1. Introduction

The organization of DNA at mesoscopic length (>30 nm) scales has been a topic of intense research in the past decade [1–19], especially after the work of Lieberman Aiden et al [2] where the authors mapped out the spatial proximity maps of DNA segments of the human genome (each segment of length 1 mega base pairs) inside the nucleus using a technique called Hi–C: high-throughput sequencing. The experimental studies provide a contact map of DNA segments [2, 6, 7, 20–22]. A contact map is a color map that shows which DNA segments, numbered \( i = 1,2,3...N_D \) are spatially close to other DNA segments \( j (j = 1,...,N_D) \) with high/low frequency.

The physics approach is to consider DNA as a polymer chain [4, 8, 15, 17, 18, 23–29, 30, 31]. The resolution of Hi–C experiments has increased to 1 kilo-BP (kilobase pair) which is still above the persistence length of naked ds-DNA (approximately, 50 nm with 150 BPs) as reported for bacterial cells [20]. However, the inside cells around 150 BPs of DNA wrap around histone-like proteins (for bacterial cells) to form a higher order structure [32], and the persistence length of a DNA polymer chain is still debated \( \text{in vivo} [33] \)
The DNA-polymer organization at large length scales can be viewed as the organization of a flexible polymer. The large length scales of ds-DNA in question are 100 nm to microns, such that a DNA-segment consisting of a kilo to mega BPs can be considered as a coarse-grained monomer in a bead-spring model of the polymer chain. Our work aims to elucidate the structure of DNA-macromolecule at this length scale, or equivalently that of a flexible polymer with added constraints. Generically we show that adding spatial constraints by cross-linking a minimal number of specific monomers along the length of chain can lead to the organization of an entire long ring-polymer into a specific structure, but there are fluctuations in the structure due to thermal energy.

Much of the research is focussed on the structure and organization of the DNA during the interphase stage [5, 8, 11, 16, 23, 24, 34–37]: the stage of the cell cycle when the cell does not divide into daughter cells. It is also known that the individual chromosome is not arranged as a random walk polymer. From the data of the contact map, we observe that some DNA segments have a much higher spatial association with other chain segments and show up as the presence of so-called topologically associated domains (TADs) [2, 7, 20] in the contact map. A DNA-macromolecule is not just a long polymer chain within a nucleus or a cell (for bacteria) but there are also various proteins and enzymes doing various functions. For example, there are DNA-binding proteins which attach two different and specific segments of DNA chain together and the enzyme topoisomerase which allows chains to cross each other by suitably cutting and rejoining chains [25]. Polymer physics principles are suitably adapted to incorporate the effects of protein and enzyme related activity when investigating the origin or reasons of formation of TADs [1, 8, 11, 14, 23, 24]. Furthermore, when studying DNA-polymer organization in the interphase stage, we assume the system to be in a state of local equilibrium so that the principles of statistical mechanics can be applied.

Studies have shown that the organization of DNA is a fractal globule rather than an equilibrium coil [23, 24]. The understanding is that segments of DNA get locally collapsed to form unentangled crumpled sections of the coil, such that within a segment there are many contacts, and fewer number of contacts between collapsed neighboring segments. These then show up as TADs in the contact map.

In recent years, there have been more detailed polymeric models which can reproduce the experimentally measured TADs for sections of DNA [15, 18, 35]. The most successful of these are the SBS (strings and binders) model [18, 38] and the loop extrusion model [15, 39, 40]. In the SBS model, monomers along a chain have the same size but have distinct affinities of attraction for freely diffusing binder-molecules. There are as many distinct kinds of binder molecules as different kinds of monomers. Monomers of the same kind but separated along the chain contour can get attached to the same binder molecule to result in the formation of loops. Some parameters, such as the number of different kinds of monomers or the number of monomers of each kind, can be optimally chosen to reproduce and fit the TADs of a particular segment of a DNA by solving a multi-dimensional optimization problem.

In the loop extrusion model, there are boundary elements (BE monomers) at specific sections along the chains. A pair of special monomers (LE-monomers), which probabilistically bind with each other to extrude loops of variable lengths by diffusing/translocating along the length of the chain, but LE-monomers are constrained to remain bounded between the BE-monomers. Again, a search through a large parameter space leads to optimal TADs with a quantitative match with experimental data. Both models seem to crucially depend on the formation and contact between suitably sized loops at appropriate locations, which in turn results in a match with experimental contact map data. Other researchers [41, 42] use optimization algorithms with weighted constraints to get an idea of the large scale structure of bio-molecules.

Instead of investigating the origin of TADs, where some headway has already been made, we ask a different question. Given the contact map, can we predict the global spatial organization of a polymer? Is there even any organization of the entire macromolecule? Note that the contact map does not give information on the spatial organization of polymer, it just gives the frequency of finding different DNA-sections in proximity. In this study, we assume that polymer-sections with the highest frequencies of contact are permanently cross linked and investigate if cross-links (CLs) above a minimal number and at special biologically well-determined locations along the chain (as determined from the contact map) play a vital role in giving shape and structure to the entire DNA-polymer.

We compare the organization of a polymer with CLs at biologically determined positions along the contour (bio-cross-links: BC) with polymer organization with an equal number of cross-links between monomers, but the monomers to be cross-linked are chosen at random (random cross-links: RC). The question is whether an equal number of minimal constraints at random positions give ‘structure’ to the polymer? We generate ten independent random configurations of cross-links (CLs) and compare the ‘structure’ of polymers for ten independent RCs and 1 BC.

It is easier to work with simpler systems, e.g. the DNA of bacteria such as *Escherichia coli* (*E. coli*) which is a ring polymer. Bacterial cells have no nucleus, the number of chromosomes is 1 or 2 per cell, and the DNAs are much shorter. Bacterial DNA also shows TADs, and they have DNA-binding proteins [7]. We choose a bead-spring flexible polymer model of *E. Coli*, a bacteria with a single chromosome for our studies. The question is how we determine if a polymer, which is expected to be unstructured, is structurally organized or not.

We start our Monte Carlo (MC) simulations from nine independent initial configurations of ring polymers without taking into account that cross-linked monomers should be in proximity. However, cross-link potentials are applied between monomer pairs from a particular CL-set. We then allow the chain to relax to equilibrium for each case in nine independent MC runs. If we observe that the DNA polymer relaxes to almost the same ‘structure’ within statistical fluctuations in each case, then we could claim that the polymer is organized. Also, we try out with fewer CLs and check the minimal number of CLs required to achieve organization of the polymer. To check for the organization, we had to come up with different
structural quantities which we describe in the next sections. Our hypothesis of special cross-links in DNA-polymers can be further established by testing it on more bacterial chromosomes. We also have results to establish that DNA-polymer of the bacteria Caulobacter crescentus gets organized into a particular structure due to the presence of CLs in appropriate locations suitable for C. crescentus [43].

We have not included confinement effects due to cell walls because we wanted to focus only on the effects of having CLs at specific locations unencumbered by any other competing effects. Also, we do not put the effects of supercoiling in our simple bead-spring polymer model because proximity effects between segments due to supercoiling, if any, should show up in the contact maps.

The organization of the manuscript is as follows: the next section, viz., section 2 discusses the model of the DNA-polymer, the computational method by which we generate initial conformations, methodology to relax polymers to local equilibrium using after which we calculate ensemble averages of statistical quantities to identify spatial organization of the polymer. The next section, section 3, discusses the statistical quantities and our results by which we arrive at our conclusions. We end with summarizing our conclusions in section 4.

2. Model and simulation method

We use Monte Carlo simulations to explore the different microstates of the DNA chain. The DNA of bacteria E. coli is a ring polymer. We model E. Coli DNA with a bead-spring model of ring polymer with \( N_{BC} = 4642 \) number of monomers in the ring. Thus, each coarse-grained monomer bead in our model represents \( 10^3 \) BPs (\( \sim 100 \) nm), which matches with the experimental resolution. The DNA model-polymer is placed near the center of the simulation box of size \((200a)^3\) with periodic boundary conditions (PBCs). The quantity \( a \) is the unit of length and is the average distance between two neighboring monomer beads along the chain contour. The box size \( L = 200a \) is chosen to be much larger than the expected maximum diameter of the polymer coil. The diameter \( \sigma \) of monomer bead is chosen to be \( 0.2a \). The Lennard-Jones potential, suitably truncated at \( r = 2^{1/6} \sigma \) and shifted (the Weeks Chandler Andersen-WCA potential), is used to model the excluded volume interaction between the monomers. A harmonic spring potential connects adjacent monomers along the chain contour. \( V(r) = \kappa(|\vec{r}|-a)^2 \), where \(|\vec{r}|\) is the distance between two monomers and \( a \) is the unit of length.

We have chosen \( \kappa = 200k_B T/a^2 \), where the thermal energy \( k_B T = 1 \) is the unit of energy in our simulations.

Using data from [21] and subsequent analysis methods which are described in detail in the appendix, we obtained the frequency of finding two segments of E. Coli-DNA spatially close to each other. We use this data from contact maps as an input to our simulations. The experimental resolution of the size of segments is \( 10^3 \) base pairs. The model monomer in our simulations represents a DNA segment exactly of the same size as the experimental resolution. We cross-link monomers whose frequency of being in spatial proximity is greater than the threshold frequency \( p_c \). Depending on the value of threshold frequency that we choose, we can have (a) \( N_{CL} = 47 \) or (b) \( N_{CL} = 159 \) pairs of monomers of the DNA-polymer that we cross-link. We bind these pairs of monomers together by an additional spring potential \( V_c = \kappa_c (r-a)^2 \) with \( \kappa_c = 200k_B T/a^2 \). The cross-linked monomers are held together at a distance of \( a \), but the different CLs can move with respect to each other as the chain explores different conformations.

The set of 47 CLs from biological contact-map data, which we refer to as BC-1 in the rest of the paper, are a subset of 159 CLs which we call BC-2. To analyze whether the overall mesoscale organization of the chain is determined primarily by a particular choice of CLs, we start our simulations from nine independent initial conditions. Instead of having different random coil initial polymer configurations with bonds between neighbouring monomers and cross-link constraints intact, we by hand design different initial conditions of positions which account for bonds along contour but not the cross-links between selected monomers. In the different initial conditions that we design, the cross-linked monomers are located at different positions with respect to each other. The distance between a pair of cross-linked monomers will be different in each of the nine different configurations and likely to be much larger than \( a \). For example, in one of the initial conditions, the monomers of the ring polymer are arranged along a circle of radius 30.73a such that one circle has 193 monomers. The circles of monomers are stacked up to form a cylinder. Note that this will lead to monomer numbered as 1 and the last monomer \( N = 4642 \) to be at a distance much larger than \( a \) though it is a ring polymer. Also, the monomers which form CLs can be at distances much larger than \( a \) as they are arranged along the cylinder. But these will come closer due to the presence of harmonic spring potentials acting between monomer-pairs as the polymer is allowed to relax during the MC run. In two other initial conditions, we arrange the monomers in circles of radius 40.92a and 36.94a. For the next three initial conditions, the monomers are arranged in squares of side 90a, 80a and 70a; these squares are then stacked up. For the last three initial conditions, we arrange monomers in equilateral triangles of side 40a, 50a, 60a and stack them to form a vertical column.

The question we then ask is: as the chains relax from their initial conditions to their equilibrium conformations in different Monte Carlo runs, do all of them organize themselves in some particular set of conformations (in a statistical sense), though the initial configurations were very different? If they do, we can expect the presence of CLs to play a significant role in the organization since a normal ring polymer is not expected to show structural organization.

We use additional techniques to allow the chain to relax slowly over \( 10^7 \) Monte Carlo iterations to its equilibrium state without allowing the system to get stuck in some entangled and metastable state. We set a spring constant of cross-links \( \kappa_c^{\text{initial}} = 0.01\kappa_c \), where \( \kappa_c = 200k_B T/a^2 \) at the start of the simulation and gradually ramp it up in steps of 0.01\( \kappa_c \) every 1000 MC steps, as the CL monomers approach each other.
in the relaxation process. In a standard Metropolis step, a monomer attempts a displacement \( \delta \overrightarrow{r} = \delta (r_1 \hat{i} + r_2 \hat{j} + r_3 \hat{k}) \) in a random direction, where \( \delta = 0.2\sigma \) and \( r_1, r_2, r_3 \) are random numbers. The attempt is accepted with Boltzmann probability. In addition, every 100 iterations, we attempt displacements with \( \delta = 1.2\sigma \). This helps chains to cross each other at times and overcome topological constraints which might arise as the chain relaxes from its initial condition. Chain crossing is also helped by the relatively small diameter of the beads compared to the bond-length.

We monitor the potential energy as the chain relaxes. The value of energy relaxes to the same value at the end of \( 10^5 \) iterations from the nine different runs, see figure 1. It gives us confidence that the chain conformations are not stuck at metastable energy minima. From this initial state, we evolve each of the nine different chain conformations in independent simulation runs over the next \( 12 \times 10^6 \) iterations and collect data to calculate and compare structural quantities. We carry out this comparison of statistical data from nine independent runs for each set of CLs, viz., chains with (a) 47 (b) 159 CLs.

In addition, we also carry out similar calculations starting from nine independent initial configurations for each of the ten distinct sets of randomly chosen position of CLs (monomers which are cross-linked together are chosen randomly from the list of monomers). The effective number of CLs for RC-1 and RC-2 correspond to the number of CLs in BC-1 and BC-2. In BC-1 and BC-2 set of CLs, there are some CLs which are not independent. For example, monomer number 16 and 17 are cross-linked to monomer 2515 and 2516, respectively. One cannot consider them as distinct CLs. The list of cross-linked monomers is given in table 1 of supplementary section table I. Hence there are fewer effective CLs than the number of CLs in BC-1 and BC-2. Thus we compare the results of our simulation from bio-CLs (BC-1 and BC-2) with an equal number of effective random CLs. In each of these random set of CLs, we have the same number of effective CLs as the ones obtained from biological data, which is less than the corresponding number of CLs in BC-1 and BC-2. Hence the list of randomly positioned CLs have just (a) 27 effective number of CLs (we refer to these as RC-1) and (b) 82 effective CLs (referred as RC-2), corresponding to 47 CLs in BC-1 and 159 CLs in BC-2. We can now compare the structural data obtained from polymer simulations using BC-1 and RC-1 on the one hand, and BC-2 and RC-2 on the other.

3. Results

Now we discuss the statistical quantities which we use to investigate the structure and conformation of the ring polymer. We aim to check if statistical quantities from nine different runs with the same set of CLs give similar results to infer that the polymer has a similar shape and conformation across runs. We further compare data from ten different RC-1 and RC-2 CL sets with data from E. Coli CL set BC-1 and BC-2, though in this manuscript we show data primarily from one representative RC set.

The first quantity we want to estimate is the size and extent of the polymer with CLs. To that end, we calculate the moment of inertia tensor \( \mathbf{I} \) with respect to the center of mass (CM) of the polymer coil and diagonalize the matrix to get its principal moments for each microstate. We then calculate the average principal moments \( I_1, I_2, I_3 \), where \( I_1 \) is the largest eigenvalue and \( I_3 \) the smallest.

In figure 2(a) we show the values of \( (I_1/I_3) \) for different random CL-sets but with the same number of CLs as RC-1 and RC-2. For each random CL-set, the average is taken over nine independent initial configurations. In subplot (b) we show \( I_1/I_3 \) for biologically determined CLs: BC-1, BC-2 for nine independent initial conditions. In plot (c) \( R_g = \sqrt{(I_1 + I_2 + I_3)/3M} \) for different random CL sets with the same number of CLs as RC-1 and RC-2 is shown and in subplot (d) we show \( R_g \) for nine independent initial conditions for BC-1 and BC-2 respectively. Here \( M \) is the sum of masses of the individual monomers \( M = \sum m_i \), \( m_i = 1 \) is the mass of each monomer. The value of \( I_1/I_3 \) is the ratio of major and minor axes and gives a measure of the shape asymmetry of the coil. Comparing the value of \( I_1/I_3 \) in figures 2(a) and (b) we see that \( I_1/I_3 \) has a lower value for all ten RC-2 sets compared to BC-2 set. A plausible explanation for this difference is given later in this paragraph and confirmed by the end of this paper. Subplots figures 2(b) and (d) show the values of \( R_g \) obtained from randomly determined CLs and biologically determined CLs. The calculated value of \( R_g \) for the ring polymer without CLs and the average value is \( \approx 12 \). The value of \( R_g \) decreases as we increase the number of CLs from BC-1/RC-1 set to BC-2/RC-2 sets; this decrease in the value with an increase in the number of effective constraints is expected. But interestingly, the change in the value of \( R_g \) as we go from BC-1 to BC-2 is distinctly less than the decrease in \( R_g \) as we go from RC-1 to RC-2. We interpret the difference between the two cases as follows: the effective CLs in BC-1 are already at critical positions along the contour which give partial organization in the DNA. On increasing the number of CLs (BC-2), the organization of the molecule improves along the already established framework. On the other hand, an increase in the number of random CLs leads to an overall shrinkage in
that for RC-2. Moreover, monomer density of CLs at the central region is more for BC-2 compared to the two cases. In contrast, the normalized density for BC-1 and RC-1, respectively from figures 3 and 4, we again see an increase in the number density of monomers and CLs in set RC-2 as compared to RC-1, whereas plot (b) is for BC-2,RC-2. $n_{CL}^*(r)$ is averaged over nine independent initial conditions and the error bar shows the standard deviation from the average.

The monomer number density of $n_{CL}(r)$ is plotted versus $r$, where $r$ is the distance of the position of the monomers from the center of mass of the DNA-polymer coil. Plot (a) is for BC-1, RC-1, whereas plot (b) is for BC-2,RC-2. $n_{CL}(r)$ is averaged over nine independent initial conditions and the error bar shows the standard deviation from the average.

The quantities $n_{CL}^*(r)$ and $n_{CL}(r)$ are calculated by calculating the average number of monomers and CLs in radial shells of width $2a$ from the CM of the coil, divided by the volume of each shell. The CL-density is further normalized by the total number of CLs for the particular case under consideration to obtain $n_{CL}^*(r)$. Data for $n_{CL}^*(r)$ and $n_{CL}(r)$ from nine independent runs are plotted for each of set of CLs: BC-1, BC-2, and one set of RC-1, RC-2 in figures 3 and 4, respectively. The small standard deviation from the average for monomer number densities and the normalized CL number density is an indication that the arrangement of monomers and CLs have relaxed to similar distributions and is independent of the starting configuration of monomers.

Comparing subplots (a) and (b) of figures 3 and 4 for BC-1 and BC-2 establishes that coils with a higher number of CLs lead to a higher density of monomers and CLs at the center of the coil. As the coil gets into a more compact coil structure lead to a higher density of monomers and CLs at the center of the polymer coil. The quantities $n_{CL}^*(r)$ and $n_{CL}(r)$ are calculated by calculating the average number of monomers and CLs in radial shells of width $2a$ from the CM of the coil, divided by the volume of each shell. The CL-density is further normalized by the total number of CLs for the particular case under consideration to obtain $n_{CL}^*(r)$. Data for $n_{CL}^*(r)$ and $n_{CL}(r)$ from nine independent runs are plotted for each of set of CLs: BC-1, BC-2, and one set of RC-1, RC-2 in figures 3 and 4, respectively. The small standard deviation from the average for monomer number densities and the normalized CL number density is an indication that the arrangement of monomers and CLs have relaxed to similar distributions and is independent of the starting configuration of monomers.

To gain some more insight into the global structural organization of the DNA-coil, the simplest question to ask is whether a particular CL is always found near the center of the coil or near the periphery of the coil. To this end, we compute the probability of each of the CLs to be found in the inner, middle, and outer regions of the DNA-coil. We use the cutoff radii $R_{inner} = 5a$, $R_{middle} = 9a$ (chosen from the knowledge of the value of $R_g \approx 8a$) and calculate the probability $P_{inner}, P_{middle}, P_{outer}$ of finding the $i$-th CL within distance $r < R_{inner}$ (inner region), $R_{inner} < r < R_{middle}$ (middle region) and $r > R_{middle}$ (outer region), respectively, from the coil’s center of mass. If the values of $P_{inner}, P_{middle}, P_{outer}$ for each CL has small deviation from the average value in each of the nine independent runs, it would indicate that the presence of CLs leads to similar organization of the DNA across independent runs. Also, we compare the probability distribution of CLs for runs with bio-CLs and random-CLs to investigate if...
bio-CLs lead to organization distinct from that obtained with random-CLs.

We carry out the same exercise for different segments of the polymer chain. The E. Coli chain with 4642 monomers is divided into 80 segments with 58 monomers in each segment, and the segments are labeled from \( i = 1, 2..., N_s \) as we move along the contour. We can then calculate the location of the CMs of each segment, and find the probability of finding the CMs in the central, middle and outer region. The segments in a random-walk polymer model (without CLs) can take any conformation, and there is no reason to believe that certain segments will preferably be found in the inner or outer regions of the coil. If the segments were completely delocalized, we would expect the polymer in different microstates to contribute to all the P_{inner}, P_{middle}, P_{outer} quantities for each segment. The question is to what extent will this basic behavior of polymer coils be modified by the presence of bio-CLs and random-CLs?

Probability data about the location of CLs and segments for BC-2 and RC-2 is given in figures 5 and 6, respectively. Data for BC-1 and RC-1 is given in the supplementary data (stacks.iop.org/JPhysCM/30/034003/mmedia) section figures 1 and 2. Furthermore, from figures 5 and 6(a)–(c) we see that some CLs (e.g. the CL with index 60) has nearly an equal probability of being in the inner or middle region of the coil, but a very low probability of being found in the outer region. For BC-2, most CLs are found in the inner and middle regions of the coil whereas for RC-2 CL set there are some CLs at the periphery; refer to figure 5. On the other hand, from figure 6 we see a larger number of segments have a finite probability of being in the outer regions for BC-2 as compared to data for RC-2. The data consistently show that the position of CL, as well as segments, are localized in space across different runs.

Having established that the CLs and segments of the DNA-polymer coil have some degree of radial organization, we try to extract more detailed structural information about the position of segments relative to each other within the coil. We calculate the probability of each CL (alternatively, each segment) to be in proximity to other CLs (alternatively, other segments). If there are no particular well-defined relative positions of CLs in each case, further \( n_{PL}^{CL}(r) \) is averaged over nine independent runs starting from initial conditions and the error bar shows the standard deviation from the average.

\[
\begin{align*}
P_{inner} & = \frac{\text{number of CLs in inner region}}{\text{total number of CLs}} \\
P_{middle} & = \frac{\text{number of CLs in middle region}}{\text{total number of CLs}} \\
P_{outer} & = \frac{\text{number of CLs in outer region}}{\text{total number of CLs}}
\end{align*}
\]

where \( r \) is the distance of the position of the CLs from the center of mass of the DNA-polymer coil. Subplot (a) is for BC-1 and RC-1, whereas subplot (b) is for BC-2 and RC-2. The number of CLs in each case is further normalized by the total number of CLs in each case. Further \( n_{PL}^{CL}(r) \) is averaged over nine independent runs starting from initial conditions and the error bar shows the standard deviation from the average.

In figure 7 we show color maps showing the average probability \( P(i,j) \) of finding each pair CLs \( ij \) at distances of \(<5a\) for BC-2, RC-2 for two independent runs. As the Monte Carlo simulation evolves, at each microstate if the distance
average, values of $x$-axis is segment index. In each case $E_\text{Coli}$ DNA coil. The segments to be found in the inner, middle and outer region of probabilities of the center of mass (CM) of 80 DNA-polymer monomers. We set these them by default, and will show up as high probabilities in the a $d <$ 5 between of CLs. A pair of CLs which are near each other along the corresponding data from two independent runs with $RC-2$ set for $BC-2$, and the bottom two color maps in figure 7 show The top two color maps of figure 7) represent data obtained, $i$ for pair $j$ is incremented. The probability $p(i,j)$ is dark in all color maps. We do this because we want to see only non-trivial correlations between different CLs. Following figure 7, the color maps show the probability of finding a pair of segment-CMs within a distance of 5$a$ for $BC-1/RC-1$ and $BC-2/RC-2$ is shown in figures 8 and 9, respectively. Note that these probability color maps give much more detailed information than a pair correlation function $g(r)$, which would just give the average distance between CLs or segment-CMs.

Monomers constituting a pair of nearby CLs are separated by less than 6 monomers along the contour. This is the reason that the diagonal $p(i,i)$ is dark in all color maps. We do this because we want to see only non-trivial correlations between different CLs. Following figure 7, the color maps show the probability of finding a pair of segment-CMs within a distance of 5$a$ for $BC-1/RC-1$ and $BC-2/RC-2$. We arrive at some conclusions by comparing different pairs of probability-color maps in figures 7–9. Firstly, comparing color maps for data from different initial conditions, e.g. compare the top two color maps in each of the figures which are for $BC-1/RC-2$ (or equivalently compare the bottom two color maps which are for $RC-1/RC-2$), shows bright and dark patches at equivalent positions in the map. Thus the same set of CLs and segments are spatially near each other in both the runs, i.e. the polymer organization is similar in both the runs.

Additional color maps from two more independent runs for each set of CLs are also given in the supplementary section for further comparison. The reference to relevant color maps in the supplementary section is given in the figure caption of each figure, and these further reiterate our conclusion that the structural organization of DNA-polymer is similar across different runs for the same set of CLs. Thus we find further evidence of our hypothesis that the set of CLs decides the large scale structure of the polymer.

Secondly, the number of bright pixels are much higher in color maps obtained using CL sets $BC-2$ and $RC-2$ (figure 9) as compared to colormaps for $BC-1$ (figure 8). This is not

**Figure 6.** Subplots (a)–(c) (in the left column) shows the probabilities of the center of mass (CM) of 80 DNA-polymer segments to be found in the inner, middle and outer region of $E.\text{Coli}$ DNA coil. The $x$-axis is segment index. In each case average, values of $P_{\text{inner}}, P_{\text{middle}}$ and $P_{\text{outer}}$ are taken over random runs starting from independent initial conditions; deviation from the average is shown as error bars. Small error bars indicate that the probability of finding CM of segment $i$ in a particular region is nearly the same across different runs. Data on subplots (d)–(f) are for the random choice of cross-link position (set $RC-2$) with 82 CLs in a chain with 4642 monomers. Each segment has 58 monomers. The dataset with fewer CLs (referred as $BC-1$ and $RC-1$, respectively) are shown in supplementary section figure 2.

$d$ between a pair of CLs is such that $d <$ 5$a$, a counter $c(i,j)$ for pair $i, j$ is incremented. The probability $p(i,j)$ at the end of the MC-run is the value of $p(i,j) = c(i,j)/N_{\text{micro}}$, where $N_{\text{micro}}$ is the number of microstates over which data is calculated for calculation. The $x$-axis and the $y$-axis represent CL indices $i, j$, and the colored pixel indicates the value of $p(i,j)$. The top two color maps of figure 7) represent data obtained for $BC-2$, and the bottom two color maps in figure 7 show corresponding data from two independent runs with $RC-2$ set of CLs. A pair of CLs which are near each other along the contour of the chain will have the distance $d <$ 5$a$ between them by default, and will show up as high probabilities in the color map. We set these $p(i,j) = 0$ in the calculation if the
surprising as more constraints due to the presence of higher number of CLs lead to relatively more compact well-defined structure and a large number of CLs (or segments) near one another. With the few bright patches for BC-1, RC-1 CL set with 27 effective CLs, one cannot clearly define the mesoscale conformation of the whole chain, though there are indications of the emergence of structure. However, a set of 82 effective CLs for BC-2, RC-2 might be enough to deduce and define the large-scale organization of DNA-polymer as we now know which segments are neighbors of a particular segment.

Figure 8. Color maps to represent probability \( p(i,j) \) to find CM of segment \( i \) spatially close to CMs of other chain segments \( j \). There are 80 segments in the \( E. \text{Coli} \) polymer with 58 monomers per segment. The top two figures are runs with BC-1 and the bottom two for RC-1. More color maps from independent runs are given in the supplementary section: figure 3.

Figure 9. Color maps to represent the probability \( p(i,j) \) to find CM of segment \( i \) spatially close to CMs of other chain segments \( j \). There are 80 segments in the \( E. \text{Coli} \) polymer with 58 monomers per segment. The top two figures are runs with BC-2 and the bottom two subplots RC-2. More color maps from independent runs are given in the supplementary section: figure 5.
Thirdly, the comparison of color maps for BC-2 and RC-2, especially in figure 9 show a different nature of the organization of the DNA polymer. For BC-2 adjacent segments show higher propensity to be together, which can be deduced by observing that there are clusters of adjacent bright pixels. Comparatively, bright pixels are scattered more randomly in the color maps for RC-2. From the color maps, we can clearly observe that there is a difference in the nature of patterns for BC-2 and RC-2.

Fourthly and importantly, the reasons for the formation of clusters of bright pixels seen in the top two color maps of figure 7 (for CLs) is not the same as that of figure 9 (for segment-CMs). To understand the bright patches of figure 7, we remind the reader that the CLs are often found adjacent to each other along the chain contour for BC-1 and BC-2. Suppose CL-i, CL-j, and CL-k are next to each other along the chain. Note that then p(i, j), p(i, k), p(j, k) has been explicitly put to zero. But if CL-m, which is far from i and j along the contour, comes within a distance of 5a from CL-i, then CL-m is also automatically close to CL-j, k and three adjacent pixels will appear in the color map, viz., p(i, m), p(j, m), p(k, m). Thus, the bigger bright patches for BC-2 in figure 7 should not be interpreted as evidence for a more organized polymer. A similar arrangement of bright/dark pixels across runs is just evidence of similar organization across different runs. We do not get big bright patches in the color maps for RC-1 and RC-2 in figure 7, because the calculations are done using only the effective number of CLs. Thereby there are no redundant CLs in the RC-1, RC-2 set of CLs. We have also checked that the organization is maintained for BC-2 if we have only the effective CLs in the BC-2 set of cross-links.

To quantify the differences in the color maps of BC-2 and RC-2 in figure 9, we calculate the number of segments, \( n_{seg}(i) \), which are near (i.e. within distance \( d < 5a \)) to the CM of the \( i \)th segment with probability \( p(i, j) > 0.05 \). That is, we count the number of non-black pixels in the color map of figure 9 for a particular segment with index \( i \). Then we divide \( n_{seg}(i) \) by the total number of segments to get \( \bar{f}(i) \), to get an estimate of the fraction of a total number of segments which approach segment \( i \) with any finite probability. It is shown in the figure 10 for RC-2 and BC-2. A cutoff of 0.05 for the value of \( p(i, j) \) is appropriate as, nonetheless, most of the color map is black and deep red going up to yellow for very few pixels. From the figure, we observe that the value of \( \bar{f}(i) \) is relatively high for BC-2 set of CLs as compared to \( \bar{f}(i) \) for bio BC-2; this suggests that for random CL-sets, many more segments can approach a particular segment for BC-2 compared to that for BC-2. We interpret this as a more spatially organized structure with BC-2 cross-links, as it has fewer but well-defined neighbors as can also be checked from the color map of figure 9. As an example, segments with indices 70–78 for BC-2 are close only to their adjacent segments (bright diagonal patch in the colormap) giving relatively very low value of \( \bar{f}(i) \) in figure 10(a).

We have also obtained color maps for the ten different sets of random CLs (data not shown), and for each CL-set we can calculate \( \bar{f}(i) \) for each segment index \( i \). Moreover, we can calculate \( f_{av} \), that is the average value of \( \bar{f}(i) \) summed over all the segment indices, i.e. \( f_{av} = (\sum \bar{f}(i))/80 \). Furthermore, we can calculate the mean of \( f_{av} \) over nine independent runs for each CL set, and thereby obtain \( \langle f \rangle \). In figure 10(b), we plot \( \langle f \rangle \) versus the random CL-set index, each set has the same number of CLs as in RC-2. We compare this data with the \( \langle f \rangle \) for the one set of biologically obtained CLs: BC-2. We clearly see that for each random CL-sets the quantity \( \langle f \rangle \) has relatively higher value than \( \langle f \rangle \) for BC-2. Observing the differences in color maps for BC-2 and RC-2, we claim that the positions of CLs along the chain for DNA are not completely random. An equivalent number of CLs in random positions also give an organized structure in that the color maps from nine independent runs look similar, but the nature of the organization is very different from the case where biological positions of CLs are chosen.

To extract further insight into the structural organization of the DNA-polymer, we would next probe whether the segments are at geometrically fixed positions with respect to each other, of course accounting for thermal fluctuations. Thereby, we next calculate the angular correlations between CLs and equivalently between segment’s CMs.

To that end, we calculate the dot product of the radial vectors from the CM of the polymer coil to the respective positions of a pair of CLs \( (i, j) \) and check if the value of \( \cos(\theta_{ij}) > 0 \) or \(< 0 \), where \( \theta_{ij} \) is the angle between the two vectors. If the value of \( \cos(\theta_{ij}) > 0 \), we can say that the two CLs are on the same side/hemisphere of the coil, and increment counter \( c_{opp}(i, j) \) by 1. If \( \cos(\theta_{ij}) < 0 \) we decrement \( c_{opp}(i, j) \) by 1. For all possible pairs of CLs, we calculate the average value of \( \langle c_{opp}(i, j) \rangle \) suitably normalized by the number of snapshots used to calculate the average. The value of \( \langle c_{opp}(i, j) \rangle \approx -1 \) would indicate that the pair of CLs \( i, j \) are always on two opposite hemispheres. A value of \( \langle c_{opp}(i, j) \rangle \approx 1 \) means that the two CLs remain on the same hemisphere. We should not interpret \( \langle c_{opp}(i, j) \rangle \approx 0 \) as we cannot claim that the average angle between the radial vectors is nearly a right angle. The reason is that if the CLs are closer to the center of the DNA-coil, small positional displacements could cause the quantity \( c_{opp}(i, j) \) to fluctuate between 1 and −1 and cause \( \langle c_{opp}(i, j) \rangle \)
to average out to zero. The $\langle c_{\text{opp}}(i,j) \rangle$ data for all pairs of CLs are given in figure 11 for BC-2/RC-2 respectively, the corresponding data for relative angular positions for the segment’s CMs are given in figures 12 and 13 for BC-1/RC-1 and BC-2/RC-2. As before, the top two color maps in all four figures are from two independent initial conditions with BC-1/BC-2 and the bottom two color maps are for two independent runs with RC-1/RC-2.

**Figure 11.** Color maps to investigate the angular location of different CLs with respect to each other. Subplots (a) and (b) are for BC-2 and (c) and (d) for RC-2 with different initial conditions, respectively. Refer to supplementary section figure 6 to compare with more color maps from independent runs.

**Figure 12.** Color maps to investigate the angular location of different DNA-polymer segments with respect to each other. Subplots (a) and (b) are for BC-1 (c) and (d) for RC-1 with different initial conditions, respectively. Refer to supplementary section figure 7 to compare with more color maps from independent runs.

In the color maps of figures 11–13 we see that there are patches of bright and dark pixels, the size of patches are larger for BC-2 compared to RC-2. As mentioned before, if $\langle c_{\text{opp}}(i,j) \rangle = 0$, which is represented by the orange/deep yellow color in the color map, we cannot predict the angular positions of the CLs/segment-CMs because of the reason
explained above. We can clearly see that the color maps from independent runs starting from different initial conditions look similar.

In figure 12, comparing segment-CM color maps in (a) and (b) (for BC-1) with (c) and (d) (for RC-1) we do not find any difference in the nature of the distribution of patches. But as the number of CLs increases as we go from BC-1 to BC-2 and RC-1 to RC-2 in figure 13, we find differences in the pattern of color maps on comparing (a) and (b) with (c) and (d) corresponding to BC-2 and RC-2 CL sets, respectively. In contrast, for color maps (a) and (b) of figure 13, we observe large patches of bright pixels as compared to the patches in (c) and (d). Large patches of bright/dark pixels for BC-2 suggest that adjacent segments along the chain contour are on the same/opposite hemispheres with respect to the CM of the coil. The small patches of bright and dark pixels in (c) and (d) for RC-2 suggest more random distribution of different segments. The polymer is organized in both BC-2 and RC-2 CL sets as color maps from independent runs look similar, but the nature of the organization is different. The reasons for large bright patches in the color maps for CL-angular positions as shown in figure 11 is not the same as for the color maps in figure 13. The reasons for the difference has been explained previously for positional correlation color maps.

Finally, we show a representative snapshot of the DNA-polymer in figure 14 (top). The polymer is colored from blue to red as we go from monomer index 1 to 4642 along the contour. This snapshot confirms what we have deduced from the previous figures of positional and angular correlations. Large sections of the chain are localized together in space. The snapshot confirms the kind of conformations expected from the color maps of angular correlation shown in figures 13(a) and (b). For example, the section marked Region-1 representing monomers around 1750 (segment index 30) is diametrically opposite Region-3 with monomer index 2990 (segment index 50). In figure 13 (a) we see the pixel corresponding to segment indices (30,50) are black. The Region-2 represents monomer numbered around 4100, segment index 71. We can see the pixels corresponding to segment indices (30,71) are yellow whereas pixels for (50,71) are white. The bottom figure shows the CL distribution in space: only one of the monomers out of the pair which constitutes a CL has been plotted.

It is interesting to observe in figure 14 (bottom) that the CLs are clumped together in space in about four aggregates. We believe that this helps in the mesoscale organization of the chain as multiple segments of the chain are pulled towards the coil’s center with multiple loops on the periphery of the coil. The peripheral loops can lead to relatively large fluctuations in the values of $I_1/I_3$ as seen in figure 2. This is further validated by figure 6(c), where we see that a large number of segments are to be found in the outer region with significant probabilities. Thus BC-2 set of cross-links leads to the reorganization of the CLs in space such that they form clusters in space with the possibility of polymer loops emanating from the CL-clusters in a rosette-like structure. We interpret that loops from a particular CL cluster would be neighbors of specific other polymer segments due to the nature of the arrangement, as opposed to spatial proximity to many segments as seen for RC-2 in figure 10 while comparing color maps for BC-2 and RC-2.

We also calculate the distribution of length of segments between two adjacent CLs in the bio and four representative random sets. It is given in figure 15. The distribution of lengths

Figure 13. Color maps to investigate the angular location of different DNA-polymer segments with respect to each other. Subplots (a) and (b) are for BC-2 (c) and (d) for RC-2 with different initial conditions, respectively. Refer to supplementary section figure 8 to compare with more color maps from independent runs.
is a fixed quantity once one chooses a particular CL set as an input to the simulation. The number of segments of length $l$ between two adjacent CLs along the chain contour has been shown on the $x$-axis (using a bin size $d\ell$ of ten monomers) and the frequency density function (FDF) $P(l)$ is plotted on the $y$-axis, where $P(l) = N_S(l)/(N_S * d\ell)$. We denote the number of segments of length $\ell$ by $N_S(l)$ and $N_S$ is the total number of segments between CLs. Thus $N_S = 82 + 2 - 1$ for RC-2 as each CL is constituted of a pair of monomers, and $l$ is essentially a count of the number of monomers between two CL-monomers along the contour. For the bio-CL, one particular monomer is attached with many other monomers (see supplementary material table 1) hence there is a peak at segment length value 0 to 10. We also observe in the randomly chosen CL sets after segment length $\approx 150$ FDF is almost zero while in the biologically obtained CL set there are a few segments until segment length $\approx 400$. This shows that in the biological CL set there are several long segments which can form bigger loops as compared to CLs chosen in a random manner.

Figure 14. Representative snapshot from our simulation of DNA-polymer with BC-2 is shown in the top figure. The color bar on the right shows the color which is used to represent the monomers numbered from 1 to 4642. The black circles show the positions of CLs. The bottom figure shows the position of CLs in space where we have removed the other monomers for better visualization. The coordinates are the same as used for the snapshot above. In bottom snapshot, we see there are approximately four clusters of CLs in space for BC-2 while the CLs are uniformly distributed in space for RC-2 (refer to supplementary section figure 9).

Figure 15. Distribution of segment length between two adjacent CLs for BC-2 set and four different random CL sets corresponding to RC-2. The $x$-axis shows the length of the segments between two CLs and the $y$-axis shows the frequency density of segment’s length.

4. Discussion

The primary and new conclusions of our study is that if particular sets of monomers in a DNA-ring polymer are held together by suitable proteins (cross-links at specific points in our model polymer), it leads to an organization of the polymer coil. A minimal number of CLs are required to be able to claim that there is a distinct organization of the DNA-polymer since we do not obtain a well-defined structure with 47 bio-CLs (equivalently 27 effective CLs). We have also tried to look for organization with the number of CLs in between BC-1 and BC-2 but we cannot identify a well defined structure. The number of effective CLs is 82 for a ring polymer of 4642 monomers, or approximately 2% of the polymer chain. Moreover, the monomers which are cross-linked in the bacterial DNA are not randomly chosen from the length of the contour and lead to an organization of the ring polymer into a particular organization which is very different and distinct compared to what is obtained using an equal number of random cross-links. Of course, the DNA polymer undergoes local conformational fluctuations due to the thermal energy but overall the structure is maintained in a statistical sense. We can deduce the presence of distinctive mesoscale organization of DNA from the calculation of three quantities: (a) radial distribution of segments, (b) positional correlations between segments and (c) angular correlations between segments. Thus we have much more detailed information on the organization of different segments than that can be obtained from pair correlation function. We have used 159 CLs for our simulation of DNA-polymer, but these should be considered as only 82 effective CLs. We can predict the 2D arrangement of different
segments relative to each other with the statistical quantities presented in this manuscript. However, the 2D maps for the DNA of E.Coli (and Caulobacter crescentus) have been presented elsewhere [43]. Note that we have used minimal information from the (experimental) contact map to predict global organization, and we have cross-checked that our prediction matches with the data from the experimental contact map; refer to [43]. We find the clusters of CLs towards the center of the coil; these CLs are pulling different segments of the chain towards the center, and many loops on the periphery, which we interpret as the rosette-like structure. We have given a possible argument for how and why the structure with relatively well localized DNA-polymer segments is achieved in a polymer, but a full understanding and systematic methodology of the choice of CL-positions from the view of polymer physics can be developed only in future, when we will have access to a larger number of contact maps for many DNAs.

To be able to compare between different runs, the contact map is normalized so that the effect of varying the number of sequenced reads is accounted for. Each sum of the number of contacts in each row and column in the matrix was normalized to 1. This provides a normalized contact map, which can now be used to elucidate the 3D structure of the genome and compare changes across different conditions.

The above analysis has been done using raw data obtained from [21].

Acknowledgments

We acknowledge the use of computational facilities provided by DBT Alliance, project numbers IA/I/12/1/500529, IA/I/11/2500290, to C Assisi, S Nadkarni, MS Madhusudhan. We also acknowledge the use of a cluster bought using DST-SERB grant no. EMR/2015/000018 to A Chatterji. We acknowledge financial support by the DBT project BT/PR16542/BID/7/654/2016 to A Chatterji.

Appendix. Generation of contact frequency map

In the field of bioinformatics, a sequence database is a biological database which is a collection of computerized nucleic acid sequences. Paired-end sequencing allows researchers to sequence both ends of a DNA-fragment to generate high-quality, alignable sequence data. Paired-end sequencing facilitates the detection of genomic reorganization and repetitive sequence elements.

In a paired end-sequencing run, the distance between the alignments of the two fragments is the length of the DNA fragment being sequenced. Aligners (software tools) use this information to better align reads when faced with a read that aligns to multiple regions such as those that may lie in a repeat region. To avoid this behavior, the reads are aligned in single end mode while keeping track of the pairs. We have employed the BWA [44] aligner to align reads as it has the best sensitivity among short read aligners.

The aligned reads are then binned at the desired resolution (or the minimum distance between restriction sites). A 2D matrix with the required number of bins is initialized. Large fractions of the reads in a 3C library were from fragments that were not cross-linked and fall into the same bin or bins adjacent to each other. These read pairs were filtered out. The counter in the bin with the coordinates indicated by the alignment of each read in the pair is incremented for all remaining reads. The filled matrix gives the total number of contacts between different parts of the genome and the resulting matrix is called the contact map.
[28] Bohn M and Heermann D W 2010 *PLoS One* **5** 1–14

[29] Barbieri M, Chotalia M, Fraser J, Lavitas L M, Dostie J, Pombo A and Nicodemi M 2012 *Proc. Natl Acad. Sci. USA* **109** 16173

[30] Goloborodko A, Marko J F and Mirny L A 2016 *Biophys. J.* **110** 2162

[31] Rubinstein M and Colby R H 2003 *Polymer Physics* (Oxford: Oxford University Press)

[32] Phillips J T R and Kondev J 2008 *Physical Biology of the Cell* (New York: Taylor and Francis)

[33] Maeshima K, Hihara S and Eltsov M 2010 *Curr. Opin. Cell Biol.* **22** 291

[34] Mateos-Langerak J *et al* 2009 *Proc. Natl Acad. Sci. USA* **106** 3812

[35] Brackley C A, Brown J M, Waithe D, Babbs C, Davies J, Hughes J R, Buckle V J and Marenduzzo D 2016 *Genome Biol.* **17** 59

[36] Rosa A, Becker N B and Everaers R 2010 *Biophys. J.* **98** 2410

[37] Benedetti F, Dorier J and Stasiak A 2014 *Nucl. Acids Res.* **42** 10425

[38] Gilbert N and Marenduzzo D 2017 *Chromosome Res.* **25** 1

[39] Alipour E and Marko J F 2012 *Nucl. Acids Res.* **40** 11202–12

[40] Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie B R, Mirny L A and Dekker J 2013 *Science* **342** 948

[41] Rousseau M, Fraser J, Ferraiuolo M A, Dostie J and Blanchette M 2011 *BMC Bioinform.* **12**

[42] Ba D, Sanyal A, Lajoie B R, Capriotti E, Byron M, Lawrence J B, Dekker J and Marti-Renom M A 2011 *Nat. Struct. Mol. Biol.* **18** 107

[43] Agarwal T, Manjunath G, Habib F and Chatterji A 2017 Origin of spatial organization of DNA-polymer in bacterial chromosomes (arXiv:1701.05770)

[44] Li H and Durbin R 2010 *Bioinformatics* **26** 589