Origin of spatial organization of DNA-polymer in bacterial chromosomes

TEJAL AGARWAL1, G. P. MANJUNATH2, FARHAT HABIB3 and APRATIM CHATTERJI1,4(a)

1 IISER-Pune - 900 NCL Innovation Park, Dr. Homi Bhaba Road, Pune-411008, India
2 IISER Mohali - Knowledge City, Sector 81, SAS Nagar, Manauli-140306, India
3 Inmobi - Cessna Business Park, Outer Ring Road, Bangalore-560103, India
4 Center for Energy Science, IISER-Pune - Dr. Homi Bhaba Road, Pune-411008, India

received 14 September 2017; accepted in final form 20 February 2018
published online 7 March 2018

PACS 87.15.ak – Monte Carlo simulations
PACS 82.35.Lr – Physical properties of polymers
PACS 82.35.Pq – Biopolymers, biopolymerization

Abstract – In vivo DNA organization at large length scales (∼100 nm) is highly debated and polymer models have proved useful to understand the principles of DNA organization. Here, we show that <2% cross-links at specific points in a ring polymer can lead to a distinct spatial organization of the polymer. The specific pairs of cross-linked monomers were extracted from contact maps of bacterial DNA. We are able to predict the structure of 2 DNAs (E. coli and Caulobacter crescentus) using Monte Carlo simulations of the bead-spring polymer with cross-links at these special positions. Simulations with cross-links at random positions along the chain show that the organization of the polymer is different in nature from the previous case. We provide some direct and some indirect experimental validation for our predicted organization of DNA-polymers.

Introduction. – It is established that DNA-polymer is not a random coil in either bacterial cells [1–3] or in eukaryotic cells [4–7]. Experimental methods such as CCC (chromosomal conformation capture) which was then further developed as 5C and then Hi-C have consistently shown the presence of topologically associated domains (TADs) in the contact maps (C-maps) of DNA chains [8–10]. The Hi-C technique gives us the C-map which is the map of frequencies of a segment of the DNA chain (say i) to be found in spatial proximity to another segment (say j) for all combinations i, j of segments along the contour length of the DNA-polymer. TADs are patches in C-maps which indicate that some segments of the chain (at 1 mega-base pair (BP) to 1 kilo-BP resolution), are found spatially close to other particular segments with higher frequencies compared to the rest of the segments.

The ds-DNA is stiff at length scales of 1 nm but can be considered to be a flexible chain at length scales beyond 100 nm [11]. The persistence length \( \ell_p \) of a naked DNA is 150 base pairs (BP) \( \equiv 50 \text{ nm} \) [12] and the value of \( \ell_p \) in vivo is debated [13]. Since the resolution of Hi-C experiments is well above this length scale [1,4], there have been focussed attempts in the last few years trying to understand the DNA organization and in particular the origin of formation of TADs from the principles of polymer physics [14–18]. A series of studies indicates that TADs in eukaryotic cells are indicative of the fractal globule organization of the polymer (as opposed to the equilibrium globule) [4,19]. Recently, more detailed polymer models with either different lengths of loops or with many distinct (coarse-grained) diffusing binder molecules which cross-link different segments of the chain have reproduced TADs of sections of a particular eukaryotic DNA by performing optimizations in the multi-parameter space. Distinct kinds of binder molecules link correspondingly distinct monomers (DNA-segments) along the chain, and the optimization parameters include the number of distinct kind of binders/monomers, the position and number of distinct monomers as well as diffusing cross-links along the contour [16,20,21].

We propose a much simpler model for shorter bacterial DNAs and ask a more general question: Does fixed cross-links (CLs) at a few specific positions along the polymer chain contour organize the polymer into a
particular architecture? If so, can we predict the global shape/structure of the DNA-polymer and does it reproduce the C-map or at least parts of it? The positions of the cross-links are chosen from the C-maps which have relatively higher frequency of interaction between segments. We cross-link a minimal number of these segments, and then computationally cross-check if the other segments of the polymer get localized in space and with respect to each other. Of course the chain can fluctuate due to thermal fluctuations but maintaining the architecture. We then compare this polymer organization with the organization obtained when a ring polymer (most bacterial DNAs are ring polymers) has an equal number of CLs at randomly chosen positions along the chain contour. We choose 10 different realizations of randomly positioned CLs. On comparing we see that nature chooses the position of CLs carefully such that the architecture of the DNA-polymer is well organized in a manner very distinct from what is obtained for a polymer with random CLs.

**Model.** We investigate the organization of two bacterial DNAs, *E. coli* and *C. crescentus*. Each has a single chromosome of length 4 mega-BP: we choose to work with shorter bacterial DNA with just one chromosome and no nuclear wall. DNA is modeled as a flexible bead-spring ring polymer (both bacterial chromosomes are ring polymers) with a harmonic spring potential $\kappa(r-a)^2$ acting between neighbouring beads; the choice of $a = 1$ ($\sim 100$ nm) sets the length scale of the problem. The excluded volume (EV) of the beads are modelled by a suitably truncated purely repulsive Lennard-Jones potential with $\sigma = 0.2a$. The *E. coli* and *C. crescentus* DNAs have 4642 and 4017 kilo-BPs, which we model by 4642 and 4017 monomers, respectively. The naked DNA Kuhn segment has just 300 BPs [12] whereas a bead represents 1000 BPs. The effect of DNA coiling around histone-like proteins occurs at smaller length scales; longer-range effects due to supercoiling, presence of plectonemes etc. should show up in the C-map and their effects get incorporated as cross-links at the length scales we consider. Moreover, bacterial DNA occupy 15–25% of cellular volume [2], so we choose to ignore confinement effects, if any. Instead, we fully focus on the role of CLs in the organization of the polymer. The introduction of CLs between segments of the chain in our model can be justified due the presence of DNA-binding proteins which are present in bacterial cells (as well as higher eukaryotic cells) [2,8,22,23].

We model the cross-links between two segments of the DNA-polymer by a harmonic potential $\kappa(r-a)^2$, where $\kappa = 200k_BT/a^2$ between two monomer beads. The two “cross-linked” monomers (CLs) are typically well separated along the contour of the model polymer. We cross-link a pair of monomers if they are found spatially close above a certain frequency in C-maps. By lowering the frequency cutoff, we can have more cross-linked monomers. For details of the analysis, refer to [24] and to the section “The method to generate C-maps…” in the Supplementary Material Supplementarymaterial.pdf (SM). So, we take 47 or 159 CLs for *E. coli*. For *C. crescentus* we take 49 or 153 CLs which we refer as BC-1 and BC-2, respectively. Cross-linked monomers are listed in table S1 in the SM and [24] (tables are obtained from the analysis of the data in [1,3]). From the table we observe that a pair of neighbouring monomers along the chain contour can get cross-linked to another pair of neighbouring monomers, hence the number of actual CLs are fewer (refer to table S1 in the SM for examples and detailed explanations). Removing such over-counts, there are 26 and 60 effective CLs for *C. crescentus*. For *E. coli* we have 27 and 82 effective CLs. We also investigate the large-scale organization of the chain when we have a set of CLs, where pairs of monomers are chosen randomly and cross-linked. A set of 26 and 60 CLs at random positions in a ring of 4017 monomers is referred as RC-1 and RC-2, respectively.

**Results.** To establish that sets of bio-CLs lead to a particular organization of the polymer, we start from 9 different initial configurations of a ring polymer, and allow the chain to *equilibrate* using Monte Carlo (MC) simulations using the Metropolis algorithm. After equilibration (inherently non-equilibrium biological systems at a certain stage of their cell cycle can be thought of to be in a state of local equilibrium), we compare statistical quantities which provide us evidences about structure and conformations of polymer chains. If we get a similar structure from all 9 runs, we could claim that CLs lead to particular sets of conformation at large length scales. We take care to choose the 9 different initial configurations of polymers in ways that we ensure that the cross-linked monomers are at very different relative positions with respect to each other (see [24] for method details). Moreover, the distance between them can be much larger than their equilibrium distance $a$. Thus the initial potential energy of the system will be high (fig. S3 in the SM), and as the system relaxes due to the presence of CLs through MC moves, the average energy of the polymers in each run should have nearly the same value at the end of the equilibration run. After equilibration, the polymer explores the phase space over 12 million iterations in each run to calculate the statistical quantities with thermal energy scale $k_BT = 1$. Small EV of the beads allows chains to cross each other, moreover we take a large MC displacement of $1.2\sigma$ in every 100 steps. These help in releasing any artificial topological constraints induced by the initial configuration. Chain crossing is justified due to the activity of topoisomerase II.

We repeat these calculations using RC-1 and RC-2 CLs and compare the polymer organization with those obtained using BC-1 and BC-2. To firmly establish that the BC-1 and BC-2 set of CLs lead to an organization of the ring polymer which is very distinct from the organization achieved with random CLs, we choose 10 independent sets of random CLs, then for each set of CLs gave 9 runs starting from 9 independent initial positions. After equilibration, we compare the differences in the large-scale
organization. For each random CL-set, RC-1 CLs are a subset of the RC-2 CL set. We show later that the reason for the distinct organization of the chain with BC-2 is in turn the very distinct spatial organization of the CLs themselves in space, which in turn comes from the choice of monomers which are cross-linked.

The primary problem is how to identify the large-scale organization of a single floppy polymer chain and come up with a prediction of the relative position of different segments, when rapid conformation changes are inherent in the system. Quantities like the pair correlation function $g(r)$ between monomers are insufficient as we would like more individualized information about the arrangement of different segments of the DNA. We use the following four quantities to identify the global organization of C. crescentus DNA-polymer; similar detailed data for E. coli is given in [24].

1) We estimate the radius of gyration $R_g$ of the DNA-polymer of C. crescentus. The value of $R_g$ obtained is $≈8a$ from all the 9 runs for BC-1, and $≈6.5a$ for BC-2. Data is shown in fig. 1(a). In contrast, $R_g$ decreases from $8a$ to $6a$, when the number of CLs are increased from RC-1 to RC-2 for each set of random CLs. The decrease is more significant for random CLs, as would be expected for a polymer chain with more constraints. A smaller relative decrease in $R_g$ as we change from BC-1 to BC-2 compared to the change from RC-1 $→$ BC-2 is the first indication of the distinct organization of the coil with bio-CLs. A ring polymer with 4017 monomers without any CLs has $R_g$ value of 11.

2) We divide the polymer into segments of 50 monomers each, and identify whether the center of mass (CM) of each segment is to be found in the inner, middle or outer section of the coil. Thereby the polymer has 80 segments. We define a segment to be in inner/middle/outer section if the distance $r$ of the segment’s CM from the CM of the coil is ($r < 5a$), ($5a < r < 9a$), ($r > 9a$), respectively. If a segment $i$ is found to be in the same section in all the 9 independent runs, we can claim that all 9 chains are similarly organized. Data in fig. 1(b)–(d), confirms and validates the above claim. As we see in fig. 1(b), the same segments are found inside the coil with higher probability, some are more likely to be found in the middle region, and the rest in the outer region. However, the values of probabilities $P_{inner}, P_{middle}, P_{outer}$ for a segment statistically fluctuate across runs.

3) Instead of calculating $g(r)$, we aim to identify which segments (say $i$) are near other segments (say $j$) with higher probability. We can calculate this probability for each pair of segments and show this in a colormap, where both the axes represent segment indices and the color of each pixel denotes the probability that the CM of segments $i$ and $j$ are within a distance of $RC < 5a$. We show data from two runs for BC-2 in fig. 2(a), (b). For comparison we also show data RC-2 in fig. 2(c), (d), respectively. Colormaps of bio-CLs and random-CLs in fig. 2 show that there is a higher probability (color red) of finding only certain segments near others, and some segments are never found in proximity of certain other segments (color black). This indicates a certain degree of organization of segments. Large fluctuations in the conformation of a polymer would result in a colormap which would be predominantly dark, indicating that there is almost equal (and small) probability of different segments to be near each other (fig. S1 in the SM). Moreover, colormaps from independent runs and same CL set show statistically similar patterns of red and dark pixels: this implies the same organization of segments in independent runs.

For BC-1/RC-1, we cannot identify a large-scale structural organization of the polymer, nor can one distinguish between the colormaps of BC-1 and RC-1. In contrast, on comparing colormaps of BC-2 and RC-2 (fig. 2) we can infer that the character of organization of the polymer is different in the two cases. We find large red patches separated by dark rows/columns in the colormaps for BC-2, whereas, for RC-2 colormaps the lighter pixels are relatively more scattered. This is further quantified in fig. 2(e), where we see that a chain segment is approached by fewer other segments for BC-2 as compared to RC-2. For each of the random CL-sets, each segment can be near a larger number of segments as can be deduced from the higher value of $(f)$ in fig. 2(f). This observation, coupled with the fact that the polymer has a higher value of $R_g$
and more segments in the outer region for BC-2 when compared with data for RC-2 (fig. S6 in the SM), implies that certain segments have well-defined neighbouring segments (more structure) as compared to the polymer with RC-2. The neighbouring segments could be far away along the contour length but are neighbours in spatial location. Thus, the positions of bio-CLs along the contour are special (not random) for BC-2, as these result in a distinctive meso-scale organization of the DNA. Representative snapshots from the simulations of the polymer for different initial conditions are plotted in fig. S6 of the SM from which we see that snapshots from different runs look statistically similar.

We emphasize that the colormaps of fig. 2 look similar to the C-maps of DNA-polymer which we use as modelling input. But the content is very different in the sense that we obtain large-scale structural correlations of the entire polymer chain from our colormaps. To reiterate, C-maps give us input about the location of CLs along the polymer contour at the length scale of monomers, our colormaps show how various segments (each of 50 monomers) are organized relative to each other in space.

4) We next focus on the relative angular position of polymer segments with respect to the CM of the DNA globule. For each pair of segments $i,j$ we calculate if the vectors $\vec{r}_i$ and $\vec{r}_j$, joining the CM of the globule to the CM of the segments $i,j$, subtend an angle of more than $\pi/2$ radians. If the angle $\theta$ between $\vec{r}_i$ and $\vec{r}_j$ is $>\pi/2$, then we interpret that the two segments lie in opposite hemispheres, else the two segments lie on the same side of the globule with respect to the CM of the polymer. We compute the average of the counter $\delta_{ij}$ for each pair of segments $i,j$ as follows: for a microstate $\delta_{ij}$ is incremented by 1 if $\cos(\theta) < 0$, and decremented by 1 if $\cos(\theta) > 0$. We plot the value of $\langle \delta_{ij} \rangle$ in fig. 3 for each pair of $i,j$ for BC-1, BC-2, RC-1, RC-2 CLs. The value of $\langle \delta_{ij} \rangle$ varies from $-1$ to $1$.

The interpretation of data presented in fig. 3 is similar to that of fig. 2. If the pixel corresponding to segments $i,j$ is bright, then they are angularly close with respect to coil CM, and a dark pixel indicates they lie predominantly on opposite hemispheres. An orange pixel corresponds to the value of $\langle \delta_{ij} \rangle$ $\approx 0$. If $\langle \delta_{ij} \rangle \approx 0$ then we cannot interpret their relative angular locations. An orange pixel does not imply $\theta \approx \pi/2$ because one can also get $\langle \delta_{ij} \rangle \approx 0$ if the segments lie close to the center of the coil and can rapidly change their relative positions by small spatial displacements. This would lead to $\langle \delta_{ij} \rangle \approx 0$. For BC-1 and RC-1 CL sets (fig. S2 in the SM), we again see that the patterns
Fig. 4: (Color online) The 2-D maps of DNA organization of *C. crescentus* and *E. coli* bacteria above the 100 nm length scale using BC-2. This is obtained using the statistical data presented in this paper, and in [24] for *E. coli*. The links between different circles indicate that colormaps show these segments to be spatially close to each other though they look separated on the map. The number inside the circle is the segment index, thus we know the relative positions of the different segments with respect to each other. The circles are colored red, yellow or white depending on the number of genes in the corresponding segment (see text).

of bright pixels are almost identical from independent runs. In contrast, the colormaps for BC-2 in fig. 3(a), (b) and RC-2 in fig. 3(c), (d) are immediately distinguishable. The BC-2 data show large patches of bright pixels indicating that adjacent segments along the contour are on the same side of the globule. The dark and bright pixels for RC-2 are relatively more distributed/scattered. Furthermore, a more detailed discussion on the reasons for different colormaps in figs. 2 and 3 for BC-2 is given in the SM.

Finally, using the aggregate of all the structural quantities presented in figs. 1, 2, 3 we are able to piece together the large-scale organization of DNA-polymers in a 2D map for both *C. crescentus* and *E. coli* (fig. 4). For details see the section “Construction of 2-D map” in the SM. Corresponding structural quantities to generate a 2-D map of *E. coli* can be found in [24].

In lieu of the experimental confirmation our predictions, we use the information content of C-maps to validate our prediction of DNA organization. To compare the positional colormaps from our simulations to experimental C-maps, we condense C-map data into a 80 × 80 matrix by a suitable coarse-graining (averaging over neighbouring bins) of the data of C-maps. We then compare the coarse-grained C-map (frequency of larger DNA-sections) with the highest probabilities (>0.5) of the simulation colormaps, see fig. 5: these show a good match. This is by no means an obvious match: we take very few (but significant) points from the C-map as cross-linked monomers in our simulations, and are able to predict from our simulations the positions of highest contact frequencies in the coarse-grained C-map. In addition, we generate additional positional information: see figs. 1, 2, 3. The difference between figs. 5(a) and (c) for *C. crescentus* is the absence of clear high-probability diagonals in our colormap. The plausible reason for this difference could be the presence of plectonemes as proposed in [1]. The effect of plectonemes is not accounted for in our model as we have taken only 153 CLs (only 60 effective CLs). But it is enough to match our simulation C-maps with experimental data even for segments far separated along the contour. Furthermore, the cutoff distance $R_0 = 5a$ we choose while generating colormaps is comparable to the $R_p \approx 4a$ of 50 segments. Thereby, only few segment CMs can occupy positions within distance $R_0$.

The list of the 47 most expressed genes of *E. coli* out of $\approx$4000 genes is given in [25]. Moreover, the position of the gene along the chain contour is also known [26]. We find that 37 out of the 47 most expressed genes are always found in the segments which remain in the peripheral region of the coil as per our simulations. The list of these 47 genes, their locations along the DNA-contour as well as their corresponding monomer and segment index in our model are given in table S2 in the SM. The spatial location of the segment is shown in figs. 1(b)–(d) and in fig. 4. This is an indirect validation of our results, as biologically it is expected that more transcriptionally active genes are found in the peripheral regions for easy access by transcription factors [27]. We did not find the list of most expressed genes for *C. crescentus*.

We can also identify the number of genes in a particular segment of 58 kilo-BP and 50 kilo-BP from [26] for *E. coli* and *C. crescentus*. In fig. 4 the segments are
colored according to the number of genes in that segment; dark color represents more genes and bright color represents fewer genes. For *E. coli* the most gene dense regions are found in the peripheral regions in simulation as well as on the same side of the globule, but for *C. crescentus* the gene dense regions are spread over the extension of the polymer. This seems to be an interesting observation whose physiological relevance might be identified in the future.

**Discussions.** — To summarize, we show that the underlying mechanism of meso-scale organization in 2 bacterial DNAs involves constraints (cross-links) at specific positions along the chain contour. Also we predict the overall 2D architecture of the genomes. We observe that the nature of the organization is different for polymers with the CLs taken from experiments and for polymers with random CL positions. Our preliminary understanding for this difference is that for bio-CLs multiple CLs get clumped together spatially (fig. S6 in the SM). As a consequence multiple segments of the chain are pulled in together by cross-links towards the center of the coil with loops remaining on the outside. This can be reconfirmed also for *E. coli* in [24]. In contrast, for a RC-2 set of CLs the CLs are scattered in space. We have also shown that a minimal number (around ~3% of monomers) of CLs are required for a polymer to get organized in a particular structure as we do not get any organization in the case of BC-1 and RC-1. We have found a lack of clear polymer organization with the number of CLs in between BC-1 and BC-2. The release of topological constraints through chain crossing could play a crucial role in the above organization. Figure 5 and table S2 in the SM provide direct and indirect validations of our predicted organization of *E. coli* and *C. crescentus* genomes as shown in fig. 4. We do not aim to exactly reproduce the DNA contact maps of the bacteria as more detailed models [14–16] do with multi-parameter optimizations. Instead we show that a bead-spring polymer model (with small EV to enable chain-crossing) produces a well-defined organization of the polymer with the hypothesis of having a minimal number of CLs at specific biologically determined positions.

***

We acknowledge computational facilities provided by DBT Alliance, project Nos. IA/I/12/1/500529, IA/I/11/2500290 to C. Assisi, S. Nadkarni, M. S. Madhusudhan and computer cluster from DST-SERB Grant No. EMR/2015/000018 and funding from DBT Grant BT/PR16542/BID/7/654/2016 to A. Chatterjee. We acknowledge useful discussions with M. S. Madhusudhan.

**REFERENCES**

[1] Le T. B. K., Imakaev M. V., Mirny L. A. and Laub M. T., *Science*, 342 (2013) 731.

[2] Jouve X., *J. Phys.: Condens. Matter*, 27 (2015) 383001.

[3] Caglierio C., Grand R. S., Jones M. B., Jin D. J. and O’Sullivan J. M., *Nucleic Acids Res.*, 41 (2013) 6058.

[4] Lieberman-Aiden E., van Berkum N. L., Williams I., Imakaev M., Ragan-McFarland E., Telling A., Amit I., Laio A., Sabo P. J., Dorschner M. O., Sandstrom R., Bernstein B., Bender M. A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L. A., Lander E. S. and Dekker J., *Science*, 326 (2009) 289.

[5] Bickmore W. A. and van Steensel B., *Cell*, 152 (2013) 1270.

[6] Sexton T., Yaffe E., Jenneberg E., Bantignies F., Leblanc B., Hoichman M., Tanay A. and Cavalli G., *Cell*, 148 (2012) 458.

[7] Tjong H., Gong K., Chen L. and Alber F., *Genome Res.*, 22 (2012) 1295.

[8] Dixon J. R., Selvaraj S., Yue F., Kim A., Li M. H. Y., Shen Y., Liu J. S. and Ren B., *Nature*, 485 (2012) 376.

[9] Halverson J. A., Smrek J., Remer K. and Grosberg A., *Rep. Prog. Phys.*, 77 (2014) 022601.

[10] Dekker J., Marti-Renom M. A. and Mirny L. A., *Nat. Rev. Genet.*, 14 (2013) 390.

[11] Phillips J. E. and Renkawitz R., *Polymer Physics* (Garland Science) 2008.

[12] Rubinstein M. and Colby R. H., *Physical Biology of the Cell* (Garland Science) 2008.

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

[14] Pombo A. and Nicodemi M., *Curr. Opin. Cell Biol.*, 28 (2014).

[15] Gilbert N. and Marenduzzo Davide, *Chromosome Res.*, 25 (2017) 1.

[16] Fudenberg G., Imakaev M., Lu C., Goloborodko A., Abdennur N. and Mirny L. A., *Cell Rep.*, 15 (2016) 2038.

[17] Mirny L. A., *Chromosome Res.*, 19 (2011) 37.

[18] Rosa A. and Everaers R., *PLOS Comput. Biol.*, 4 (2008) e1000153.

[19] Imakaev M. V., Tchourine K. M., Nechaev S. K. and Mirny L. A., *Soft Matter*, 11 (2015) 665.

[20] Barheri M., Chotalia M., Fraser J., Lavitas L.-M., Dostie J., Pombo A. and Nicodemi M., *Proc. Natl. Acad. Sci. U.S.A.*, 109 (2012) 16173.

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

[22] Phillips J. E. and Corces V. G., *Cell*, 137 (2009) 1194.

[23] Ohielson R., Bartkuhn M. and Renkawitz R., *Chromosoma*, 119 (2010) 351.

[24] Agarwal T., Manjunath G., Abid F., Vaddavalli P. L. and Chatterji A., *J. Phys.: Condens. Matter*, 30 (2018) 034003.

[25] Karlin S., Mizrak J., Campbell A. and Kaiser D., *J. Bacteriol.*, 183 (2001) 5025.

[26] Online available database https://www.ncbi.nlm.nih.gov.