Heterogeneous interactions and polymer entropy decide organization and dynamics of chromatin domains

Kiran Kumari,[1,2,3] J. Ravi Prakash,[3] and Ranjith Padinhateeri[2]

[1]*IITB-Monash Research Academy, Indian Institute of Technology Bombay, Mumbai, Maharashtra - 400076, India*

[2]*Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai 400076, India*

[3]*Department of Chemical Engineering, Monash University, Melbourne, VIC 3800, Australia*

Abstract

Chromatin is known to be organized into multiple domains of varying sizes and compaction. While these domains are often imagined as static structures, they are highly dynamic and show cell-to-cell variability. Since processes such as gene regulation and DNA replication occur in the context of these domains, it is important to understand their organization, fluctuation and dynamics. To simulate chromatin domains, one requires knowledge of interaction strengths among chromatin segments. Here, we derive interaction strength parameters from experimentally known contact maps and use them to predict chromatin organization and dynamics. Taking two domains on the human chromosome as examples, we investigate its 3D organization, size/shape fluctuations, and dynamics of different segments within a domain, accounting for hydrodynamic effects. Considering different cell types, we quantify changes in interaction strengths and chromatin shape fluctuations in different epigenetic states. Perturbing the interaction strengths systematically, we further investigate how epigenetic-like changes can alter the spatio-temporal nature of the domains. Our results show that heterogeneous weak interactions are crucial in determining the organization of the domains. Computing effective stiffness and relaxation times, we investigate how perturbations in interactions affect the solid-like and liquid-like nature of chromatin domains. Quantifying dynamics of chromatin segments within a domain, we show how the competition between polymer entropy and interaction energy influence the timescales of loop formation and maintenance of stable loops.

PACS numbers:

INTRODUCTION

Chromatin is a long polymer made of DNA and proteins; its four-dimensional (spatial and temporal) organization is crucial in regulating cellular processes like transcription, replication and DNA repair [1–4]. How the chromatin self-organises spatially and temporally is a question of active research. Recent developments in chromosome conformation capture (Hi-C) [4–9] and microscopy experiments [10–12] help us investigate contacts among different segments and overall 3D organization of the chromatin polymer. Such experiments so far suggest that intra-chromatin interactions lead to the formation of different types of domains within a single chromosome [6, 7, 11–15].

It has been hypothesized that loop extrusion, phase separation or an interplay between both could be the mechanism for the formation of these domains [14, 16–18]. In the phase separation picture, interactions mediated by the different combinations of histone modifications and chromatin binding proteins lead to the formation of compartments/domains [17, 19, 20]. In the loop extrusion picture, chromatin regions are actively brought together with the help of proteins like cohesins/condensins and held together by CCCTC-binding factor (CTCF) [13, 14, 21, 22].

While an interplay between loop extrusion and phase separation would determine the formation of many of the domains, it has been found that not all the domains have CTCF-dependent loops [13, 23]. Recently, it has been shown that in the absence of loop extruding factors, chromatin does still form domains and execute the necessary biological function [11, 17, 23–25] indicating that micro phase-separation might be an important mechanism. Phase separation is also known to bring together certain enhancers and promoters, segregating them from other regions [19, 20, 26]. In certain cases, as far as biological function is concerned, there is an ongoing debate whether actual contact is crucial or proximity — closeness in 3D without being physically in contact — would suffice [3, 17, 24, 27]. To probe this further, it is important to go beyond contacts and examine the whole configurational space of chromatin polymer.

Widely used experimental methods so far provide us only static snapshots of chromatin contacts. Large Hi-C contact values are sometimes assumed to represent static contacts holding together different segments of chromatin. However, note that experimentally inferred contact probability values for nearly all chromatin segments are very small ($<< 1$) [13]. This

implies that the contacts will be often broken, and the chromatin polymer conformation can be highly dynamic. Even in a steady state, there are likely to be large fluctuations, cell to cell and temporal variabilities. While there have been several attempts to understand the static 3D configurations and their average properties, the dynamic nature of chromatin domains remains unclear.

Complementing recent experimental efforts, there have been several computational studies investigating chromatin organization [21, 22, 28–52]. Many studies carried out "forward" simulations where the chromatin structure is predicted starting with a set of interaction strengths [40, 41, 43, 44]. However, we do not know the optimal values of interaction strengths *a priori*. One requires an inverse method that can obtain the optimal interaction strengths such that experimentally known properties are recovered. Only a few studies has been carried out to optimize the interaction energy parameters to converge to the experimental contact [31, 48] We have recently developed an Inverse Brownian Dynamics (IBD) method to extract potential energy parameters that are consistent with HiC/5C experiments [30]. The method explicitly computes intra-chromatin interactions for every segment pair. Using the intra-chromatin interactions computed from this method, in this work, we go beyond the static picture and study chromatin fluctuations and dynamics perturbing epigenetic states. We use Brownian dynamics (BD) simulations with hydrodynamic interactions (HI) for this study. Going beyond contact frequencies, we compute the full distance distribution between all pairs of segments and examine the cooperative nature of folding. We then study the dynamics of the domain, compute relaxation times, loop formation times and contact times. We investigate how the interaction strengths and polymer entropy influence these measurable quantities. Finally, we discuss the significance of these findings.

MODEL AND METHODS

We consider chromatin as a bead spring chain having optimal intra-chromatin interactions derived from 5C and Hi-C data using an Inverse Brownian Dynamics (IBD) algorithm [30]. The total energy of the chromatin bead spring chain, made of N beads, is $U = U^{\mathrm{S}} + U^{\mathrm{SDK}}$ where $U^{\mathrm{S}} = \sum_i \frac{H}{2}(|\boldsymbol{r}_i - \boldsymbol{r}_{i+1}| - r_0)^2$ is the spring potential between the adjacent beads i and $(i+1)$, $\boldsymbol{r}_i$ is the position vector of bead i, r_0 is the natural length and H is the stiffness of the spring [53]. To mimic protein-mediated interactions between bead-pairs,

we use the Soddemann-Duenweg-Kremer (U^{SDK}) potential whose repulsive part is modelled by the Weeks-Chandler-Anderson potential $4\left[\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}+\frac{1}{4}\right]-\epsilon_{ij}$ upto the range $2^{1/6}\sigma$. Here σ is the parameter that determines the minima of the potential, $r_{ij}=|\boldsymbol{r}_i-\boldsymbol{r}_j|$ is the distance and ϵ_{ij} is an independent parameter representing the attractive interaction strength between beads i and j. The repulsive part is unaffected by the choice of potential parameter ϵ_{ij}. The attractive part is modelled by $\frac{1}{2}\epsilon_{ij}\left[\cos\left(\alpha r_{ij}^2+\beta\right)-1\right]$ which, unlike the Lennard-Jones potential, smoothly reaches zero at the cut off radius $r_c=1.82\sigma$ [54, 55]. Parameters α and β control r_c (see Table S1 in supplementary information (SI)). We simulate the chromatin polymer using Brownian dynamics simulations, where the time evolution of the bead positions are governed by an Itô stochastic differential equation. This simulation accounts for hydrodynamics interaction which is computed by the regularized Rotne-Prager-Yamakawa tensor. (See sec. S1) For the simulation, all the length and time scales are non-dimensionalised with $l_H=\sqrt{k_{\mathrm{B}}T/H}$ and $\lambda_H=\zeta/4H$, respectively where T is the absolute temperature, k_{B} is the Boltzmann constant, and $\zeta=6\pi\eta_s a$ is the Stokes friction coefficient of a spherical bead of radius a with η_s being the solvent viscosity. All non-dimensional quantities are indicated with the asterik (*) symbol such as the non-dimensional position $r_i^*=r_i/l_H$.

For chromatin, we do not know the interaction energy strengths *à priori*. Given 5C/HiC data, we performed an inverse calculation and obtained the optimal interaction strengths (ϵ_{ij}) between bead i and bead j, that are consistent with the experimental contact probability map [30]. The inverse algorithm works as follows: we start with the initial guess values of interaction strengths, simulate the polymer following the conventional forward Brownian dynamics method and obtain the simulated contact probabilities in the steady-state (IBD flow chart is shown in Fig. S1). Based on a statistical method, the interaction strengths (ϵ_{ij}) are revised for the next iteration, depending upon the difference between the simulated and the known experimental contact probabilities (see sec. S1). We perform several iterations of the loop (i.e., BD simulation, calculation of contact probability and revision of ϵ_{ij}) until the error between the simulated and experimental contact probabilities is less than a predetermined tolerance value. At the end of the IBD process, we obtain the optimised interaction strength values (ϵ_{ij}) that are independent of the initial guess values (also see Kumari et al. [30]). For each iterative step, we have generated 1000 trajectories and collected 100 samples from each trajectory, making it an ensemble of 10^5 configurations.

In this manuscript, we have studied two different loci in the human genome. We primarily focus on human α-globin gene locus (Chr:5, ENm008 region); we have simulated it starting with the experimental data available from Baù et al. [9], in two different cell types: in the GM12878 cell type where the α-globin gene is in its repressed state (throughout the manuscript we will refer to this state as the "OFF" state), and in the K562 cell type where the gene is in its active state (throughout the manuscript we will refer this state as the "ON" state) The optimised ϵ_{ij} values for these states are shown in Fig. S2 (also see Kumari et al. [30]). We have repeated this study on a different locus (human Chr.7, a 500 kb region) for two cell types: IMR90 and K562. The different cell types for the same region represent different epigenetic states of the domain.

RESULTS AND DISCUSSION

Prediction of interaction energies between chromatin segments

As a first step, we have analysed the optimised interaction strengths (ϵ_{ij}), predicted using our IBD method, for α-globin gene locus for two different cell types (OFF and ON epigenetic states). The frequency distribution of the optimised ϵ_{ij} values is shown in Fig. 1(a). In the ON state, the smaller values of ϵ_{ij} are more frequent, while the OFF state has a wide distribution of ϵ_{ij}.

We then investigated the relation between the optimised ϵ_{ij} values (from IBD) and the contact probability values (p_{ij}) from experiments. We find that, for any given bead pair, contact probability value alone does not determine ϵ_{ij}. Studying the data in detail, we propose the following empirical relation that can predict the interaction strength of a given bead pair:

$$\epsilon_{ij} = \epsilon_0 + \alpha \log\left(\frac{p_{ij}}{p_{min}(i-j)}\right), \tag{1}$$

where $p_{min}(i-j)$ is the minimum non-zero value of the contact probability of all beads having genomic distance $|i-j|$, and the other two parameters $\alpha \approx 1$, $\epsilon_0 \approx 0$ are numerical constants (see S2). In Fig. 1(b) & (c), we present the comparison between the ϵ_{ij} predicted from the above empirical relation and the optimised ϵ_{ij} from IBD for ON and OFF states respectively. One can see that the formula predicts the ϵ_{ij} values reasonably well. The ϵ_{ij} from IBD simulation has larger spread; however, both have a similar behavior and very close

numerical values. See Fig. S3 and sec. S2 for more details. This formula would greatly help in generating excellent initial guess values of ϵ_{ij} and will lead to faster convergence of the IBD.

Next, we computed the average 3D spatial distance ($r^*_{ij} = |\mathbf{r}^*_i - \mathbf{r}^*_j|$) between different pairs of chromatin segments as a function of the corresponding 1D genomic separation ($s_{ij} = |i-j|$) for α-globin gene (OFF state), and have compared with a "control" simulation of a self-avoiding walk (SAW) polymer with no attractive interaction ($\epsilon_{ij} = 0$). While the SAW results in $\nu = 0.6$ as expected, the OFF state has a scaling exponent $\nu = 0.38$ suggesting a near closed packing (see Fig. S4 and sec. S3). 3D distance computations help us to fix parameters in the model. We used published FISH data for distance between a single pair of α globin segments and deduced $l_H = 36$nm, and using the estimates of diffusion time scales we deduced $\lambda_H = 0.1$ (see Fig. S5 and sec. S4). Throughout this paper, $l_H = 36$nm and $\lambda_H = 0.1$s are used to convert all non-dimensional lengths and times, respectively, into standard units. We will present quantities in both non-dimensional and dimensional units. The reasons for the choice of both these specific values are discussed in greater detail in sec. S3 and S4

Distance distributions and cooperative nature of chromatin folding

Even though the average distance between two chromatin segments is often used to represent chromatin organization, this may not describe the accurate biological picture in a dynamic, heterogeneous context. We compute the distribution of the 3D distance between different segments, $p(r^*)$, as it captures the maximum information about variability and fluctuations of chromatin. As a control, we computed the $p(r^*)$ for a SAW and it agrees well with the known analytical expression of des Cloizeaux, $p(\boldsymbol{r}^*) = C(r^*)^{\theta+2}e^{-(Kr^*)^{1/(1-\nu)}}$ [56] . Here ν is the Flory exponent, θ is a geometrical exponent and the coefficients C and K are functions of θ [56]. Recent work has led to an accurate estimation for these constants which are discussed in sec. S5. Fig. 1(d) shows the validation with intermediate beads of a SAW (with $\nu = 0.6, \theta = 0.81, K = 1.17, C = 2.05$; also see Fig. S6).

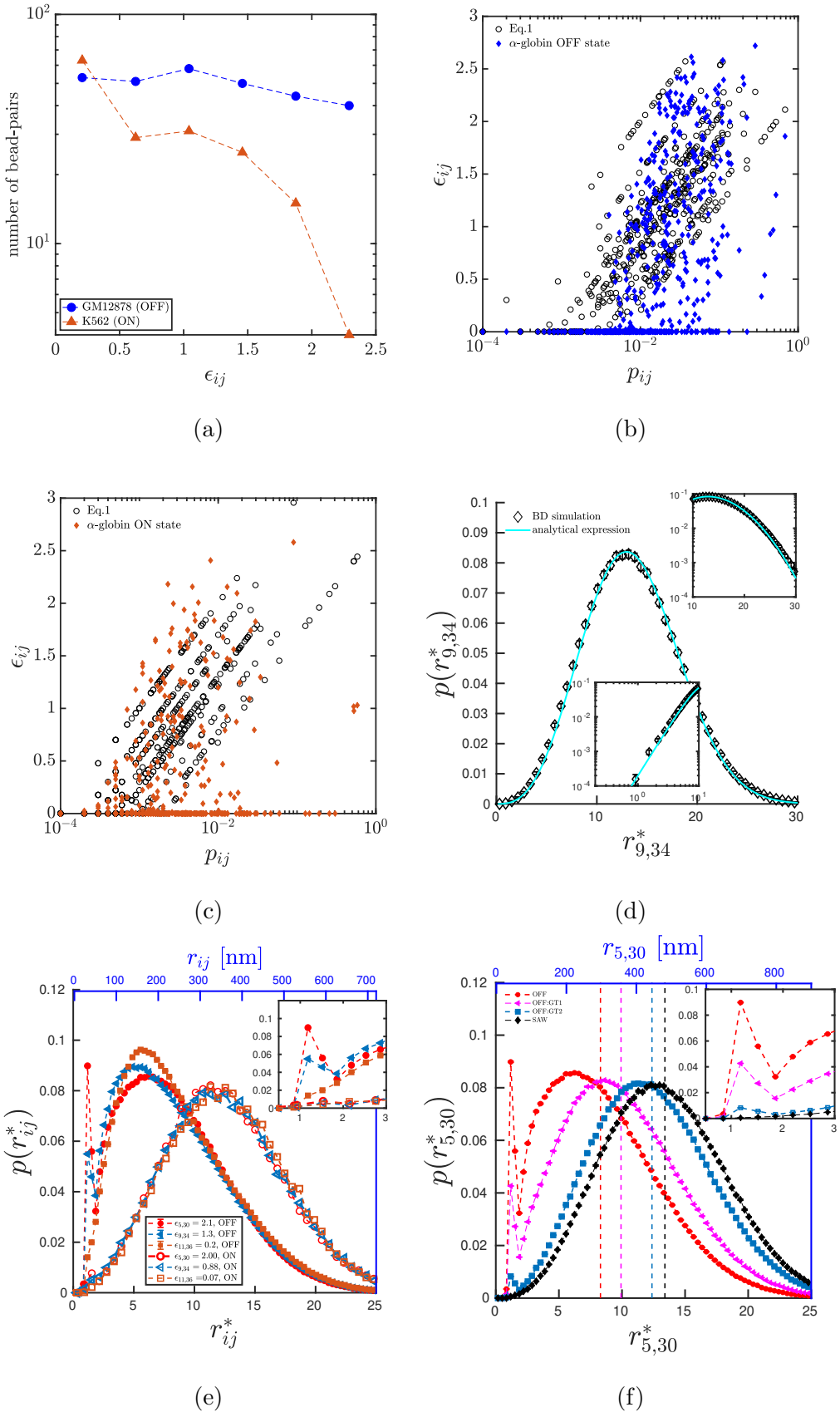

number of bead-pairs
GM12878 (OFF)
K562 (ON)
ϵ_{ij}
(a)
Eq.1
α-globin OFF state
ϵ_{ij}
p_{ij}
(b)
Eq.1
α-globin ON state
ϵ_{ij}
p_{ij}
(c)
BD simulation
analytical expression
$p(r^*_{9,34})$
$r^*_{9,34}$
(d)
r_{ij} [nm]
$p(r^*_{ij})$
r^*_{ij}
(e)
$r_{5,30}$ [nm]
OFF
OFF-GT1
OFF-GT2
SAW
$p(r^*_{5,30})$
$r^*_{5,30}$
(f)

FIG. 1

FIG. 1: **Distribution of interaction strengths and 3D distances**: (a) Distributions of the optimised interaction strengths (ϵ_{ij}) for the ON and OFF states of the α-globin gene locus. The ON state is populated with smaller values of ϵ_{ij} while the OFF state has nearly a flat distribution with similar occurrence of all values of ϵ_{ij}. (b-c)Prediction of the relation between interaction strength and contact probability from the IBD simulation(filled diamonds) and from the eq. 1(black open circles) for the OFF state(b), and ON state(c) of $\alpha-$globin. (d) Distance probability distribution $p(r^*_{ij})$ from simulations compared with the analytical des Cloizeaux expression [56] for a pair of beads 9 and 34 of a SAW polymer of total length $N = 50$. Insets show the zoomed in view of the fit for small (bottom inset) and large (top inset) values of r^*. (e) $p(r^*_{ij})$ for various bead pairs with different interaction strengths ϵ_{ij}, but same $s_{ij} = 25$ in the the OFF state (filled symbols) and the ON state (empty symbols). Inset: the small r^* regions is zoomed in the inset to highlight the interaction-driven peak in the OFF state. (f) Comparison of $p(r^*_{ij})$ between different perturbed epigenetic-like states. OFF: state with all interactions in GM12878 (red), OFF GT1: when weak interactions are ignored; i.e., only with $\epsilon_{ij} > 1k_{\mathrm{B}}T$ interactions (pink), OFF GT2: when only very high interactions ($\epsilon_{ij} > 2k_{\mathrm{B}}T$) are accounted for (blue), SAW: control simulation with no attractive interaction (black). Vertical dashed lines represent $\langle r^*_{ij} \rangle$ corresponding to each distribution.

We then studied the $p(r^*)$ for the α-globin gene locus in both GM12878 and K562 cell types. Examining various segments 250kb apart along the chain backbone (25 beads) for the OFF state, we find that all the distributions have a broad peak near their respective average distances (Fig. 1(e)), with an overall asymmetric tilt towards smaller r^* values, when compared with SAW (See some sample snapshots in Fig.2). For bead pairs having high ϵ values, a sharp peak emerges near $r^* \approx r_c$ — we call this an "attraction-driven peak" as it is within the attractive range of the potential. The height of the peak is correlated with the strength of attraction (ϵ). However, the average distances (vertical lines in Fig. 1(d)) appear independent of ϵ_{ij}. This difference is also reflected in the cumulative distribution function as shown in Fig. S7. Similarly, we computed $p(r^*)$ for the ON state as well. Interestingly, we do not see the attraction-driven peak near $r^* \approx r_c$. The shape of the curve is similar to the SAW showing more symmetric distribution and independent of the identity of the bead-

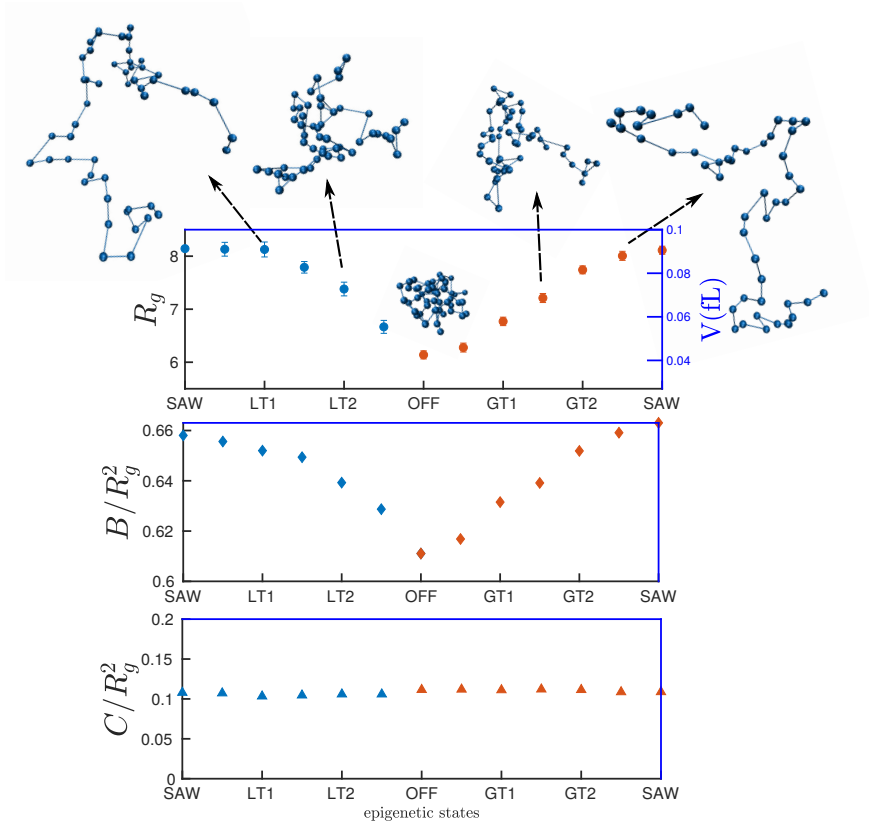


FIG. 2: **Interaction strengths determining shape properties of chromatin domain:** Upper panel: Size of the chromatin domain (R_g) as we perturb interaction strengths. x-axis represents different interaction states with extreme ends representing the control (SAW) polymer and the OFF state in the middle. LT1 (LTx) indicates that all interactions below $1k_BT$ (xk_BT) are present in the polymer. Similarly GT1 (GTx) indicates that all interactions above $1k_BT$ (xk_BT) are present. The right y-axis indicates volume of the chromatin domain in femtolitre. Snapshots from simulations at various epigenetic states are shown around the perimeter of the graph. Bottom two panels represent the normalized asphericity (B/R_g^2) and acylindricity (C/R_g^2), respectively. See sec. S7 for the definition of B and C. The x-axis is the same in all panels.

pairs – it depends mainly on the genomic separation. Together, these results imply that average distances between bead pairs may not represent the complete picture of chromatin organisation; understanding the whole distribution is necessary.

Given that we have the optimal interaction strengths that satisfy the experimentally known contact probability constraints [30], we can answer the following important question:

Are the measurable properties of a given bead-pair (e.g. $r_{5,30}$) solely determined by the interaction between those two particular beads ($\epsilon_{5,30}$) or are they influenced by the interactions among other bead-pairs as well? To answer this, we adopted the following strategy: we systematically switched off the attractive interaction among certain bead pairs and computed probability distributions and other polymer properties. We simulated polymers for the following four cases: (i) all interactions are considered – GM12878 (OFF), (ii) only those interactions above $1k_BT$ are considered – we call it OFF:GT1 – all weak interactions ($< 1k_BT$) are switched off here, (iii) only strong interactions above $2k_BT$ are considered – OFF:GT2 – all weak and medium interactions ($< 2k_BT$) are switched off (iv) all attractive interactions are switched off – the SAW polymer. These four cases can be imagined as four different epigenetic-like states – states having different interaction strengths due to underlying epigenetic variations (see sec. S6 and Fig. S8). Fig. 1(f) shows $p(r_{5,30})$ for all the four cases. When we switch off the weak interactions below $1k_BT$ (OFF:GT1), compared to the OFF state, the height of the interaction-driven peak of the distribution decreases and overall, the polymer swells resulting in the shift of the second peak (compare pink and red curves in Fig. 1(f)). This implies that weak interactions having strengths comparable to thermal fluctuations can also influence the contact probability and polymer configurations. If we keep only the highly prominent interactions and neglect all interactions below $2k_BT$ (OFF:GT2), the interaction-driven peak further diminishes, and the distribution function approaches the SAW distribution (compare blue with other curves in Fig. 1(f)). Note that the interaction between beads 5 and 30 is present ($\epsilon_{5,30} = 2.09$) in all the cases except in the SAW case. These results suggest that the measurable properties for a given bead-pair (e.g. $r_{5,30}$) depends not only on the attraction strength of that particular bead pair but also on the interactions of the whole polymer chain. This result implies that all bead pairs collectively/cooperatively contribute in determining the relative position for a particular bead pair.

Relevance of weak and strong interactions in altering the volume and shape of the chromatin domains

The above picture suggests that the chromatin folding is influenced by the collective behaviour of all beads having different interaction strengths. To examine the nature of

collective behaviour, we probed a property of the whole polymer, namely the radius of gyration (R_g) defined in sec. S7. To understand how folding is affected by different epigenetic states, we did the following. We started with a polymer having no interactions (SAW), added weak interactions (small ϵ) that exist between beads in the OFF state as the first step, equilibrated, computed R_g and sequentially added stronger interactions between beads step by step ($\epsilon < 0.5, \epsilon < 1.0, ..., \epsilon < 2$ and so on, denoted as LT1, LT2 etc.), until the OFF state (GM12878) is reached. Each step was equilibrated and R_g was computed (see Fig:2, top panel). From R_g, we have also computed the volume $V = (4/3)\pi R_g^3$ of the chromatin domain as shown in the right side y-axis. As seen from the figure, adding very weak interactions does not change R_g much. However, adding intermediate interactions significantly reduces the R_g, and it saturates as the interactions get stronger, resulting in a sigmoidal-like curve showing signatures of cooperative/collective behaviour. Since we have equilibrated the polymer for each set of ϵ values, the LHS of the curve can be interpreted in two ways: folding the polymer by adding stronger and stronger interaction starting with a completely unfolded state or equivalently unfolding the polymer by removing the stronger interaction starting with a completely folded OFF state. One can also ask how the polymer would fold if one adds strong interactions (larger ϵ) as the first step, starting with SAW, and then add weaker interaction sequentially step by step (denoted as GT1, GT2 etc.). This is shown in the RHS of Fig. 2 (orange symbols). The whole curve suggests that having prominent interactions alone or weaker interactions alone may not take the system closer to its full equilibrium state. We also show typical snapshots of 3D chromatin configurations corresponding to different epigenetic states. As expected, the OFF state is compact, and the volume of the domain increases as we go towards the SAW state. The approximate two fold-change in volume between the two extreme states seen here are roughly the same order as the density change observed experimentally [57].

To quantify how the shape of the chromatin domain changes with epigenetic states, we computed the asphericity (B) and the acylindricity (C) parameters (see sec. S7 for definition). Asphericity quantifies the extent of deviation from a spherical shape. If a polymer is coiled with the average shape of a sphere, $B = 0$. Here a positive B value suggests that even in the OFF state, the chromatin domain is not a perfect sphere. As we go from OFF to SAW, the asphericity increases by ≈65% as shown in sec. S7. However, the asphericity scaled with the polymer size (B/R_g^2) changes by ≈10%. Similar to R_g, we have shown the GT (RHS, orange

symbols) and LT (LHS, blue symbols) cases for the asphericity too. Even though both sides are monotonically increasing, note that LT cases are not equivalent to the GT cases. We also compute the acylindricity parameter that quantifies the extent of the deviation from a perfect cylinder. Here too, $C > 0$ values suggest that the chromatin domain is not a perfect cylinder (see the lower panel of Fig. 2 and sec. S7). Even though the acylindricity is monotonically increasing (shown in Fig. S9) from the OFF state to a SAW, it is increasing in proportion to the size of the polymer. Hence the scaled acylindricity (C/R_g^2) is nearly a constant, as shown in the lower panel of Fig. 2.

Weak interactions are crucial in recovering 3D organization of chromatin

To demonstrate our method and results on a different chromatin domain, we studied another 500 kbp domain (137,250 kbp-137,750 kbp) on human Chr7. Starting from Hi-C data (10kb resolution) from Rao et al. [13], we performed IBD and computed optimal interaction strengths (see details in sec. S8) for cell types describing different epigenetic states (cell lines IMR90 and K562) for the same location. As the Hi-C data is denser compared to the 5C, we converged for the Hi-C contacts in a step-wise manner. In the first step (shown in Fig. 3 for IMR90), we optimized only for the stronger contacts (prominent peaks) in the Hi-C data. The prominent (strong) contacts are defined as follows: for each value of the $|i-j|$ line (line parallel to the diagonal representing all equidistant bead pairs), the mean and standard deviation for that off-diagonal line is computed as $\bar{p}_{|i-j|}$ and $\sigma_{|i-j|}$, respectively. If the contact probability for a matrix element in that $|i-j|$ off-diagonal line is greater than this mean + standard deviation ($p_{ij} > \bar{p}_{|i-j|} + \sigma_{|i-j|}$), it is considered as a prominent probability value. In the next step, we added weaker contacts such that all contacts greater than the mean ($p_{ij} > \bar{p}_{|i-j|}$) are present and achieved the IBD convergence. Finally, we included the rest of the contacts (all contacts) and obtained the optimal interaction strengths that can reproduce the complete experimental HiC map (Fig. 3). It is interesting to note that in the first and second steps when weak contacts (contacts smaller than the mean) are not present, the contact probability from our simulations did not recover the complete contact map. This step-wise optimization process illustrates that the weak contacts are equally important, and the prominent contacts alone cannot reproduce the complete Hi-C map. This is consistent with what we found for the α-globin domain in the earlier section. A similar analysis for

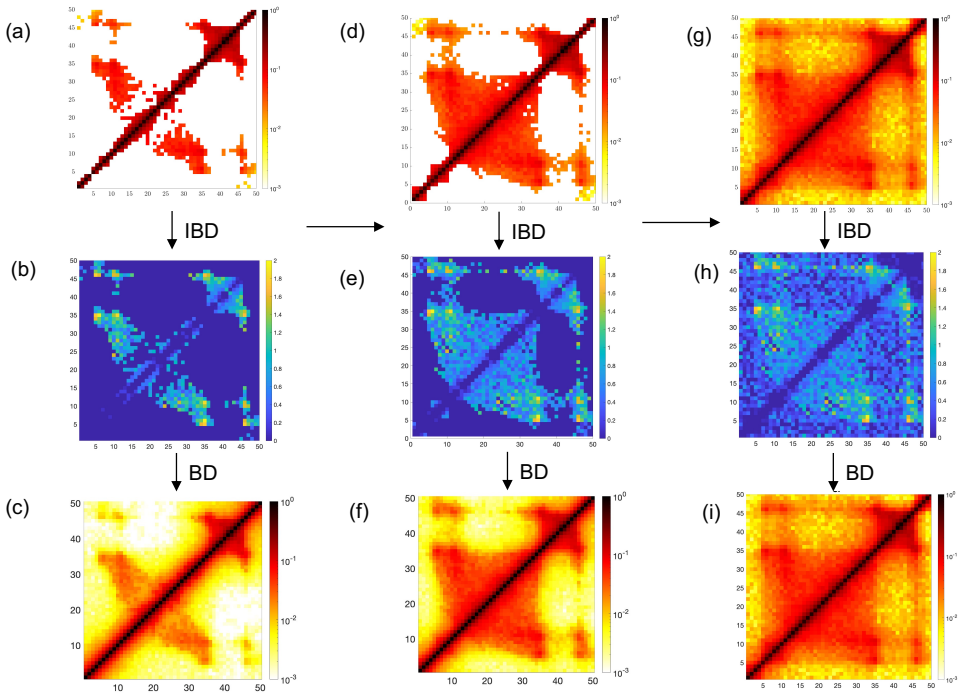
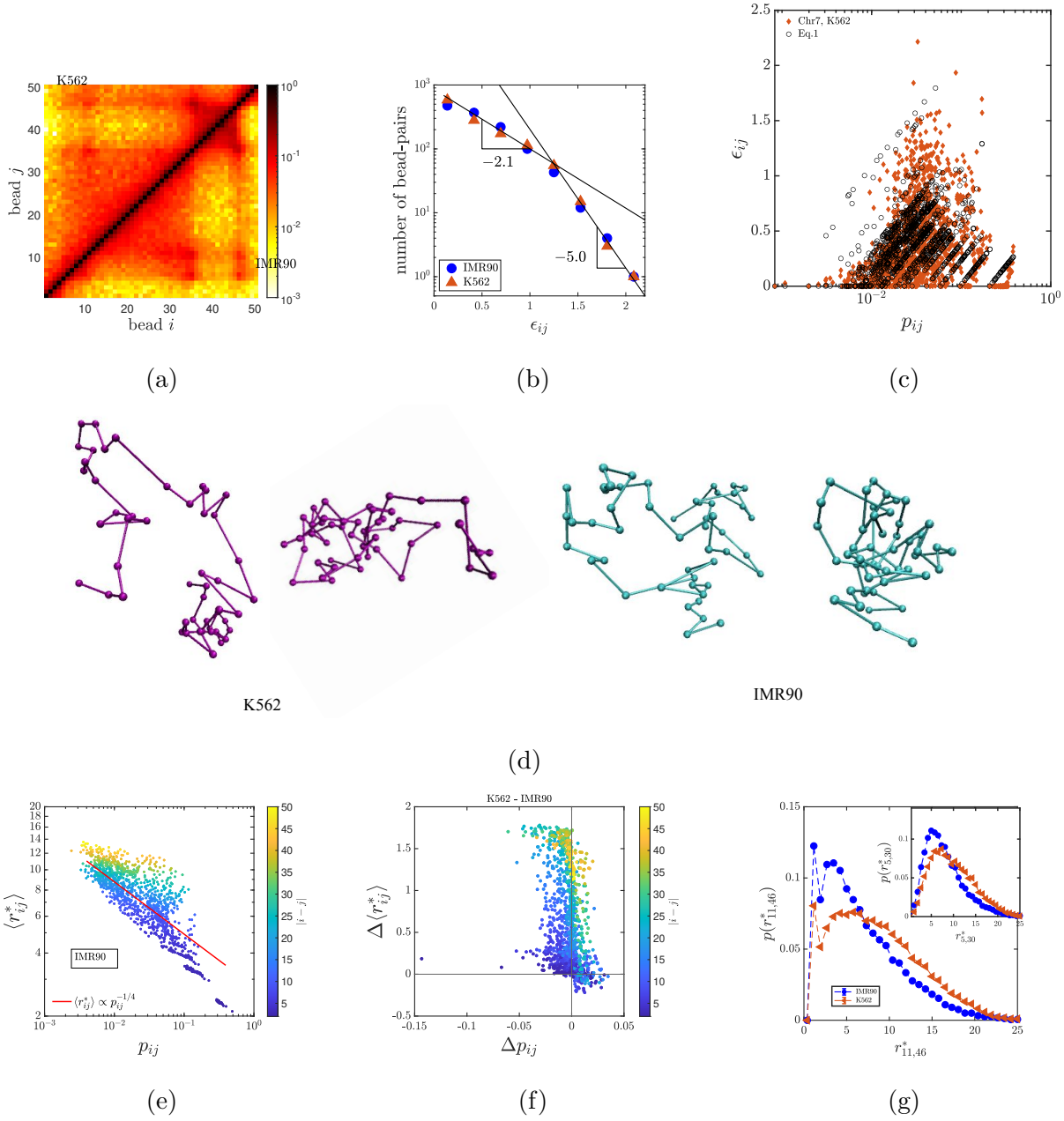


FIG. 3: **Step-wise IBD showing the importance of weak interactions:** Stepwise IBD process for the Chr7 region in the IMR90 cell line. The IBD optimisation is presented in 3 steps. (I) First column: input is only the prominent probability values (higher than one standard deviation from the mean. i.e., $p_{ij} > \bar{p}_{|i-j|} + \sigma_{|i-j|}$) show in (a) based on the peak detection algorithm (see text and S8). (b) The corresponding optimised ϵ_{ij} values, and (c) the recovered contact probabilities. (II) Second column: after the first optimisation step, all peaks above the average probability values (shown in (d)) are fed as input (i.e., $p_{ij} > \bar{p}_{|i-j|}$). The corresponding optimised interaction strengths in (e) are improved with some weak interactions appearing in this 2nd step. However note that the recovered contact probability simulated with prominent interactions in (f) is not comparable to the full contact probability in (g). (III) Third column: the complete Hi-C matrix in (g) is fed as input. At the end of this third step the whole contact probability matrix was recovered in (i), and the corresponding optimal interaction strengths are predicted in (h).

K562 is shown in Fig. S10, and a comparison between ϵ_{ij} for both the cell lines are shown in Fig. S11.



(a) (b) (c)

(d)

(e) (f) (g)

FIG. 4

FIG. 4: **Chromatin configuration and fluctuations for different epigenetic states**: (a) The recovered contact probabilities (p_{ij}) for two epigenetic states/different cell types, K562 (upper triangle) and IMR90 (lower triangle) for a 500 kbp domain (137,250 kbp-137,750 kbp) on human Chr7. Contact maps for both the cell lines are roughly similar. However, the difference can be seen around certain bead pair like 10-45. (b) The distributions of the optimised interaction strengths (ϵ_{ij}) for the corresponding epigenetic states (cell types). Both the distributions are peaked at low values of ϵ_{ij} and are sharply decaying. Two exponential lines are given as a guide to the eye. (c) Prediction of the relation between interaction strength and contact probability from the IBD simulation(filled diamonds) and eq. 1(black open circles) for the Chr.7 region from K562 cell line (d) Two randomly selected 3D configurations for each epigenetic state (K562 and IMR90). (e) Our prediction of the mean 3D distance between every pair of beads (r_{ij}) as a function of their corresponding contact probability (p_{ij}) for one of the epigenetic states(IMR90). The color indicates genomic distance $|i-j|$ between the pair of beads (in units of bead size, see sidebar). Curve with a power-law relation $r_{ij} \propto p_{ij}^{(-1/4)}$ is plotted (red line) as a guide to the eye. (f) The difference between K562 and IMR90 epigenetic states. $\Delta r_{ij} = r_{ij}(\text{K562}) - r_{ij}(\text{IMR90})$ is predominantly positive suggesting that K562 is slightly more extended (see text). (g) Probability distribution of 3D distance, $p(r^*_{ij})$, for bead pairs $11-36$ (main figure) and $5-30$ (inset). The former pair has high contact probability and shows a double peak. The latter (inset) is for a randomly selected pair. Both the main figure and inset suggest that K562 is slightly more extended than IMR90.

Distribution of interaction strengths, 3D organization and fluctuations of chromatin domain for different epigenetic states

Now that we have used the IBD method and obtained optimized ϵ_{ij} for the new domain, we can analyze the statistical properties of ϵ_{ij}, use ϵ_{ij} values in a forward simulation, and obtain 3D organization of the domain for both the epigenetic states (IMR90 and K562). First, we present contact probability recovered from our simulation(Fig. 4(a)) for both the epigenetic states (upper triangle is K562 and the lower triangle is IMR90). Since the IBD

algorithm iteratively reduces the difference between the input HiC contact probability (experimentally obtained) and the recovered contact probability, this map from the simulation is identical to the contact map from Hi-C up to a very low tolerance (see Method). Note that contact maps for both the cell types are roughly similar. However, there are some important differences for some of the contact pairs (e.g., 10-45). We aim to examine whether these small differences in contact probability would lead to detectable differences in 3D distance and other quantities. In Fig. 4(b), we present the distribution of optimized interaction strength for both the cell lines (IMR90 and K562). All the distributions are peaked at smaller values and decrease gradually, reminiscent of exponential decay. The interaction strength per bead-pair is in the range 0-2 k_{B}T. Here too we used eq.1 and predicted the ϵ_{ij} values. As seen from Fig. 4(c), the prediction provides a reasonable comparison with what is obtained from the IBD simulation, indicating that the eq.1 works well for different cell types and different experimental datasets (more details in sec. S2).

We now present representative 3D configurations for both the cell lines in Fig. 4(d). To quantify the shape and size of this region, we computed R_g, asphericity(B) and acylindicity(C). As represented in Table I, R_g shows that K562 is more extended compared to IMR90. In other words, the epigenetic changes, appearing as small differences in contact maps and ϵ_{ij}, are affecting the 3D configuration of chromatin in these cell types differently. The parameters B and C suggest that IMR90 is more spherical and K562 is more cylindrical compared to each other.

Our simulations can predict the precise relation between contact probability of any segment pair and its corresponding average 3D distance. Fig. 4(e) shows this relation for the domain we studied from the cell line IMR90 (see Fig. S12 for K562). Interestingly, it shows a broad distribution of 3D distance for a given contact probability, similar to what is observed in α-globin gene locus [30]. Even though a single power law curve cannot describe the whole data, the fit to the broad data gives the relation $r_{ij} \sim p_{ij}^{-1/4}$. While the average behaviour of the curve is consistent with the earlier work [58, 59], our simulations uniquely predicts the variability in the 3D distances. In Fig. 4(f), we show how the contact probabilities ($\Delta p_{ij} = p_{ij}(\mathrm{K562}) - p_{ij}(\mathrm{IMR90})$) and corresponding 3D distances ($\Delta r_{ij} = r_{ij}(\mathrm{K562}) - r_{ij}(\mathrm{IMR90})$) differ between both the cell types. The p_{ij} differences are small. However, a small change in Δp_{ij} (≈ 0.05) can change the 3D distance by $\approx$ 2units ($\approx$ 60nm) for certain pair of beads (beads having high $|i-j|$, green/yellow color). Intu-

TABLE I: Various shape properties based on the eigenvalues of the gyration tensor G are described here for cell lines IMR90 and K562 for the domain we studied in chromosome 7. See sec. S7 for definition of these symbols.

Shape property	cell line IMR90	cell line K562
R_g^2	30.25	40.20
B	17.50	25.67
B/R_g^2	0.54	0.64
C	3.52	4.20
C/R_g^2	0.116	0.104
prolateness	-0.970	-0.974
anisotropy	0.478	0.538

itively, when Δp_{ij} is positive, the Δr_{ij} should be negative and vice versa. This is clearly seen in the second and fourth quadrants. The first quadrant, where both Δr_{ij} and Δp_{ij} are positive points to those bead pairs that are compacted/extended due to possible collective effects arising from other beads nearby.

Since a wide range of 3D distances are possible for any given contact probability, we computed the whole 3D distance distribution $p(r^*)$ between different segments for both the cell type (see Fig. 4(g)). For certain segment pairs, here too, we observe double-peaked distributions, showing attraction-driven peak as in the case of the α globin OFF state. For certain other segment pairs, the distribution has only a single peak, as shown in the inset of Fig. 4(g). However, in all the cases, the distributions are skewed towards the lower values of r_{ij} suggesting more compaction compared to SAW. It is also evident that the K562 is more extended than IMR90 once again, showing that small changes in contact probability can have interesting differences in 3D distances.

Prediction of probability of triple contacts

Experimental contact probability data such as 5C and Hi-C can only provide pair-wise contacts. However, our simulations can go beyond pair-wise interaction and can provide multi-body contacts. Here in this section, we investigate three-body contacts — simultane-

ous contacts between three beads in our simulations — for various segments in different cell types. That is, we compute the probability of three beads (at positions r_i, r_j and r_k) coming together within a distance less than the cutoff radius (i.e. $r_{ij} < r_c$, $r_{ik} < r_c$, and $r_{jk} < r_c$). First, we investigate the triple contacts within the α-globin domain. Figs. 5(a), (b) and (c) show the triple body contact with bead number 10, 15 and bead 40 as reference points, respectively. That is, the probability that bead 10 is in contact with any other two beads (i and j represented in x and y axis, respectively) is shown in Fig. 5(a) for both the ON (K562) and OFF (GM12878) states in lower and upper triangles of the matrices, respectively. Since, bead 10 is the reference point ($k = 10$), the vertical lines at beads 10 ($i = 10$) indicates the triple contact among $k = 10$ (the reference point), $i = 10$, and j. Since the reference and i are the same, this reduces to a pair-wise contact. Similarly, the horizontal line at bead 10 indicates the contact among $k = 10$ (the reference point), i, and $j = 10$. All other matrix elements show the triple contact that bead 10 makes with different pairs of beads (i and j). Interestingly, the OFF (GM12878) state has many more triple contacts as compared to the ON (K562) state. This is consistent with our earlier observation that the OFF state is highly compact while the ON state is close to SAW polymer.

We repeated this analysis for different beads in the domain of our interest in Chr7. Fig. 5(d)-(f) show the triple contact with beads 11, 35 and 46 as reference points, respectively. Here too, we represent the K562 and IMR90 in upper and lower triangles of the matrices, respectively. In all three cases, IMR90 shows more triple contacts compared to the K562 cell line, specifically in the region bead 30 to bead 40.

Estimation of stiffness and drag properties from domain relaxation times and fluctuations

Whether chromatin is liquid-like, solid-like or gel-like has been a matter of intense discussion in the recent literature [10, 25, 60, 61]. In the phase separation picture, chromatin segments are thought to be "liquid-like", dynamically exploring various configurations. Given that our model can study the stochastic nature of formation and breakage of bonds, and polymer dynamics, consistent with what is observed in Hi-C experiments, below we compute relaxation times and fluctuations of the chromatin domain and estimate effective elastic and drag properties. To demonstrate our results, we use α-globin domain as an example.

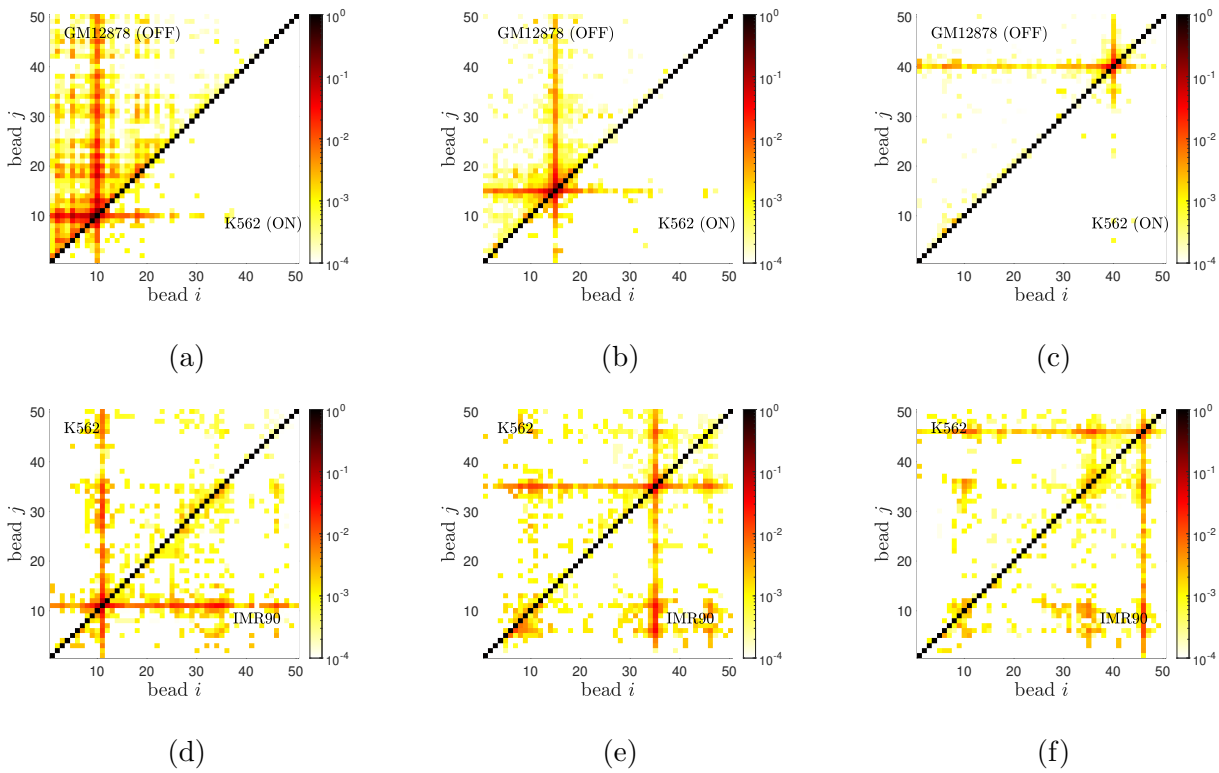


FIG. 5: **Triple contact probabilities**: Probability of contact between 3 chromatin segments simultaneously, computed from simulations. (a) Triple contact probabilities of bead 10 (reference point) with any other two beads for the α-globin gene locus in the OFF state (upper triangle) and ON state (lower triangle). (b) & (c) are similar to (a) with reference points as beads 15 and bead 40, respectively. (d) Triple contact probabilities of bead 11 (reference point) with any other two beads on the Chr.7 region for K562 (upper triangle) and IMR90 (lower triangle). (e) & (f) are similar to (d) with reference points as beads 35 and bead 46, respectively.

First, we computed the end-to-end vector autocorrelation function $\langle \mathbf{R}^*_{\mathrm{E}}(0) \cdot \mathbf{R}^*_{\mathrm{E}}(t^*) \rangle \,/\, \langle \mathbf{R}^{*2}_{\mathrm{E}}(0) \rangle$ where $\mathbf{R}^*_{\mathrm{E}} = (\mathbf{r}^*_1 - \mathbf{r}^*_{50})$ and extracted the longest relaxation time τ^* with and without hydrodynamic interactions (HI). The autocorrelation decay computed with HI is shown in Fig. 6(a) and no-HI case is shown in Fig. S13. Fig. 6(b) shows that the relaxation times for all the perturbed states are lower with HI, as observed previously for the protein folding simulations [62, 63]. We also computed the R_g autocorrelation function and observed similar behaviour. All results presented in this paper

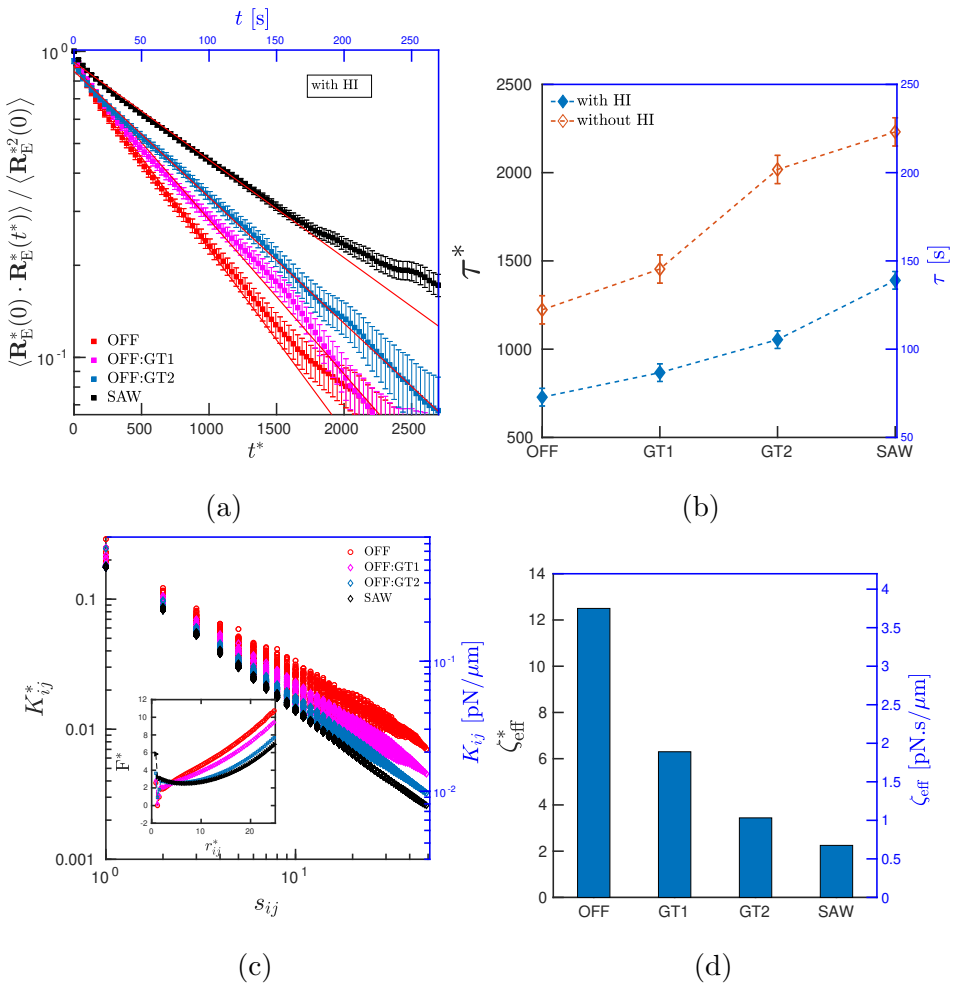


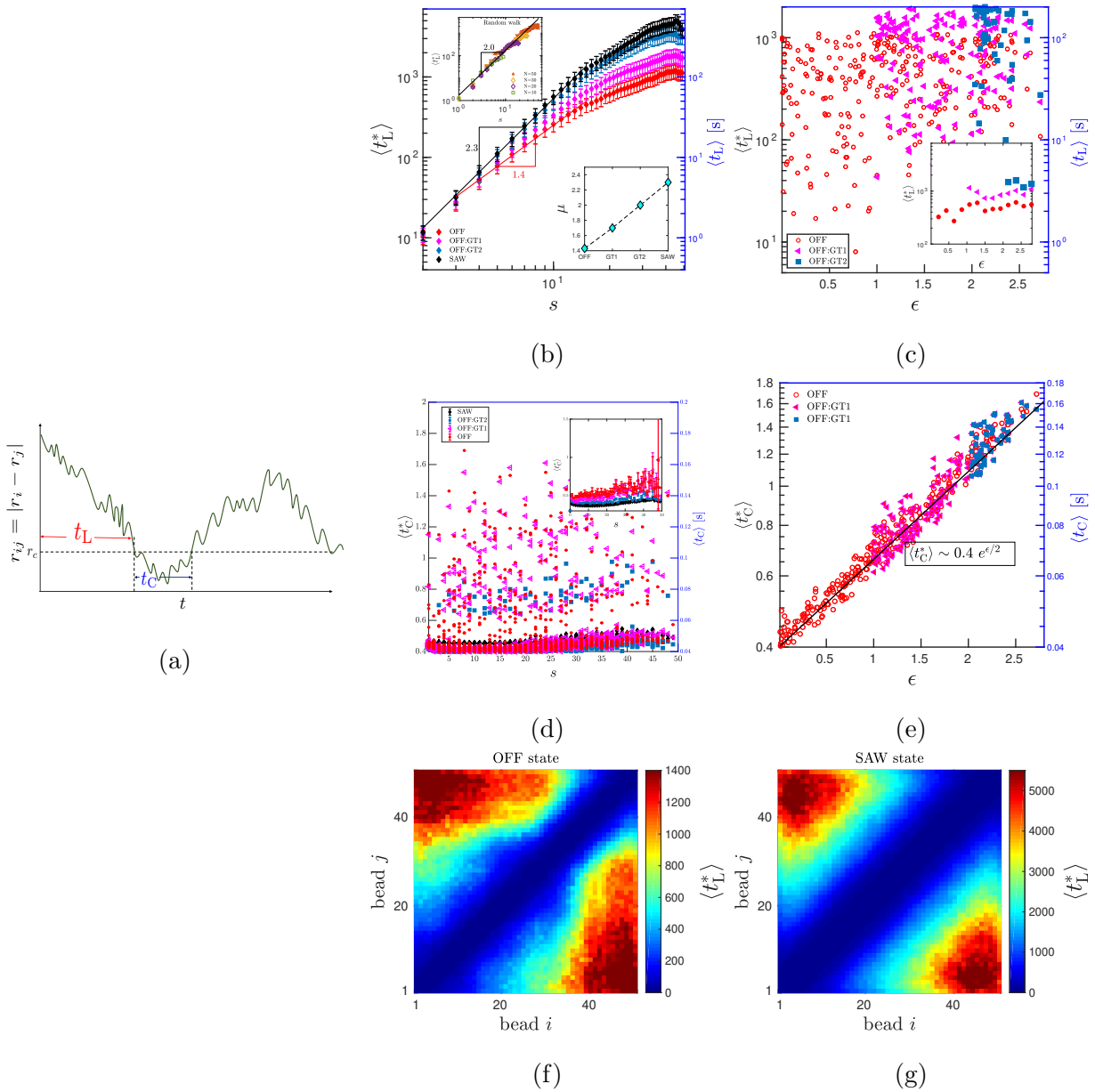
FIG. 6: **Stiffness and relaxation time characterising domain properties:** (a) Exponential decay of end-to-end auto-correlation function with time for four epigenetic states computed with HI. (b) Relaxation times (τ) with and without HI reveal that HI helps in relaxing the polymer faster. (c) Effective stiffness between all the bead-pairs; OFF state is more stiff compared to less interacting states and SAW. Inset: Free energy as a function of bead pair distance $r_{5,30}$. (d) Effective viscous drag felt by different chromatin states. OFF state chromatin (with stronger interactions) is more stiff and has a higher viscous drag.

are computed with HI unless stated otherwise. The chromatin in the OFF state has a lower relaxation time compared to SAW chromatin (See Sec. S9 for more details). Experiments measuring mean square displacement (MSD) have reported that repressed chromatin regions have an unexpected higher diffusivity compared to the non-repressed chromatin [10, 64];

considering that diffusivity is inversely proportional to the relaxation time, there is some similarity between the experimentally reported results and our findings. To understand this behaviour, we investigate the elastic and drag properties of the chromatin domain. From the measurement of fluctuations of each bead-pair we can compute an effective stiffness defined as $K_{ij} = k_{\mathrm{B}}T/\langle|\mathbf{r}_i - \mathbf{r}_j|^{\mathbf{2}}\rangle$ (Fig. 6(c)). As expected, the highly interacting OFF state is stiffer than the other epigenetic states, including SAW. This can also be understood from the free energy as a function of bead pair distance $F^*(r^*) = -\ln(p(r^*)/4\pi r^{*2})$ (see inset). The above behaviour is consistent with $K^*_{ij} \sim \dfrac{\partial^2 F^*_{ij}}{\partial r^2_{ij}}$ and stiffness (K^*_{ij}) of different epigenetic states do show similar behaviour. Since timescales in such problems are inversely proportional to the stiffness, the observed lesser time is explained by the higher stiffness. For the known stiffness and relaxation times, we can compute an effective drag coefficient defined as $\zeta^*_{\mathrm{eff}} = \tau^* \times K^*$. Taking the effective stiffness of the end beads ($K^*_{1,50}$), we find that the drag for the OFF state is higher than the other states suggesting that the existence of larger attractive interactions reduces its ability to reorganize. Both the stiffness and drag are greater for the OFF state than the SAW, but they combine to lead to a faster relaxation time for the OFF state. Our findings are consistent with the recent experimental report that highly interacting chromatin shows reduced mobility as measured by Fluorescence Recovery After Photobleaching (FRAP) technique, revealing the gel-like nature of chromatin [25, 61].

Interplay between interaction energy and polymer entropy influences the dynamics of chromatin domain

While we have gained insights into steady-state fluctuations and distance distributions, how the interactions would affect chromatin dynamics can be further probed [44, 65, 66]. We know that contacts between chromatin segments are dynamic; proteins that form contacts bind and dissociate, resulting in stochastic formation and breakage of contacts. This opens up interesting questions: How long do two beads remain in contact (looped)? When loops break and beads diffuse away, how long does it take for the bead pairs to come back in contact? What are the factors (interaction strengths, polymer entropy etc.) dictating the phenomena of dynamic contacts?



(a)
(b)
(c)
(d)
(e)
(f)
(g)
OFF state
SAW state
bead i
bead j
$\langle t_L^* \rangle$
$\langle t_C^* \rangle$
$\langle t_C^* \rangle \sim 0.4\, e^{\epsilon/2}$
OFF
OFF:GT1
OFF:GT2
SAW
Random walk
$r_{ij} = |r_i - r_j|$
t_L
t_C
r_c
t
s
ϵ
μ

FIG. 7

FIG. 7: **Interaction energy and genomic seperation dictate the behavior of loop formation time $\langle t_{\mathrm{L}}^{*} \rangle$ and contact time $\langle t_{\mathrm{C}}^{*} \rangle$:** (a) schematic representation of the distance between two beads in a single trajectory showing t_{L}^{*} and t_{c}^{*}. (b) $\langle t_{\mathrm{L}}^{*} \rangle$ has a power law scaling with genomic length ($\langle t_{\mathrm{L}}^{*} \rangle \sim s^{\mu}$) with exponent varying from 1.4 (OFF state) to 2.3 (SAW) for different chromatin states. The exponents are shown in lower inset. The upper inset shows $\langle t_{\mathrm{L}}^{*} \rangle$ for various N values for a random walk polymer. (c) $\langle t_{\mathrm{L}}^{*} \rangle$ as a function of interaction strength with each point representing a bead pair. Note the huge spread in $\langle t_{\mathrm{L}}^{*} \rangle$. Inset: $\langle t_{\mathrm{L}}^{*} \rangle$ binned and averaged over all bead pairs having same ϵ showing minimal influence of ϵ. (d) $\langle t_{\mathrm{C}}^{*} \rangle$ as a function of s with each point representing a bead pair. Here too, note the spread. Inset: $\langle t_{\mathrm{C}}^{*} \rangle$ binned and averaged over all bead pairs having the same s showing minimal dependence on the segment length. (e) $\langle t_{\mathrm{C}}^{*} \rangle$ increases exponentially with the interaction strength. $\langle t_{\mathrm{L}}^{*} \rangle$ for all the bead pairs as a heatmap for (f) OFF and (g) SAW states.

To study the temporal nature of chromatin, similar to Ghosh and Jost [44], we define loop formation time (t_{L}^{*}) and contact time (t_{C}^{*}) for all bead-pairs. t_{L}^{*} is defined as the time taken for a pair of beads to meet ($r_{ij}^{*} < r_{\mathrm{C}}^{*}$) for the first time, starting from a random equilibrium configuration. t_{C}^{*} is defined as the duration that the bead-pairs remain looped/in contact. A schematic representation of a typical time trajectory of 3D distance indicating t_{L}^{*} and t_{C}^{*} is shown in Fig. 7(a) and the actual data from our simulation, as an example, is shown in Fig. S14. Corresponding average quantities are defined by $\langle t_{\mathrm{L}}^{*} \rangle$ and $\langle t_{\mathrm{C}}^{*} \rangle$, respectively.

Two possible factors that can influence these temporal quantities are interaction strengths (ϵ) and polymer entropy. Since two beads having a larger segment length between them will have higher entropy, it is expected that the time to come into contact is longer. In other words, the time of looping is expected to be dictated by polymer entropy. To validate this hypothesis, we looked at $\langle t_{\mathrm{L}}^{*} \rangle$ as a function of the genomic length with and without HI.

As shown in Fig. 7(b) $\langle t_{\mathrm{L}}^{*} \rangle$ monotonically increases with s showing a power law behavior $\langle t_{\mathrm{L}}^{*} \rangle \sim s^{\mu}$. As a control, we matched our $\langle t_{\mathrm{L}}^{*} \rangle$ results with the previously known exponents $\mu \approx 2.3$ for SAW and $\mu = 2.0$ for a random polymer (see top inset) [67]. By simulating various chain lengths ($N = 10, 20, ...$), we can infer that the deviation from the power law for large s is due to finite chain effects (top inset). We have also computed $\langle t_{\mathrm{L}}^{*} \rangle$ for all

the other epigenetic states revealing $2.3 > \mu \geq 1.4$. The OFF state having all interactions shows the smallest exponent of 1.4. As we remove interactions from the system, μ gradually approaches the SAW limit. The scaling appears independent of HI as shown in Fig. S15. The change in power law may also be understood by looking at the free energy plotted in Fig. 6(c) inset. One can see that the free energy has a higher tilt in the OFF state compared to the other states, implying that the bead-pairs can move along the landscape quicker in the OFF state. The results for $\langle t_{\mathrm{L}}^{*} \rangle$ suggests that even in the absence of loop extrusion, the looping time is not too long (seconds to minutes). This also indicates that the micro phase-separation could be a viable mechanism for bringing together chromatin segments and possibly explains the experimentally observed fact that chromatin is functional even in the absence of loop extruding factors [11, 23, 24]. We then examined how the interaction strength influences $\langle t_{\mathrm{L}}^{*} \rangle$, and found that there is a huge spread in the $\langle t_{\mathrm{L}}^{*} \rangle$ values, for a given ϵ (Fig. 7(c)), with the average showing a mild dependence on ϵ (inset).

Interestingly the values of $\langle t_{\mathrm{C}}^{*} \rangle$ are nearly independent of genomic separation (Fig. 7(d)). Here too, there is huge variability among different bead pairs, with the inset showing the behaviour when the segment length is averaged over all pairs having the same s. However, the interaction strength significantly alters the $\langle t_{\mathrm{C}}^{*} \rangle$ (Fig. 7(e)) showing an exponential increase (see sec. S9 for a discussion on the relation between t_{C}^{*} and ϵ). This suggests that the $\langle t_{\mathrm{L}}^{*} \rangle$ is determined by the interplay between entropy (resulting from genomic separation) and energy (interaction strength). Once bead pairs come in contact, $\langle t_{\mathrm{C}}^{*} \rangle$ is dominated by the interaction strength.

For the OFF and SAW states, we also show $\langle t_{\mathrm{L}}^{*} \rangle$ between all pairs of beads as a heatmap (see Fig. 7(f) & (g)). One can quickly note that the range of SAW time scales is much higher than that of the OFF state. This is the consequence of higher μ for the SAW compared to the OFF state. In the SAW, one can observe that the times are similar for all points having the same distance away from the diagonal (a line parallel to the diagonal axis), suggesting that what matters, in this case, is the inter bead distance (s). In contrast, in the OFF state, there is heterogeneity and curvy colour contours suggesting that the time values are not just a function of segment length alone but also the identity (interaction strength) of the individual bead pairs. In other words, If entropy is the only thing that mattered, the time values will depend only on the genomic distance (like in the no-interaction SAW case). But here, as observed in Fig. 7(c), the interaction strength values also play a role, albeit small

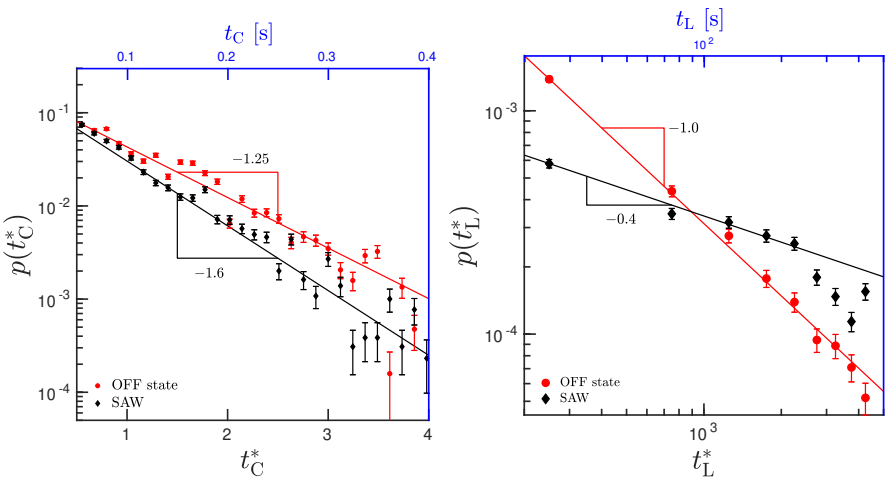
compared to the role of entropy. Hence the curvy colour contours in Fig. 7(f) once again points to the interplay between energy and entropy.

Nature of loop formation and contact time distributions

So far, we have studied the average loop formation times and contact times; however, should one assume that the average values describe these quantities completely? To answer this, similar to $p(r^*)$, here we have investigated the nature of the distribution of the temporal quantities. In Fig. 8(a) and (b), we present the probability distributions of contact ($p(t^*_{\mathrm{C}})$) and loop formation ($p(t^*_{\mathrm{L}})$) times, respectively. We observe that $p(t^*_{\mathrm{C}}) \sim \exp\left(-t^*_{\mathrm{C}}/\tau_c\right)$ with the average time τ_c that depends on the epigenetic state (SAW: $\tau_c = 1/1.6$, OFF: $\tau_c = 1/1.25$). τ_c is small for the SAW and it increases as we add interactions to the system. However, interestingly, the probability of loop formation time (t^*_{L}) has a power law decay ($p(t^*_{\mathrm{L}}) \sim (t^*_{\mathrm{L}})^{-\gamma}$) (Also see Fig. S16). This suggests that there is a huge diversity in loop formation times, and the average looping time alone may not be sufficient to describe the loop formation phenomena. We find that the epigenetic states alter the slope of the distribution (SAW: $\gamma = 0.4$, OFF: $\gamma = 1.0$), keeping the overall nature the same. Comparison of these two distributions reveals that quantitatively the t^*_{L} is much larger than τ^*_c, indicating that chromatin segments take longer to come into contact but stay in contact for a short time. Interestingly, earlier studies for yeast and drosophila have reported broadly comparable behaviour [44]. Also, see Fig. S16 for a discussion on finite size effects introducing exponential tails in the distribution.

Suggestions for experiments to test our predictions

Since we take HiC-like data as input and predict positional fluctuations and dynamics of chromatin segments, microscopy is the ideal method to test our predictions [10, 11, 64]. All $p(r)$ predictions (Fig. 1) may be tested either via live (without fixing) microscopy experiments or by collecting a large number of frozen snapshots of segment-locations via FISH or equivalent methods. Imaging experiments may also estimate the volume occupied by a domain (Fig. 2). From the positional fluctuation data, one can also obtain the effective stiffness as described earlier in this paper (Fig. 6). To measure the time-dependent quantities



(a) (b)

FIG. 8: Distribution function for contact time (t_C^*) and loop formation time (t_L^*) for a specific bead-pair (bead 5 and bead 30) in SAW and OFF state are shown in (a) and (b), respectively.

(Fig. 7, Fig. 8), apart from live microscopy experiments, one may also design appropriate FRET pairs that can probe quantities like the contact time [68]. Obtaining all of these quantities for different epigenetic states would facilitate comparison with our predictions.

DISCUSSION

In this paper, we have investigated the dynamics of chromatin domains, using an inverse Brownian dynamics model including hydrodynamic interactions, and presented the corresponding results. In this section, we discuss several aspects of the model and results in greater detail. In particular, we would like to highlight the novel aspects of our model in the context of earlier work in this area. In our work, we focus on chromatin organization and dynamics within a domain, using parameters obtained via inverse optimization, consistent with chromatin conformation capture experiments.

In the literature, there have been many remarkable studies to simulate chromatin organization within the nucleus. A number of models have used the forward method to obtain 3D configurations and dynamics of chromatin [22, 28, 41, 69, 70]. In the forward method, authors estimate the set of model parameters and the system is simulated to obtain time-

dependent/steady-state configurations, typically using Monte Carlo or coarse-grained Molecular dynamics methods. Such studies have been used to obtain 3D organization of yeast genome, regions of drosophila, mouse and human genome [28, 71–73]. Many of these studies point out interesting statistical properties of the chromatin polymer, such as the scaling of contact probabilities with genomic separation and root mean square spatial distance. They also investigate the organization of the active/inactive regions within a chromatin/within the nucleus [22, 41, 69, 71, 74].

There have been a few models that compute chromatin organization starting with the Hi-C data. Some of these models assume a mathematical relation between the contact probability and mean 3D distance and covert the contact matrix to a distance matrix. A subgroup of these models employ a harmonic interaction between various bead pairs and find optimal spring constants, leading to a Gaussian model [47, 58, 72]. Another approach has been to predict a population of static configurations by iteratively inserting contacts and finding the optimal structures that maximize a likelihood function connecting the contact matrix and the model [45, 46]. A different approach has been to employ a coarse-grained polymer model explicitly accounting for binder proteins; these models obtain the optimal number of protein types and binding sites that matched with Hi-C data [38]. While these models have the details about the protein binding locations, they do not optimize for the interaction strengths directly; once they obtain the optimal number of protein types and bindings sites, the concentration and interaction strengths of proteins are varied as separate parameters. An alternative approach has been to start with the epigenetic (histone modification/protein binding) data and compute interaction strengths that can generate 3D configurations [34, 71].

While these models provide many exciting results and predictions, another approach is to optimise for short-range interaction potentials representing "bonds" that can be broken when segments pairs are far away. Giorgetti et al. and our own work have attempted to do this [30, 31, 49]. While Giorgetti et al. do Monte Carlo simulation of a region of mouse X-chromosome to compute the 3D distance between segments, our focus is to study a wide range of static and dynamic properties of chromatin domains in human chromosomes, using Brownian dynamics, with hydrodynamic interactions. Furthermore, Giorgetti et al. use a square well potential; we use the SDK potential that has attractive and repulsive parts that smoothly vary along r, reaching zero at r_c. There have been a couple of interesting

polymer simulation studies on α globin gene locus [70, 75]. Both the studies investigate the mouse α globin domain, which appears to have a very different contact matrix from the human α-globin that we investigate here. Our studies are complementary to these studies as we predict the interaction strengths, 3D distance distributions and dynamics of human α globin locus that have not been investigated so far.

To study dynamics, there have been many interesting studies that compute mean squared displacement (MSD) of a loci as a function of time and showed whether regions have diffusive or sub-diffusive nature depending on the organism, nature of interactions, and length scale [40, 44, 47, 50, 66, 72, 73, 76]. In the context of yeast and regions of drosophila, mouse, and the human genome, there have been investigations of first meeting times and contact times [44, 66, 76]. Here, we also compute similar quantities and show how the single chromatin domain behaves, taking two loci as examples, with optimised parameters and what the corresponding timescales are. While the precise timescale would vary depending on the nature of the interaction/domain, some features like the power law distributions are common to the earlier observations in drosophila [44].

Since many methods use epigenetic datasets to model chromatin, we would like to argue that using Hi-C data alone has the following advantage. In many a context — for example, when studies are done under different drug-treated conditions — we may not have the epigenetic data. We may not know the precise protein concentrations, affinities of proteins, organization of proteins and histone marks under all these conditions. Hence obtaining optimal intra-chromatin interaction strengths for all bead pairs with Hi-C data alone has its advantage and may be useful to study various cell states. In this work, given the contact probability data – without any further information – we predict the intra-chromatin interactions that are a net result of different epigenetic marks, their intensities, different protein/DNA interactions and distributions/concentrations of proteins.

Another point worthy of discussion is whether considering a small domain is justifiable or not. First of all, it is computationally expensive to simulate the whole chromatin at higher resolution with hydrodynamic interactions; moreover, from a biology point of view, there are sound arguments that suggest that simulating a small domain is reasonable. From the chromatin contact probability data, we know that chromatin is organised into multiple domains (also often called TADs). This data suggest that one domain has much less probability (negligible probability) to interact with another domain. Hence their inter-domain

interaction can be assumed to be negligible. Since our aim here is to probe the dynamics of beads within a domain, it is reasonable to simulate a single domain alone. Also, note that the gene regulation and cellular processes are thought to be happening within a domain. Interestingly, recent experimental data has shown that when the chromatin polymer is cut and each domain is separated out, most of those domains are stable and behave very similar to how they behaved within the whole chromatin, indicating that each domain is independent [15]. This justifies our investigation of each chromatin domain as a separate entity. Interestingly, the intra-chromatin contacts from HiC will shrink the domain and make it dense. We find that the volumic density within the domain — defined as the ratio of the segment length to the volume of the domain — is 0.005bp/nm^3. This is the same order of volumic density as mentioned in Ghosh and Jost [44].

One of the important conclusions of polymer physics is that the temporal quantities are greatly affected by HI [62, 63, 77–84]. Earlier works that used HI investigated the active dynamics of a long polymer with the aim of investigating correlated motion due to activity [74]. While these studies gave insights about the potential large scale movements inside the nucleus due to activity, the dynamics of chromatin in the length scale relevant for gene regulation (scale of domain/TADs) is not investigated here. To the best of our knowledge, this work is the first attempt to study the chromatin dynamics with HI at the length scale of a domain. Here, we show that while the relaxation times do depend on HI, other quantities on the length scale of the domain remain less affected by HI. However, there are interesting questions that remain to be addressed on how HI affects dynamics in a crowded environment, given that the concentration of macromolecules in a cell is in an unentangled semi-dilute regime, where HI is known to be important. However, investigating dynamics in a crowded environment is a difficult problem given that HI is a long-ranged and many-body interaction. The problem of the role of screening of HI is also important in this context. Our current study could be the first step towards this direction.

Quantities calculated here have immense physical and biological significance. The finding of two peaks in the distribution function is a novel aspect that can have ramifications. As mentioned earlier, there is an ongoing debate in the field about whether gene regulation requires actual physical contact between two regulatory segments or only the proximity would suffice. Cellular processes such as transport of proteins from one region to another (e.g. enhancer-promoter), spreading of histone modifications in the 3D space etc., would

crucially depend on $p(r)$ [85, 86]. For example, given r, one can compute the time (τ_p) for proteins/enzymes to diffuse from location r_i to r_j. The mean time would depend on the distribution as $\langle \tau_p \rangle = \int \tau_p p(r) dr$. However, apart from a distance among segments, the accessibility would depend on the local compactness and diffusivity too. That is, compactness of the domain (Fig. 2) and effective viscous drag (Fig. 6) together with $p(r)$(Fig. 1) would be crucial for understanding how physics of chromatin would affect biological function. Given that phase separation is argued to be one of the important factors determining domain formation, our study reveals how the interplay between intra-chromatin interactions and polymer dynamics would affect loop formation and contact times.

CONCLUSION

Even though there is a great improvement in our understanding of the static nature of chromatin organization, very little is known about the dynamics, which is a crucial aspect of in vivo chromatin. With the advancement of technology, it is now possible to experimentally probe the fluctuations and dynamics of chromatin polymer. However, the main challenge to simulate the dynamics of chromatin is that we do not know the interaction strength parameters among different segments. We have overcome this challenge by using an inverse technique and obtained optimal interaction strengths between all chromatin segments, and used it to investigate the dynamics of a chromatin domain.

We summarize our key findings: (i) We investigated the 3D organization of two chromatin domains for two different epigenetic states in each case. (ii) Starting from 5C/HiC data, we predicted the optimal intra-chromatin interactions strengths for all cases revealing how epigenetic changes would affect the interactions and 3D organization of chromatin. (iii) Going beyond the average properties, we computed the distance probability distribution; we observe a double peak – an interaction-driven peak and an entropy-dominated peak – depending on the epigenetic state and chromatin segment pair identity. (iv) Introducing perturbations in optimized interaction strength values to systematically mimic epigenetic-like states, we show how perturbations would alter $p(r)$; the distance distribution between a given bead pair depends on the interaction strength of all other pairs suggesting the cooperative nature of chromatin folding. (v) Volume and the shape properties of the chromatin domain depends on the magnitude of interaction strengths present, whether it is an epige-

netic state or perturbed state. The OFF state of α-globin gene is highly collapsed/compact, more spherical compared to the extended, less spherical ON state or SAW. (vi) Our simulations investigated beyond the pairwise contact probability information and predicted the probability of three segments coming together (triple contact). (vii) The relaxation time of the domain is dependent on the magnitude of the interaction strengths in the domain. Counter-intuitively, the relaxation time of a highly interacting OFF state is much shorter than that of a non-interacting SAW polymer. We explain this phenomenon by computing the effective stiffness of the domain from polymer fluctuations. We also show that the OFF state has a higher effective drag. (viii) We study dynamics accounting for crucial hydrodynamic interactions; we show that HI has a significant influence on the relaxation time of the chromatin domain. With HI, the domain takes half the time to relax as compared to the no-HI case. (ix) We compute the loop formation time and the time for the looped bead pairs to remain in contact. We show that average looping time has different scaling with genomic separation, depending on the nature of the chromatin states having different interaction strengths. The looping times show a power law distribution indicating multiple timescales that might be involved with looping. On the other hand, the contact time has an exponential distribution.

This study can be further extended genomewide to examine various gene loci and investigate the fluctuations and dynamics of all domains in the genome. Such polymer models are useful for examining aspects like the spread of histone modifications and accessibility of the domains. We hope that this study will catalyse new experimental and computational studies examining the interplay between epigenetics and polymer dynamics.

Acknowledgments We thank Burkhard Dünweg, Dibyendu Das and Rajarshi Chakraborty for enlightening discussions. The work was supported by the MonARCH and SpaceTime computational facilities of Monash University and IIT Bombay, respectively. We also acknowledge the funding and general support received from the IITB-Monash Research Academy, DST, SERB and DBT India.

Conflict of Interests: The authors declare that they have no conflict of interest.

[1] W. A. Bickmore, Annu. Rev. Genomics Hum. Genet. **14**, 67 (2013).

[2] B. D. Pope, T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, et al., Nature **515**, 402 (2014).

[3] P.-C. Wei, C.-S. Lee, Z. Du, B. Schwer, Y. Zhang, J. Kao, J. Zurita, and F. W. Alt, PNAS **115**, 1919 (2018).

[4] A. Sanyal, B. R. Lajoie, G. Jain, and J. Dekker, Nature **489**, 109 (2012).

[5] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, et al., Science **326**, 289 (2009).

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

[7] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, et al., Nature **485**, 381 (2012).

[8] T. Nagano, Y. Lubling, C. Várnai, C. Dudley, W. Leung, Y. Baran, N. M. Cohen, S. Wingett, P. Fraser, and A. Tanay, Nature **547**, 61 (2017).

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

[10] T. Nozaki, R. Imai, M. Tanbo, R. Nagashima, S. Tamura, T. Tani, Y. Joti, M. Tomita, K. Hibino, M. T. Kanemaki, et al., Mol. Cell **67**, 282 (2017).

[11] B. Bintu, L. J. Mateo, J.-H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang, Science **362** (2018).

[12] Q. Szabo, A. Donjon, I. Jerković, G. L. Papadopoulos, T. Cheutin, B. Bonev, E. P. Nora, B. G. Bruneau, F. Bantignies, and G. Cavalli, Nat. Genet. **52**, 1151 (2020).

[13] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, et al., Cell **159**, 1665 (2014).

[14] M. J. Rowley and V. G. Corces, Nat. Rev. Genet. **19**, 789 (2018).

[15] H. Belaghzal, T. Borrman, A. D. Stephens, D. L. Lafontaine, S. V. Venev, Z. Weng, J. F. Marko, and J. Dekker, Nat. Genet. pp. 1–12 (2021).

[16] E. Alipour and J. F. Marko, Nucleic Acids Res. **40**, 11202 (2012).

[17] M. Mir, W. Bickmore, E. E. Furlong, and G. Narlikar, Development **146**, dev182766 (2019).

[18] J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, and L. A. Mirny, Proceedings of the National Academy of Sciences **115**, E6697 (2018).

[19] S. J. Nair, L. Yang, D. Meluzzi, S. Oh, F. Yang, M. J. Friedman, S. Wang, T. Suter, I. Alsha-

reedah, A. Gamliel, et al., Nat. Struct. Mol. Biol. **26**, 193 (2019).

[20] D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp, Cell **169**, 13 (2017).

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

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

[23] A. Kaushal, G. Mohana, J. Dorier, I. Özdemir, A. Omer, P. Cousin, A. Semenova, M. Taschner, O. Dergai, F. Marzetta, et al., Nat. Commun. **12**, 1 (2021).

[24] N. S. Benabdallah, I. Williamson, R. S. Illingworth, L. Kane, S. Boyle, D. Sengupta, G. R. Grimes, P. Therizols, and W. A. Bickmore, Mol. Cell **76**, 473 (2019).

[25] B. A. Gibson, L. K. Doolittle, M. W. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, and M. K. Rosen, Cell **179**, 470 (2019).

[26] K. Shrinivas, B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V. Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, R. A. Young, et al., Mol. Cell **75**, 549 (2019).

[27] Y. Zhang, R. P. McCord, Y.-J. Ho, B. R. Lajoie, D. G. Hildebrand, A. C. Simon, M. S. Becker, F. W. Alt, and J. Dekker, Cell **148**, 908 (2012).

[28] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic Acids Res. **42**, 9553 (2014).

[29] J. J. Parmar and R. Padinhateeri, Curr. Opin. Struct. Biol. **64**, 111 (2020).

[30] K. Kumari, B. Duenweg, R. Padinhateeri, and J. R. Prakash, Biophys. J. **118**, 2193 (2020).

[31] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, Cell **157**, 950 (2014).

[32] S. Bianco, D. G. Lupiáñez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, et al., Nat. Genet. **50**, 662 (2018).

[33] G. D. Bascom, C. G. Myers, and T. Schlick, PNAS **116**, 4955 (2019).

[34] M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, and J. N. Onuchic, PNAS **113**, 12168 (2016).

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

[36] M. Di Stefano, J. Paulsen, D. Jost, and M. A. Marti-Renom, Curr. Opin. Genet. Dev. **67**, 25 (2021).

[37] C. T. Clarkson, E. A. Deeks, R. Samarista, H. Mamayusupova, V. B. Zhurkin, and V. B. Teif, Nucleic Acids Res. **47**, 11181 (2019).

[38] M. Conte, L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, and M. Nicodemi, Nat. Com-

mun. **11**, 1 (2020).

[39] Y. Qi and B. Zhang, PLoS Comput. Biol. **15**, e1007024 (2019).

[40] G. Shi, L. Liu, C. Hyeon, and D. Thirumalai, Nat. Commun. **9**, 1 (2018).

[41] Q. MacPherson, B. Beltran, and A. J. Spakowitz, PNAS **115**, 12739 (2018).

[42] G. Bajpai and R. Padinhateeri, Biophys. J. **118**, 207 (2020).

[43] C. A. Brackey, D. Marenduzzo, and N. Gilbert, Nature Methods **17**, 767 (2020).

[44] S. K. Ghosh and D. Jost, PLoS computational biology **14**, e1006159 (2018).

[45] H. Tjong, W. Li, R. Kalhor, C. Dai, S. Hao, K. Gong, Y. Zhou, H. Li, X. J. Zhou, M. A. Le Gros, et al., Proceedings of the National Academy of Sciences **113**, E1663 (2016).

[46] N. Hua, H. Tjong, H. Shin, K. Gong, X. J. Zhou, and F. Alber, Nature protocols **13**, 915 (2018).

[47] S. Shinkai, M. Nakagawa, T. Sugawara, Y. Togashi, H. Ochiai, R. Nakato, Y. Taniguchi, and S. Onami, NAR genomics and bioinformatics **2**, lqaa020 (2020).

[48] G. Tiana, F. Villa, Y. Zhan, R. Capelli, C. Paissoni, P. Sormanni, E. Heard, L. Giorgetti, and R. Meloni, Computer Physics Communications **186**, 93 (2015).

[49] G. Tiana, A. Amitai, T. Pollex, T. Piolot, D. Holcman, E. Heard, and L. Giorgetti, Biophysical journal **110**, 1234 (2016).

[50] M. M. Tortora, H. Salari, and D. Jost, Current opinion in genetics & development **61**, 37 (2020).

[51] M. Falk, Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, H. Leonhardt, B. Joffe, J. Dekker, G. Fudenberg, I. Solovei, et al., Nature **570**, 395 (2019).

[52] G. Bajpai, D. A. Pavlov, D. Lorber, T. Volk, and S. Safran, Elife **10**, e63976 (2021).

[53] R. Bird, C. Curtiss, R. Armstrong, and O. Hassager, Dynamics of polymeric liquids, kinetic theory (volume 2) (1987).

[54] T. Soddemann, B. Dünweg, and K. Kremer, Eur. Phys. J. E **6**, 409 (2001).

[55] A. Santra, K. Kumari, R. Padinhateeri, B. Dünweg, and J. R. Prakash, Soft Matter **15**, 7876 (2019).

[56] J. Des Cloizeaux, J. Phys **41**, 223 (1980).

[57] R. Imai, T. Nozaki, T. Tani, K. Kaizu, K. Hibino, S. Ide, S. Tamura, K. Takahashi, M. Shribak, and K. Maeshima, Mol. Biol. Cell **28**, 3349 (2017).

[58] G. Shi and D. Thirumalai, Physical Review X **11**, 011051 (2021).

[59] S. Wang, J.-H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C.-t. Wu, and X. Zhuang, Science **353**, 598 (2016).

[60] K. Maeshima, S. Tamura, J. C. Hansen, and Y. Itoh, Curr. Opin. Cell Biol. **64**, 77 (2020).

[61] H. Strickfaden, T. O. Tolsma, A. Sharma, D. A. Underhill, J. C. Hansen, and M. J. Hendzel, Cell (2020).

[62] T. T. Pham, M. Bajaj, and J. R. Prakash, Soft Matter **4**, 1196 (2008).

[63] T. T. Pham, B. Duenweg, and J. R. Prakash, Macromolecules **43**, 10084 (2010).

[64] T. Germier, S. Kocanova, N. Walther, A. Bancaud, H. A. Shaban, H. Sellou, A. Z. Politi, J. Ellenberg, F. Gallardo, and K. Bystricky, Biophys. J. **113**, 1383 (2017).

[65] A. Amitai and D. Holcman, Physical Review E **97**, 032417 (2018).

[66] Y. Zhang and O. K. Dudko, Annual review of biophysics **45**, 117 (2016).

[67] N. M. Toan, G. Morrison, C. Hyeon, and D. Thirumalai, J. Phys. Chem. B **112**, 6094 (2008).

[68] J. G. Yang, T. S. Madrid, E. Sevastopoulos, and G. J. Narlikar, Nat. Struct. Mol. Biol. **13**, 1078 (2006).

[69] S. H. Sandholtz, Q. MacPherson, and A. J. Spakowitz, PNAS **117**, 20423 (2020).

[70] C. A. Brackley, J. M. Brown, D. Waithe, C. Babbs, J. Davies, J. R. Hughes, V. J. Buckle, and D. Marenduzzo, Genome biology **17**, 1 (2016).

[71] L. Liu, G. Shi, D. Thirumalai, and C. Hyeon, PLoS computational biology **14**, e1006617 (2018).

[72] O. Shukron and D. Holcman, PLoS computational biology **13**, e1005469 (2017).

[73] M. Di Pierro, D. A. Potoyan, P. G. Wolynes, and J. N. Onuchic, Proceedings of the National Academy of Sciences **115**, 7753 (2018).

[74] D. Saintillan, M. J. Shelley, and A. Zidovska, Proceedings of the National Academy of Sciences **115**, 11442 (2018).

[75] A. M. Chiariello, S. Bianco, A. M. Oudelaar, A. Esposito, C. Annunziatella, L. Fiorillo, M. Conte, A. Corrado, A. Prisco, M. S. Larke, et al., Cell Reports **30**, 2125 (2020).

[76] N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, and C. Murre, Nature communications **10**, 1 (2019).

[77] J. R. Prakash, Korea-Aust. Rheol. J. **21**, 245 (2009).

[78] J. R. Prakash, Curr. Opin. Colloid Interface Sci. **43**, 63 (2019).

[79] R. Prabhakar, C. Sasmal, D. A. Nguyen, T. Sridhar, and J. R. Prakash, Phys. Rev. Fluids **2**,

011301 (2017).

[80] R. Prabhakar and J. R. Prakash, J. Non-Newtonian Fluid Mech. **116**, 163 (2004).

[81] C. M. Schroeder, J. Rheol. **62**, 371 (2018).

[82] P. Sunthar and J. R. Prakash, Europhys. Lett. **75**, 77 (2006).

[83] P. Sunthar and J. R. Prakash, Macromolecules **38**, 617 (2005).

[84] C. M. Schroeder, E. S. Shaqfeh, and S. Chu, Macromolecules **37**, 9242 (2004).

[85] M. Katava, G. Shi, and D. Thirumalai, bioRxiv (2021).

[86] D. Jost and C. Vaillant, Nucleic acids research **46**, 2252 (2018).