are fitted data. Ori2 and Ter3 have sub-diffusive exponents 0.49 and 0.54 respectively. (d) Power law fitting same as (c), with a lag time of \((10 - 100)^{-1}\). Here Ori2 and Ter3 have MSD exponent 0.40 and 0.54 respectively, less than as mentioned in (c). For all the plots MSD and time are in units of \(\sigma^2\) and \(t\) respectively and these are in log scale.
exponents lie higher than $\alpha = 0.5$ and for long lag time this median values lies between $\alpha = 0.4$ to $\alpha = 0.5$ for most of the loci. This results are well corroborated with experimental findings by Javer et al [25], which reported that these median values lie around 0.4. In a similar fashion, figure 3 (e) and 3 (f) depict the histogram plot of MSD exponent for all trajectories of 30 loci for two different lag times. For short lag time, the mean value of the exponent $\alpha_{\text{mean}} \sim 0.502 \pm 0.067$ and for long lag time this value is lower $\alpha_{\text{mean}} \sim 0.447 \pm 0.158$. In both cases, the distribution of MSD exponent ($\alpha$) values is very wide. Taken together, our Hi-C integrated computer model elucidates two key traits of $E. \ coli$ chromosome: a) slow sub-diffusive dynamics and b) heterogeneous chromosomal loci mobility. These results are consistent with previous experiments, [21, 25] and prompt us to elucidate the underlying mechanism of these features.

C. Hi-C contacts as the origin of the heterogeneity in $E. \ coli$ chromosomal diffusivity

The preceding discussions pointed out to a significant loci-coordinate-dependent sub-diffusive motion of chromosome, largely consistent with the previous measurements by Javer et al [25]. These experimentally consistent observations of heterogeneous mobility across chromosomal loci in our computer model demands a more microscopic physical interpretation. We noted that one of the key features of the current model, that sets it apart from other excluded-volume interaction model or other generic polymer-based ‘Rouse model’, is the integral presence of ‘Hi-C’ interaction potentials to capture the experimentally obtained inter-beads contact matrices. We speculated that these Hi-C contact might have a role in dictating the loci-dependent heterogeneity. Accordingly, to dissect its specific role, if any, we performed a fresh set of control dynamical simulations by removing Hi-C contacts from our computer model.

Figure 4 (a) and 4 (b) plot the MSD exponents of all loci, as derived from the simulations trajectories of the model in absence of Hi-C interaction potential, at short and long two lag times. For the purpose of comparison, we also reproduced the MSD exponent of same set of loci as derived in simulations with Hi-C contacts (figure 4 (c) and 4 (d)). The comparison suggests that the MSD exponents are uniformly similar across all loci in absence of Hi-C contacts. However, turning on these interactions in the model recapitulates the loci-dependence of MSD exponents. On a similar spirit, to investigate this loci-specific heterogeneity in diffusion coefficients, we have also calculated this diffusion constant value from simulated trajectories in the absence of Hi-C interactions. Figure S2(a) and Figure S2(b) demonstrate the appar-
ent diffusion coefficient values as a function of chromosomal coordinates, in the absence of Hi-C interactions with lag time \((0.1 - 10)\tau_{BD}\) and \((10 - 100)\tau_{BD}\) respectively. When compared with the results of simulated data in presence of Hi-C contacts (Figure S2(c) and Figure S2(d)), it is obvious that without Hi-C contacts, all the loci display nearly the same diffusion constant values, with two different time regions. In combination, these analysis, based on control simulations, dissect the decisive role of specific Hi-C contacts in capturing the inherent loci-dependent mobility.

The resolution of our model and the explicit incorporation of the Hi-C interaction potential allow us to track specific inter-loci contacts and pinpoint the underlying reason for the coordinate-dependent loci dynamics. In the current protocol, the inter-genomic distance is modeled as inversely proportional the Hi-C contact probability \((D_{ij} = \sigma/P_{ij})\). Accordingly, the number of interbeads contacts connected to each loci gets modulated. In particular, Ori1 and Ori2 loci show one of the higher mobilities, while most of the Ter loci (Ter2, Ter3, Ter4, Ter5, Ter6). Diff, NSR3 and NSL3 show the slowest mobility. On the other hand, some loci (NSR3, Right4, Right5, Ter3, Ter4, NSL2, NSL3 and Ori3) are showing higher exponent \((\alpha > 0.5)\) values and some of them (Right1 and Ter1) are showing lower exponent \((\alpha < 0.4)\) value, but the values of these exponents lie around \(\alpha \sim 0.44 \pm 0.158\), for longer lag time.

To get a closer insight on why Ter macro-domain (MD) demonstrates comparatively lower mobility compared to other MDs, we have computed the fraction of connections between different MDs and NS regions in our model. Figure S3 highlights the heat map of the fraction of Hi-C contact formed within MDs and NS regions. This suggests that the fraction of intra-chromosomal connections within the Ter MD is relatively higher compared to other MDs, which slows down the mobility of Ter and its loci. On a related note, prior investigations [6, 29] refer that the presence of nucleotide associated protein MatP in the Ter domain plays an important role in insulating it from the other MDs and making it more compact [30]. We wanted to explore if the current model can predict the role that MatP would have played in modulating the loci dynamics of E. coli. Accordingly, we made use of Liyo et al’s experimentally measured Hi-C data of mutated bacteria devoid of MatP [13] and performed a fresh set of simulations for capturing the dynamics of MatP-devoid chromosome (ΔMatP). (see SI Model and Methods) Figure S4(a) and S4(b) depict the diffusion constant values of all loci with wild type (WT) and ΔMatP mutant at two different lag times, \((0.1 - 10)\tau_{BD}\) and \((10 - 100)\tau_{BD}\) respectively. For mutant MatP case, most of the loci (except some loci in Ter MD) show lower diffusion mobility compared to the WT. We believe that this mostly results from the fact that the absence of MatP reduces the insulation of Ter from the other MDs, which introduces additional contacts between MDs, especially with Ter MD, thereby further slowing down the mobility of chromosomal loci in ΔMatP chromosome. To verify the compactness of Ter MD, we have calculated distance \((D_{ij})\), between \(i^{th}\) and \(j^{th}\) beads in Ter MD for WT and ΔMatP cases, and plotted the distribution of this \(D_{ij}\) in figure S5. From this figure, it is clear that the mean value and standard deviation are large for ΔMatP compare to WT, which apprises that Ter MD is less compact for ΔMatP compare to WT. We have also reported the total number of connections in our model between different MDs and NS regions in table S1 and table S2 for WT and ΔMatP respectively.

In a bid to investigate how frequent and which particular pair of chromosomal loci meet with each other for the first time, we have calculated first passage time (FPT) for the chromosomal loci from each of the total 200 trajectories. Figure 5 (a) and (b) compare the distribution of FPT for all the loci with all trajectories simulated with Hi-C and without Hi-C interaction respectively. In presence of Hi-C interactions, numerous loci meet each other significantly and more quickly for the first time compared to that in absence of Hi-C, resulting in much shorter mean first passage time (MFPT) of 27.14τ_{BD}. (compared to al-
most double value in absence of Hi-C). More importantly, a wide distribution of FPT (with long tail) (5 (a)), evident in presence of Hi-C, reflects an inherently heterogeneous, Hi-C interaction-modulated encounter between the loci. Figures 5 (c) and (d) demonstrate the network plots of FPT of chromosomal loci in presence and in absence of Hi-C respectively, with the thickness of the lines representing the value of average FPT. The presence of non-specificity, largely uniform encounters between the loci is clearly evident when the Hi-C interactions are turned off.

D. Tracking the origin of sub-diffusive motion: Spatial and Temporal Coherence

Correlation functions are popular approaches, which provide dynamical insights for a system. Mathematically, velocity auto-correlation function (VAF) is defined as:

\[ C^{\Delta t}_{\text{v}}(\tau) = \langle \vec{v}(t+\tau) \cdot \vec{v}(t) \rangle_{t}, \]

where angular brackets denote the time average and \( v(t) = \frac{1}{\Delta t} (\vec{r}(t+\Delta t) - \vec{r}(t)), \) \( \Delta t \) are the different time intervals for velocity calculation and \( \tau \) is the lag time. Physically VAF provides the strength of correlation between two velocities of the particle separated by a time interval \( \tau \). Here we have calculated an ensemble averaged VAF for the genomic midpoint of the chromosome. Previously experimental and simulations studies on \textit{E. coli} Chromosome \cite{21, 22, 31} have revealed the presence of a negative correlation peak at \( \tau = \Delta t \) \cite{32}, which slowly goes to zero (\( \tau \gg \Delta T \)), due to the viscoelastic nature of cytoplasm. Figure 6 (a) and 6 (b) manifest VAF of \textit{E. coli} chromosome for small and large time intervals \( \Delta t \) respectively. The observation of negative correlation peak in our model is qualitatively consistent with the previous experiments and implies that our model effectively mimics a viscoelastic cytoplasm, even without explicit incorporation of viscoelasticity as well as a memory kernel \cite{33}. We believe that the presence of long-range inter-bead connectivity in the form of ‘Hi-C’ contacts, is an essential trait of our model, is decisive in furnishing the memory and viscoelasticity in an implicit way. Instead of the genomic midpoint of the chromosome, we have also calculated VAF for Ori2 and Ter3 loci (figure S6 (a), (b), (c), and (d) in SI), which also show the same type of behavior.

Motivated by Zidovska et al.’s work \cite{34}, where they have used displacement correlation spectroscopy (DCS) to study Spatio-temporal evolution of the global chromatin dynamics in vivo in nuclei of human HeLa cells, we also used the same method for spatial correlation calculation. Spatial correlation of the chromosomal loci defined as (equation 1 \cite{35})

\[ C^{\Delta t}_{s}(r) = \left\langle \frac{\sum_{i>j} [\Delta \vec{r}_i(t; \Delta t) \cdot \Delta \vec{r}_j(t; \Delta t)] \delta (r_{i,j}(t) - r)}{\sum_{i>j} \delta (r_{i,j}(t) - r)} \right\rangle_t \]
FIG. 6. Velocity auto-correlation function (VAF) as a function of lag time for different time interval (a) $\Delta t = (1.0 - 9.0) \tau_{BD}$ and (b) $\Delta t = (10 - 90) \tau_{BD}$. Both curve shows that there is a negative correlation peak at $\tau = \Delta t$ and slowly goes to zero ($\tau >> \Delta t$), due to the viscoelastic nature of cytoplasm. (c) Spatial correlation $C^{\Delta t}_{\mathbf{r}}(r)$ as a function of distance $r$ for different time interval $\Delta t$. As time difference $\Delta t$ increases, the correlation function decay more slowly. (d) Spatial correlation $C^{\Delta t}_{\mathbf{r}}(r)$ fitted with exponentially decay function as $f(r) = A \exp(-r/\zeta)$ Where $A$ is the constant value and $\zeta$ is correlation length. $\zeta$ increases with small values of $\Delta t$, but for large values of $\Delta t$, $\zeta$ values are almost saturated. (e) Vector field of displacement vector for time interval $\Delta t = 1$. (f) Vector field of displacement vector for time interval $\Delta t = 90$. Color bars represent the angle ($\theta$) of the displacement vector. For a small value of time interval, $\Delta t$ displacement vectors are more irregular compare to large time interval $\Delta t$. All length and time values are in units of $\sigma$ and $\tau_{BD}$ respectively.

Where $i$, $j$ are the monomer indices, $\Delta \mathbf{r}_i(t; \Delta t) = \mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)$ and $\Delta t$ are the differences in time. Physically this correlation function provides the information on how the displacements of the chromosomal loci, separated by a distance $r$, over time interval $\Delta t$, are correlated. Figure 6 (c) shows this spatial correlation $C^{\Delta t}_{\mathbf{r}}(r)$ as a function of $r$. As time difference $\Delta t$ increases, the correlation function decays more slowly. We have also fitted $C^{\Delta t}_{\mathbf{r}}(r)$ by exponentially decay function as equation 2

$$f(r) = A \exp(-r/\zeta)$$ (2)

Where $A$ is a constant and $\zeta$ is the dimension of length, called correlation length (All fitting plots and parameters are given in Figure S7 and table S3 in SI). Figure 6 (d) demonstrates the correlation length ($\zeta$) as a function of time difference $\Delta t$. For small values of $\Delta t$, $\zeta$ values are increasing with $\Delta t$, but for large values of $\Delta t$, $\zeta$ values are almost saturated. To quantify the spatial correlation, we have projected the displacement vector ($\Delta \mathbf{r}_i(t; 0; \Delta t)$ with $\Delta t = 1 \tau_{BD}$ and $\Delta t = 90 \tau_{BD}$) onto the xy plane with a condition ($-\sigma \leq z \leq \sigma$), and represented them by a vector field. Figure 6 (e) and 6 (f) depicts the vector field of displacement vector with $\Delta t = 1 \tau_{BD}$ and $\Delta t = 90 \tau_{BD}$ respectively. From these two figures it is clear that displacement vectors are more random for small $\Delta t$ compared to large $\Delta t$, which assist to increase the correlation length with $\Delta t$.

E. The Heterogeneity in loci mobility is robust in presence of active noise

Generally, in normal diffusion, we assumed that the systems are in thermal equilibrium, but in a real scenario, there are numerous ATP-dependent biological activities (transport, metabolism) which prompts the cells to proceed far from equilibrium. A series of numerous past experiments [36–41] on eukaryotic cells and polymer model [42–45], have divulged the role of these biological active processes that initiate non-thermal fluctuations. On a related note, an important experiment by Weber at al. [24] has revealed that ATP-dependent fluctuation (non-thermal) control the in vivo dynamics of E. coli chromosomal loci. However, the investigation of dynamics of chromosome inside the untreated and ATP-depleted E. coli cells suggested that the sub-diffusive scaling exponent ($\alpha_{untreated} = 0.39, \alpha_{treated} = 0.40$) is almost unchanged irrespective of presence or absence of ATP. However the MSD and $D_{app}$ (apparent diffusion coefficient) values are significantly reduced in ATP-depleted cell.

In light of present investigations’ observation that chromosomal loci display a heterogeneous and coordinate-dependent mobility, we wanted to explore if this result would be replicated in presence of active noise, which is a better representative of in-vivo environment. To introduce this into our computer model, we increased the strength of the thermal noise and performed a fresh set of dynamical simulations of the chromosome. In particular, the fluctuation dissipation theorem was accordingly
FIG. 7. (a) Log-log plot of MSD values as a function of lag time for Ori2 loci with active (red) and passive (blue) noise. (b) Log-log plot of MSD values as a function of lag time for Ter3 loci with active (red) and passive (blue) noise. MSD values of these loci are significantly different when compared with active and passive noise.

changed to \( \langle \vec{R}(t) \cdot \vec{R}(t') \rangle = 12D\delta(t - t') \) [35, 46–48] instead of the usual expression \( \langle \vec{R}(t) \cdot \vec{R}(t') \rangle = 6D\delta(t - t') \).

We term our original simulations (which obey fluctuation dissipation theorem) as ‘active’ and current simulations (which violate fluctuation dissipation theorem) as ‘passive’. Figure 7 (a) and figure 7 (b) show the MSD value as a function of lag time for two loci Ori2 and Ter3 respectively, with active and passive noise. From these two figures, it is evident that the MSD values of these loci are notably higher in presence of ‘active’ noise than that of ‘passive’ noise. A more complete comparison across all 30 loci is presented in figure 8. We find that the apparent diffusion constant (\( D_{app} \)) of each chromosomal locus is significantly increased (almost double) in presence of ‘active noise’ (figure 8 (a-b)) in both short \((0.1 - 10)\tau_{BD}\) and long \((10 - 100)\tau_{BD}\) time lag respectively. However, very interestingly, as demonstrated by Figure 8 (c) and figure 8 (d), the respective sub-diffusive exponent (\( \alpha \)) for individual loci remains unchanged in presence of active and passive noise, at both short \((0.1 - 10)\tau_{BD}\) and long \((10 - 100)\tau_{BD}\) time lags. Together these results indicate that due to the active noise or ATP-dependent biological activities, mobility of chromosomal loci change remarkably but sub-diffusive exponents are almost fixed. The observation of a robust MSD exponent irrespective of active or passive cytoplasm is consistent with previous experiment and might be due to the fact that viscoelastic nature of cytoplasm, which is effectively captured by our model’s Hi-C interactions, does not change notably even in presence of active noise. Nonetheless, figure 8 also asserts that the inherent heterogeneity and loci-dependence of the mobility and MSD exponents, as found in otherwise ‘passive’ simulations, are rigorously preserved even in presence of ‘active’ noise.

III. CONCLUDING REMARKS

Our Hi-C data integrated beads-spring polymer model with Brownian dynamics simulations, reveals spatio-temporal dynamics which provides a close-view interpretation of previous experimental and theoretical findings on prokaryotic [21–25] chromosomal loci. We observed that chromosomal loci are moving sub-diffusively but there is clear heterogeneity in sub-diffusive exponents with respect to genomic co-ordinates as well as time. Especially Ori2 and Ter3 differ significantly in their individual displacements. The mobility of each and every loci is crucially dependent on chromosomal co-ordinates. We have also calculated the mean value of sub-diffusive exponent by taking \( \alpha \) value from each loci and making a histogram. The mean value is \( \alpha_{mean} \sim 0.502 \pm 0.067 \) for small lag time and \( \alpha_{mean} \sim 0.447 \pm 0.158 \) for large lag time.

To deduce the origin of the heterogeneous, coordinate-dependent diffusion, we have run a fresh set of control simulations by turning off the Hi-C interaction potentials. We found that, in absence of Hi-C contacts, chromosomal loci are moving homogeneously and their mobility are unchanged with respect to genomic co-ordinates. This observation provides strong evidence of long-range inter-loci communications, as manifested by the ‘Hi-C interactions’, as one of key modulators of the local dynamics of the bacterial chromosome. The observation of a negative peak in velocity auto correlation function a around particular time difference \( (\tau = \Delta t) \) brings out the underlying viscoelastic nature effectively furnished by the Hi-C integrated model.

Quite gratifyingly, the sub-diffusive exponents are robust with respect to active and passive noise, while the mobility of loci changes significantly in presence of active noise. This robust MSD exponent indicates that in presence of active noise, viscoelastic nature of cytoplasm, effectively captured by our model interaction, does not change notably.

The self-organization of E. coli chromosome into non-overlapping macro-domains has remained a key structural hallmark of this archetypal bacteria. However, how a domain-separated nucleoid would be influencing the dynamics of the chromosome has remained a key question for a while. This is especially complex, considering the intricacies related to segregation and the loci mobilities. In this respect, while the recent experimental demonstration of coordinate-dependent loci movements have been a intriguing discovery, underlying origin has so far been elusive. The current work brings out the importance of inter-loci encounter, statistically represented by Hi-C contacts, in dictating the heterogeneity of the local, coordinate-dependent chromosomal dynamics. The analysis of statistical encounters in the form of first passage time of contacts (figure 5) and the graded extent of macro-domain-dependent fraction of contacts (figure S3), provide a fair idea on how local chromosomal dynamics is governed by the long-range contacts. The dynamical analysis on MatP-devour chromosome points towards a crucial role that nucleotide-associated proteins (NAPs) might play a role in the local chromosomal dynamics. Overall, the current set of investigations reported in this article enriches our understanding of E. coli’s chromosomal dynamics, as well as the extent of information Hi-C captures.

Future extension of the present work via introduction of other macro-molecules such as ribosomes and proteins into the model would enable investigations into the diffusion and localization of such macro-molecules [49, 50] at different levels of chromosomal compaction which are brought about by multiple factors such as varying cell sizes [51], induced stress on the cell [49] and mutations.
IV. MODEL AND METHODS

A. Model Details

In Hi-C measurement, the cells are in an ensemble of different replication stages with their respective cell cycles. Here we are interested in short-time chromosomal loci dynamics in minimal medium (wt30MM), i.e. there is only one single chromosome and no replication fork. We model the E. coli chromosome as a bead-spring polymer chain, with each bead representing 5 × 10^3 bp (5 kb), the same as our previous work [14]. This resolution is also the same as Hi-C interaction maps reported by Lioy et al [13]. So the number of beads present in the polymer chain is 928 (∼6.4 × 10^6).

B. Interaction Potentials

To mimic the confinement in the E. coli cell, we have taken the polymer chain in a spherocylindrical confinement with average length L = 3.08 μm (including two end caps) and diameter d = 0.82 μm. By taking approximate volume fraction of chromosome φv = 0.1 [52], we have calculated the bead diameter σ to be 0.06731 μm (SI). Non-bonded interactions of the polymer beads are modeled by the repulsive part of the Lennard Jones (LJ) potential i.e. \( V_{nb}(r) = 4ε(σ/r)^12 \). Where \( ε \) is in the unit of \( k_B T \) and \( r \) is in the unit of \( σ \). Bonded interactions between adjacent beads have been modeled by harmonic springs with a spring constant \( k_{spring} = 300k_B T/σ^2 \). In a similar fashion, Hi-C contacts are also modeled as harmonic springs with distance dependent force constants and probability dependent bond lengths.

From the Hi-C contact probability matrix we can calculate the distance matrix \( D_{ij} \) as

\[
D_{ij} = σ/P_{ij}
\]  

(3)

where \( i \) and \( j \) are row and column index of the matrix respectively. From this distance matrix we define a restraining potential between a pair of Hi-C contacts at a separation of \( r_{ij} \), as \( V_{Hi−C}(r_{ij}) = \frac{1}{2}k_{ij}(D_{ij} - r_{ij})^2 \). Here, \( k_{ij} \) is distance dependent force constant that can be calculated from equation 4.

\[
k_{ij} = k_0e^{-\frac{(D_{ij} − σ)^2}{σ^2}}
\]  

(4)

Here \( P \) is a sparse matrix with a large number of elements that are 0. So from equation 3, it implies that lots of elements become infinity. To avoid this situation we have modeled \( k_{ij} \) as a Gaussian function. This relation implies that the force constant becomes smaller for larger distances and the \( k_0 \) value is the upper bound of the force constant. We have used a value of \( k_0 = 10^{12} \). For optimization of \( w \) we have calculated the Pearson Correlation coefficient between the experimental and simulated contact probability matrices with varying \( w^2 \). Pearson correlation coefficient values show a maximum (88%) at \( w^2 = 0.3 \). For the spherocylindrical confinement mimicking the cell wall, we have used a restraining potential in the form of 5.

\[
V_{res}(r; R_0) = \frac{1}{2}k_{res}|\vec{r} - \vec{R}_0|^2\mathbf{H} (|\vec{r} - \vec{R}_0|)
\]  

(5)

Here \( R_0 \) is the center of the spherocylindrical confinement and \( k_{res} \) is the force constant depicting the relative ‘softness’ of the confinement. We have used a value of \( k_{res} = 310k_B T/σ^2 \). \( \mathbf{H} \) is a step function that will activate if a polymer bead gets out from the spherocylindrical confinement. Therefore the Hamiltonian can be written as 6.

\[
V_{tot} = V_b(r) + V_{nb}(r) + V_{Hi−C}(r_{ij}) + V_{res}(r; R_0)
\]  

(6)
Where $V_b(r)$, $V_{nb}(r)$, $V_{H-C}(r_i)$, and $V_{res}(r; R_b)$ are bonded, non-bonded, Hi-C restraining, and confinement restraining potential respectively.

### C. Simulation Details

To study the dynamics of chromosomal loci, here we have used Brownian Dynamics (BD) simulations (over-damped condition), by integrating equation 7 (Euler scheme)

$$
\frac{d\vec{r}_i}{dt} = -\frac{D_i}{k_B T} \nabla V_{tot} + \vec{R}_i(t)
$$

(7)

Where $D_i$ is diffusion coefficient of $i$-th bead and $\vec{R}_i(t)$ is a random noise, satisfying fluctuation dissipation theorem i.e $\langle \vec{R}_i(t) \cdot \vec{R}_j(t') \rangle = 6D_i \delta_{ij} \delta(t - t')$. We can calculate $D_i$ from the Stokes-Einstein equation

$$
D_i = \frac{k_B T \eta}{6\pi r_i}
$$

Where $\eta$ is the E. coli cytosol viscosity and value is 17.5 Pa-s $[53, 54]$ (26000 times greater than the viscosity of water). Here all simulations time in the unit of $\tau_{BD} = \frac{\sigma^2}{\kappa T} \approx 12s$ (SI for unit conversion). Integration of the 7 was performed with a time step $\delta t = 1 \times 10^{-4} \tau_{BD}$ by setting $k_B T = 1$. Here we have described the simulations details for wt30MM in presence of Hi-C. For wt30MM in absence of Hi-C, $\Delta$MatP and active noise simulation protocol see SI.

All simulations were performed using the open source package GROMACS 5.0.6 [55] and we have modified the source code to introduce the spherocylindrical confinement. We have taken 200 different initial configuration (ensemble) of the polymer chain and minimized the energy of topology by steepest descent algorithm. After energy minimization, we allow the chain for equilibrium with $1 \times 10^5 \tau_{BD}$ time steps. Then we collected the data with time step $1 \times 10^3 \tau_{BD}$ (data dumping frequency).

### V. ACKNOWLEDGMENTS

We sincerely acknowledge Dr. V. Liouy for providing us the average cell size of the cells grown in MM at 30°C used for their study. We acknowledge the computation facilities provided by TIFR Centre for Interdisciplinary Sciences, India. We acknowledge the support of the Department of Atomic Energy, Government of India, under Project Identification No. RTI 4007.


