Inferring interphase chromosomal structure from multiplexed fluorescence in situ hybridization data: a unified picture from human and mouse cells

Loucif Remini, Midas Segers, Andrea Parmeggiani, Enrico Carlon

Inferring interphase chromosomal structure from multiplexed fluorescence in situ hybridization data: a unified picture from human and mouse cells

Loucif Remini,[1] Midas Segers,[2] Andrea Parmeggiani,[1] and Enrico Carlon[2]
[1)] *Laboratoire Charles Coulomb (L2C), Univ Montpellier, CNRS, Montpellier, France*
[2)] *Soft Matter and Biophysics, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium*

(Dated: 17 October 2024)

We analyze multiplexed fluorescence in situ hybridization (m-FISH) data for human and mouse cell lines. The m-FISH technique uses fluorescently-labeled single stranded probes which hybridize to specific chromosomal regions, thereby allowing the measurement of the spatial positions of up to ~ 100 tagged sites for several thousands interphase chromosomes. Our analysis focuses on a wide range of different cell lines and two distinct organisms and provides a unified picture of chromatin structure for scales ranging from 5 kb (kilo bases) up to 2 Mb (mega bases), thus covering a genomic region of almost three orders of magnitude. Confirming recent analysis (Remini et al., Phys. Rev. E **109**, 024408 (2024)) we show that there are two characteristic arrangements of chromatin referred to as phase α (crumpled globule) and β (looped domain) and discuss the physical properties of these phases. We show that a simple heterogeneous random walk model captures the main behavior observed in experiments and bring considerable insights on chromosomal structure.

I. INTRODUCTION

The spatial organization of chromosomes within the cell nucleus plays a crucial role in regulating gene expression and other nuclear processes[1]. Understanding the folding principles of chromatin, the complex of DNA and proteins constituting chromosomes, is essential for deciphering the mechanisms underlying genome function and is a topic which has attracted considerable interest in the recent years[2–8]. Recent advances in experimental techniques, such as multiplexed fluorescence in situ hybridization (m-FISH), have enabled high-resolution imaging of chromatin structure, providing several insights into its spatial arrangement[5,9,10]. Unlike the High-throughput Chromosome Conformation Capture (Hi-C) technique[11], which allows to identify contact regions for the whole genome in a single experiment, m-FISH provides a more limited information on spatial position of some selected, fluorescently tagged, genomic loci. Current m-FISH experiments can tag about ~ 100 sites, probing several thousands cells/chromosomes in a single run[9,12,13]. m-FISH data display a high correlation with Hi-C contact frequency matrices[9].

In our previous work[14], we analyzed m-FISH data from Bintu et al.[9] to investigate the spatial organization of chromatin within specific regions of human chromosome 21. By examining the probability distributions of distances between pairs of labeled genomic sites, we identified two coexisting chromatin conformations, referred to as the α and β phases, arranged in alternated micro-domains. The geometric properties of these phases were inferred from scaling laws[14] revealing a crumpled globule behavior[15] for the α phase and a localized, possibly looped, domain, for the β phase.

We extend this analysis here to novel experimental data sets for mouse chromosome 19[16], which confirms the microphase separation scenario reported earlier for human chromosome 21[14]. The more recent mouse m-FISH data[16] have an increased genomic resolution which allows us to explore the scaling behavior of the two phases at shorter scales compared to the earlier study[14]. Combining mouse and human m-FISH data we infer a general unified picture for chromatin structure for a range of genomic scales encompassing almost three order of magnitudes of genomic distances (from ~ 5 kb up to ~ 2 Mb), such picture is supported by a simple heterogeneous random walk model. In addition, we explore several features of the chromosomes organization as its intrinsic non-markovianity, the short scale behavior of contact probabilities, and multi-point distribution functions. Overall the analysis reported here illustrates the power and advantages of quantitative high resolution microscopy combined with a statistical physics modeling framework for data analysis. While currently restricted to specific selected chromosomal regions of a few organisms, we expect that future m-FISH experiments will shed lights and bring further insights on chromosomal organization.

This paper is organized as follows. In Sec. II we review the main features of the analysis of the m-FISH technique, discussing the distance distributions, contact probabilities, gyration radii and other quantities which point to the existence of distinct phases of chromatin. In Sec. III we provide a unified description of the chromatin organization highlighting different scaling regimes at different length scales. We introduce a heterogeneous random walk model which reproduces the main features observed in experiments. Some open issues are also discussed in Sec. III.

II. ANALYSIS OF MULTIPLEXED FISH DATA

In a m-FISH experiment the position of a set of $1 \le i \le N$ fluorescently tagged genomic loci are recorded for M cells in a single experimental run[9,13,16]. This results in three dimensional coordinates $[x_i^{(k)}, y_i^{(k)}, z_i^{(k)}]$ denoting the position of the tag i for the chromosome $1 \le k \le M$. Typical experimental spatial accuracy is of 50 nm.

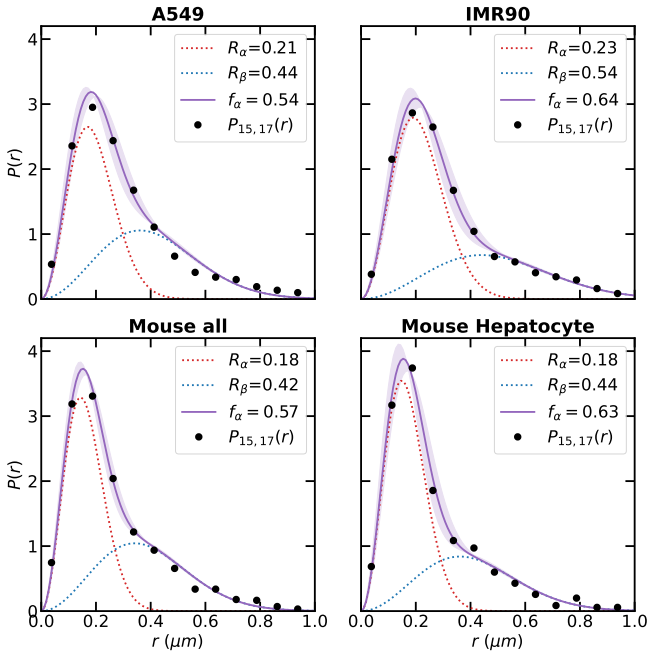

FIG. 1. Black circles: m-FISH experimental distance distributions over M distinct cells between two fixed tags i and j as specified by the labels (i,j) shown in the panel. (Top) data from Bintu et al[9] and (Bottom) data from Liu et al[16]. Red and blue dashed lines are the gaussian components $f_\alpha g(r;R_\alpha)$ and $(1-f_\alpha)g(r;R_\beta)$ representing the contribution of the two phases, see Eqs. (2) and (3). Purple lines are the sum of the two contributions, with shaded areas indicating the fits uncertainties. The fitted values for R_α, R_β and f_α are reported in the graphs. We find consistent values for $f_\alpha \approx 0.6$ for the two organisms. Distance distributions for other datasets and different tags are given in Supplementary Material.

A. Two phases from distance distributions analysis

Our previous analysis of m-FISH data[14] focused on pairwise distance of two given tags $1 \le i,j \le N$ in chromosome k defined as

$$r_{ij}^{(k)} \equiv \left[\left(x_i^{(k)}-x_j^{(k)}\right)^2+\left(y_i^{(k)}-y_j^{(k)}\right)^2+\left(z_i^{(k)}-z_j^{(k)}\right)^2\right]^{1/2} \tag{1}$$

Experimental distributions over the M chromosome replicates for a few selected tags and experiments are shown in Fig. 1 as solid circles (See Supplementary Figures S1, S2, and S3 for more examples). The analysis reveals a distribution of distances over the $1 \le k \le M$ different cells which can be fitted by the combination of two gaussians[14]

$$P_{ij}(r) = f_\alpha\, g(r;R_\alpha) + (1-f_\alpha)\, g(r;R_\beta) \tag{2}$$

where we defined

$$g(r;R) \;\equiv\; 4\pi r^2 \left(\frac{3}{2\pi R^2}\right)^{3/2} e^{-3r^2/(2R^2)} \tag{3}$$

the radial distance distribution, which includes a factor $4\pi r^2$ obtained by integrating over polar angles. The distribution $g(r;R)$ contains R as single fitting parameter, which implies a three parameters fit for Eq. (2): R_α and R_β are the characteristic radii of the two phases, and $0 \le f_\alpha \le 1$ is the fraction of phase α. The analysis with Eq. (2) is done for every pair of tags i and j, encompassing a total of $N(N-1)/2$ pairs, resulting in radii $R_\alpha(i,j)$, $R_\beta(i,j)$ and fraction $f_\alpha(i,j)$ which depend on the tagged sites considered i and j. Figure 1 shows experimental distributions for different pairs of tags (circles), as well as fits of Eq. (2). Data are from Refs. 9 and 16 (See Supplementary Material for details about the experimental data). The former reference reports m-FISH experiments in human chromosome 21 with ~ 80 fluorescent tags placed at consecutive genomic sites separated by 30 kb. The latter one is a study of mouse chromosome 19 measuring 19 different tags separated by a 5 kb distance along the genome, thus allowing to explore the chromatin spatial arrangement at higher resolution. At short genomic distances ($s \lesssim 200$ kb), the fit with a single gaussian distribution clearly fails[14], but at larger genomic scales the radii R_α and R_β tend to approach each other and the distance distributions tend to merge into Eq. (3), see Supplementary Material for more examples.

B. Scaling laws for phases α and β

For any fixed tag i, one could analyze the scaling behavior of the characteristic radii of the two phases using the following ansatz

$$R_\alpha \sim |i-j|^{\nu_\alpha} \sim s^{\nu_\alpha}, \qquad R_\beta \sim |i-j|^{\nu_\beta} \sim s^{\nu_\beta} \tag{4}$$

with s the genomic distance between tags i and j measured in bases (for equally spaced tags $|i-j|$ and s are proportional to each other). Figure 2 shows a log-log scale plot of the radii $\overline{R}_\alpha$ and $\overline{R}_\beta$ obtained from averaging over different tags positions for any given data set. We show in the same figure human chromosome 21 data from Bintu et al[9] covering a genomic range of $s = 30$ kb to 2 Mb and mouse data from Liu et al[16] which has higher resolution, but more limited range $s = 5$ kb to 95 kb. The mouse data set considered in this papers is an ensemble cells from all tissues studied in Liu et al[16]. Scaling for different mouse cells are shown in Fig S4. Figure 2 shows a good numerical overlap for $\overline{R}_\beta$ between human and mouse data. We note a discontinuous jump in $\overline{R}_\alpha$ in the two organisms, indicating that the α phase is more compact in the human chromosome 21 as compared to the mouse dataset. The resulting scaling exponent, as inferred by the slope in the log-log scale, seems to be consistent in both cases.

1. *The α phase is consistent with crumpled globule behavior*

The emerging picture from the analysis of multiple loci is that the phase α has scaling properties consistent with those of a crumpled globule, characterized by an exponent $\nu_\alpha = 1/3$ (see Fig. 2, red symbols). The crumpled globule is a long-lived metastable state obtained from quenching a long polymer from a good solvent condition to a temperature below its

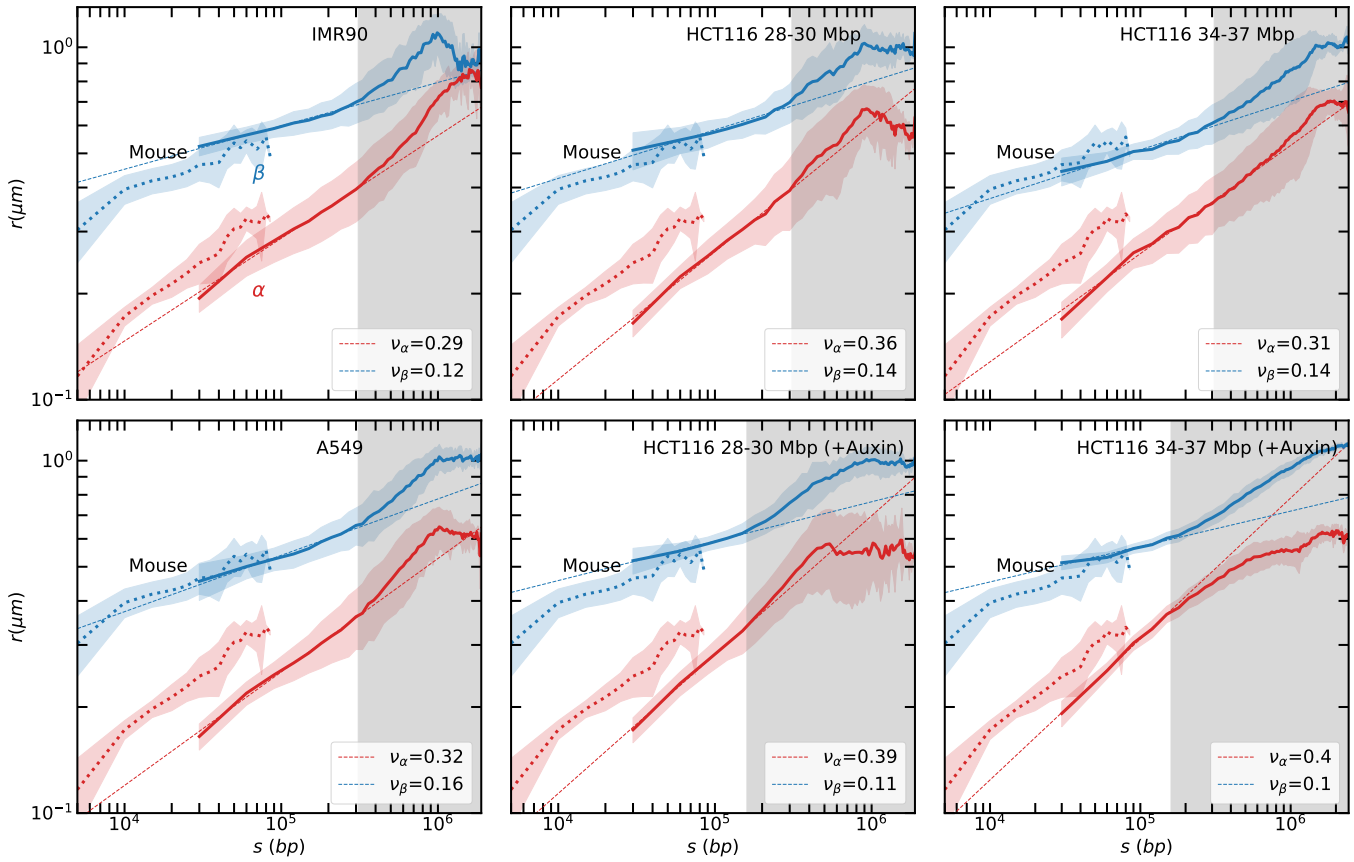

FIG. 2. Log-log plot of average radii $\overline{R}_\alpha$ (red lines) and $\overline{R}_\beta$ (blue lines) vs. s for different human cell lines as well as mouse data, with the shaded areas giving the uncertainties over the averages. Human cell lines are indicated in the graphs (IMR90, HCT116, A549 ... see Supplementary Material for dataset specifications) and are from Bintu et al[9]. They correspond to a genomic range 30 kb $\leq s \leq$ 2.5 Mb. The mouse data reported are averages over cells of different types from experiments from Liu et al[16] corresponding to 5 kb $\leq s \leq$ 95 kb. Thin dashed lines are power-law fits (see Eq. (4)) over $s \leq 350$ kb region for untreated samples and $s \leq 160$ kb for auxin treated ones, see Section II E for details. Fitted values for the scaling exponents ν_α and ν_β reported in each graph.

θ-point. The quenching leads to a crumpling of the polymer to longer and longer scales until a stationary state is reached, which is referred to as crumpled or fractal globule[15]. Differently from its equilibrium counterpart, e.g. a polymer below its θ-point undergoing a collapse transition, the crumpled globule is unknotted and also less dense. This is believed to be advantageous for the chromosomal structure as it facilitates the accessibility to different parts of the long chromosomes in living cells[17] by enzymes and other regulatory factors. The crumpled globule was identified in early Hi-C experiments[11], which reported a contact probability $p_c(s)$ scaling as $p_c(s) \sim s^{-1}$, asymptotically for large genomic distances s.

2. *The β phase is strongly localized*

For the phase β (Fig. 2, blue symbols) most of the dataset and tags give consistently a small estimate for the scaling exponent $\nu_\beta \approx 0.15$ suggesting a confined phase, such as a chromatin conformation with multiple loops. The return to the loops origin yields a characteristic scaling radius which does not grow with the genomic distance $R_\beta \sim s^0$. Multiple loops with disjoint origins may lead to a small increase of the radius $\overline{R}_\beta$ with the genomic distance, as discussed in Sec. III. We note that the mouse and human data for $\overline{R}_\beta$ agree well with each other in the genomic region 30 kb $\leq s \leq$ 95 kb where the two data sets overlap. In the vicinity of the shortest length scale $s \approx 5$ kb, the mouse data for $\overline{R}_\beta$ bend towards higher slopes, a behavior that suggests a change in scaling, which will be discussed more extensively in Sec. III.

3. *Microphase separation in α and β domains*

Unlike most common thermodynamic phase separating systems (such as liquid-gas) in which each phases becomes macroscopically extended, here the chromatin separates in *microphases* i.e. in alternating finite-sized domains of phase α and phase β. A typical example of microphase separation is given by block copolymers, where the material spontaneously organizes into distinct, nanometer-scale domains or

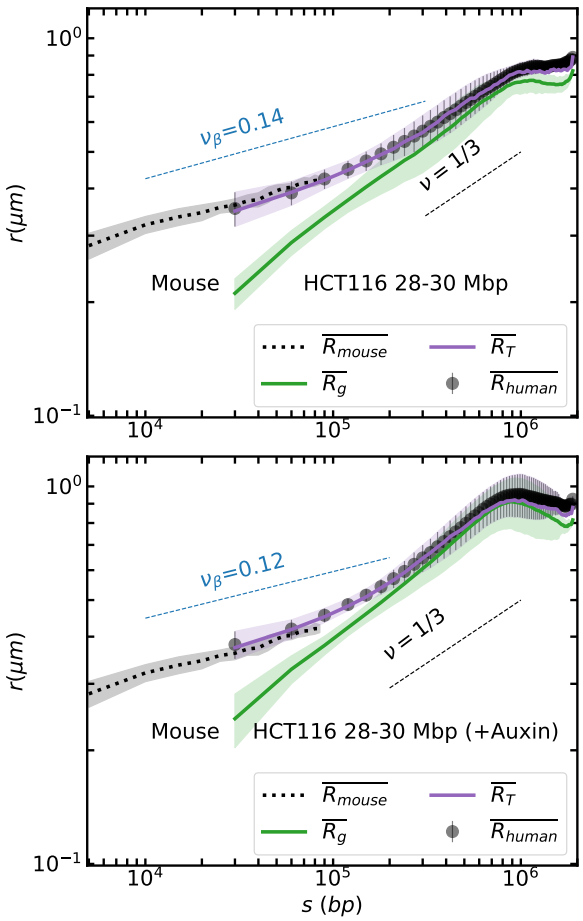

FIG. 3. Comparison of the mean experimental mean squared radius of the raw data from Bintu et al[9] $\overline{R_{human}(s)}$ (black) with the reconstructed radius $\overline{R_g(s)}$ from the single-phase model (green) and the total radius $\overline{R_T(s)}$ from the two-phase model for untreated (Top) and auxin-treated (Bottom) HCT116 cell line. The dotted black line is the the mean squared radius extracted from all mouse cells data of Liu et al[16] covering a higher genomic resolution interval. The thin dashed lines are the scaling exponents of the β phase (blue) and the crumpled globule (black). More cell lines are shown in Figures S5 and S6 in the Supplementary Material.

microphases as lamellar phase. Complete phase separation is not possible due to the finite length of each domain. The typical width of α and β domains can be inferred from the behavior of the local radius of gyration[14] and is about ≈ 150 kb. The Supplementary Material reports several plots of local radius of gyration for single cells for different data sets for human[9] and mouse[16] cell lines.

C. Total radius

Given the probability density function of the two-phase model provided by Eq. (2) one obtains the total mean square radius R_T^2 for any two given tags i and j as a linear combination of R_α^2 and R_β^2:

$$R_T^2(s) \equiv \langle r^2 \rangle = f_\alpha R_\alpha^2(s) + (1-f_\alpha) R_\beta^2(s) \qquad (5)$$

As we have seen from the analysis of Fig. 2 at short scales $R_\beta \gg R_\alpha$, we therefore expect $R_T \sim R_\beta$, i.e. the short scale behavior of the total radius is dominated by the phase β. This is indeed observed in Fig. 3 showing R_T for human and mouse data. There is a remarkable overlap between these two datasets within the genomic distances covered by both experiments resolutions. The mean-squared radius calculated from the raw human cell data is accurately reproduced by the two-phase model $\overline{R_T}$ (represented by the purple solid line), fitting well within its statistical uncertainty across all scales. The microphase coexistence model, as previously discussed, becomes invalid beyond the average width of the coexisting phases. This transition is illustrated in Fig. 3, where both the experimental and reconstructed radii converge and align with the one-phase model fit (shown in green), which scales with a factor of $1/3$. This indicates a convergence towards a crumpled globule organization. In wild-type cells, this transition occurs at approximately ≈ 300 kb, while in auxin-treated cells, it occurs at around ≈ 200 kb. The faster convergence to the one-phase model in auxin-treated cells might be due to the depletion of cohesin-mediated loops, which typically span $100-200$ kb[18,19], potentially leading to a reduction in the domain size of the β phase in these samples. Note that in some cases in the very large s limit $\overline{R_{human}}$ does not grow further, possibly due to the effect of territorial confinement of the chromosome.

D. Contact probability

Hi-C data focus on contact probability $p_c(s)$ of any two given genomic sites separated by a genomic distance s. One can infer the behavior of $p_c(s)$ from the two phase model (2) by integrating $P_{ij}(r)$ for distances $r \leq \lambda$ with λ a small length scale such that ($\lambda \ll R_\alpha, R_\beta$)

$$p_c(s) = \sqrt{\frac{6}{\pi}} \lambda^3 \left[\frac{f_\alpha}{R_\alpha^3(s)} + \frac{1-f_\alpha}{R_\beta^3(s)} \right] \qquad (6)$$

Unlike the case of the total radius, here the contact probability at short distances is dominated by the α phase. Indeed $R_\alpha \ll R_\beta$ at small s implies $p_c(s) \sim R_\alpha^{-3}(s) \sim s^{-1}$. When s grows R_α approaches R_β and this should lead to a slower decay of $p_c(s)$ with s. Finally, for even longer s, the two phase model breaks down and in analogy with the total radius scaling we expect a crumpled globule scaling $\sim s^{-1}$. Figure 4 shows a plot of $p_c(s)$ inferred from m-FISH data, estimated from the fraction of samples in which tags have a spatial distance $r < 0.01$ μm. Crumpled globule behavior predicts a scaling $p_c(s) \sim s^{-1}$, which has been observed in Hi-C data at scales above $s > 1$ Mb. At shorter genomic distances, in the range $s \approx 100$ kb, deviations from the $1/s$ power-law behavior have been reported[19–21], resulting in a characteristic shoulder-like profile as illustrated in the inset of Fig. 4. This

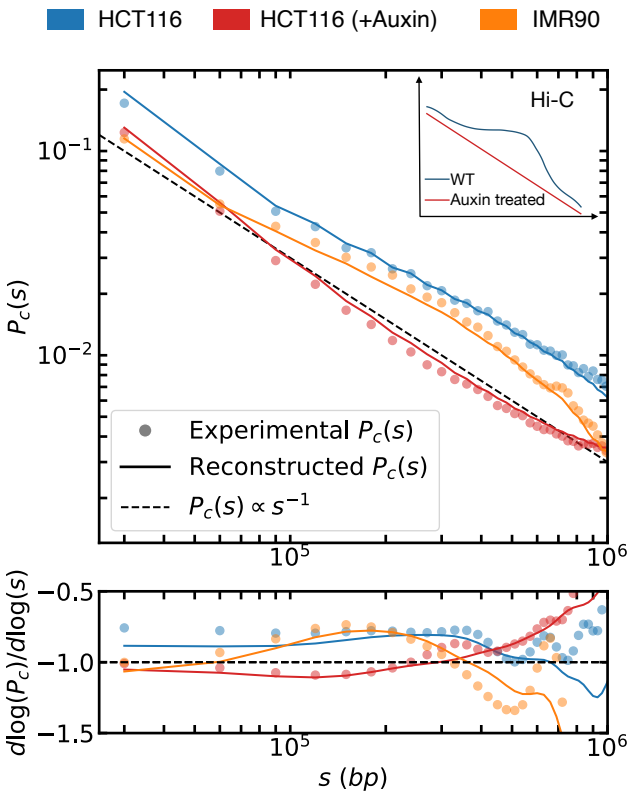

FIG. 4. (Top) Contact probability decay curves, P_c, as a function of genomic distance s, are shown for wild-type HCT116 (blue), IMR90 (orange), and auxin-treated HCT116 cells (red). The circles represent data inferred from raw experimental m-FISH histograms, while the solid lines correspond to predictions from the two-phase model. The inset provides a schematic comparison of contact probability decay with respect to s, as observed in Hi-C experiments, for untreated versus auxin-treated cells, based on findings reported in[19–21]. (Bottom) Log-derivatives of the contact probabilities displayed in the main figure, highlighting the deviation from the $1/s$ power-law behavior in untreated cell lines.

shoulder has been attributed to the effect of cohesin-mediated loops forming on top of a crumpled globule[19,20].

E. Non-markovianity

One of the advantages of m-FISH data, is that one can perform the analysis of correlations between more than two chromatin sites[9]. We discuss here these correlations focusing on conditional distributions $P_{ij}(r|r_{kn} > r_c)$ and $P_{ij}(r|r_{kn} < r_c)$, defined as the distributions for distances between two tags i and j, with a constrain on the distance between the tags k and n, which is required to be larger or smaller than some threshold value r_c. Figure 5 shows plots of $P_{ij}(r|r_{i,i-1} > r_c)$ (in blue) and $P_{ij}(r|r_{i,i-1} < r_c)$ (in red) for $j = i+1, i+2\ldots$ averaged over all i for HCT116 (untreated), HCT116 (+auxin 6h) and IMR90 cell-lines. The fact that we observe different distribution $P_{ij}(r|r_{i,i-1} > \bar{r})$ and $P_{ij}(r|r_{i,i-1} < \bar{r})$ is a proof of the existence of correlations between distances across multiple distinct sites. The two conditional distributions differ from the unconditional one $P_{ij}(r)$ (black) as follows: $P_{ij}(r|r_{i,i-1} > r_c)$ shows an enhancement of the β phase, while the distribution $P_{ij}(r|r_{i,i-1} < r_c)$ shows an enhancement of the α phase. As tag separation $\Delta n \equiv |j-i|$ increases, these phase-specific enhancements gradually diminish, leading to the eventual convergence of all three distributions as the system transitions across multiple α and β domains. This convergence is especially noticeable in the auxin-treated HCT116 cells (see $\Delta n = 13$), suggesting a more rapid loss of correlation as compared to untreated cell lines. We quantify the convergence of the two conditional probabilities through the mean squared deviation between the two distributions $\Gamma(\Delta s)$, with Δs being the genomic distance between tags i and j expressed in numbers of base-pairs.

$$\Gamma(\Delta s) \equiv \langle (P_{ij}(r|r_{i-1,i} > r_c) - P_{ij}(r|r_{i-1,i} < r_c))^2 \rangle \xrightarrow{\Delta s \gg 1} e^{-\Delta s/\xi} \quad (7)$$

where $\langle . \rangle$ refers to an averaging over all tag positions i. The decay of Γ for large values values of Δs is well captured by an exponential decay as illustrated in Fig.5c. The decay length ξ in wild-type cell lines of HCT116 and IMR90 is approximately 350 ± 30 kb and 300 ± 30 kb respectively. However, for the auxin treated HCT116 cell line Γ shows a strongly diminished decay length of about 130 ± 30 kb. The uncertainty on the numerical values of ξ arises from its dependency on r_c chosen between 0.1 μm and 0.6 μm. Although quantitatively ξ depends on the choice of r_c, the qualitative picture remains unaltered, untreated cell lines show correlations across larger genomic distances compared to the auxin treated cell line. This rapid loss of correlation in the auxin treated case might reflect microdomains of smaller sizes as compared to the untreated cell lines. This result aligns with the interval of validity of the two-phase model discussed in Section II C and provides a rough estimate of the microdomain sizes. The shorter decay length in the auxin-treated cells suggests that the depletion of cohesin-mediated loops leads to a more compact domain structure, corroborating the findings from the analysis of the mean squared experimental radius $\overline{R_{human}}$.

III. DISCUSSION

A. Microphase separation

The organization of interphase chromosomes is believed to be driven by the combined effect of two main distinct processes[22]: one based on affinity-mediated interactions and another relying on active (i.e. ATP-dependent) processes such as loop extrusion. Chromatin folding models based on the first type of processes have been proposed as due to binders bridging between different chromatin sections[23] (see Fig. 6(a)), due

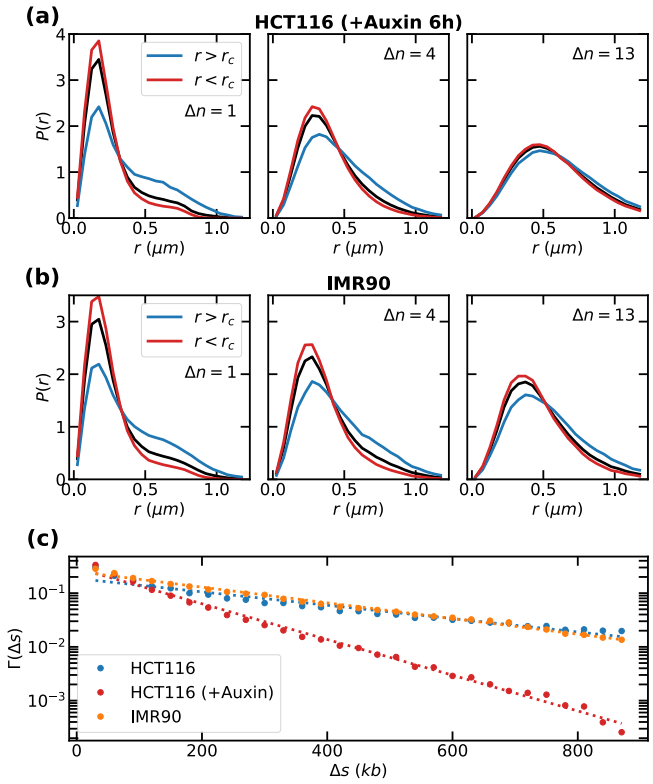

FIG. 5. Non-Markovianity analysis of chromatin configuration in HCT116, auxin-treated HCT116, and IMR90 cell lines. (a) and (b) probability density distributions of Auxin -treated HCT116 and IMR90 $P_{i,i+\Delta n}(r)$ averaged over all tag positions. The black lines represent the total distance distributions. The red and blue curves depict the conditional probability distributions based on the step $(i-1,i)$, where the blue curve shows $P_{i,i+\Delta n}(r \mid r_{i-1,i} > r_c)$ and the red curve represents $P_{i,i+\Delta n}(r \mid r_{i-1,i} < r_c)$, with r_c as a cutoff distance. This demonstrates that larger (smaller) $r_{i-1,i}$ correlates with larger (smaller) $r_{i,i+\Delta n}$, indicating the formation of microdomains (α- and β-phases). (c) linear-log plot of the mean squared deviation $\Gamma(\Delta s)$ between the two conditional distributions.

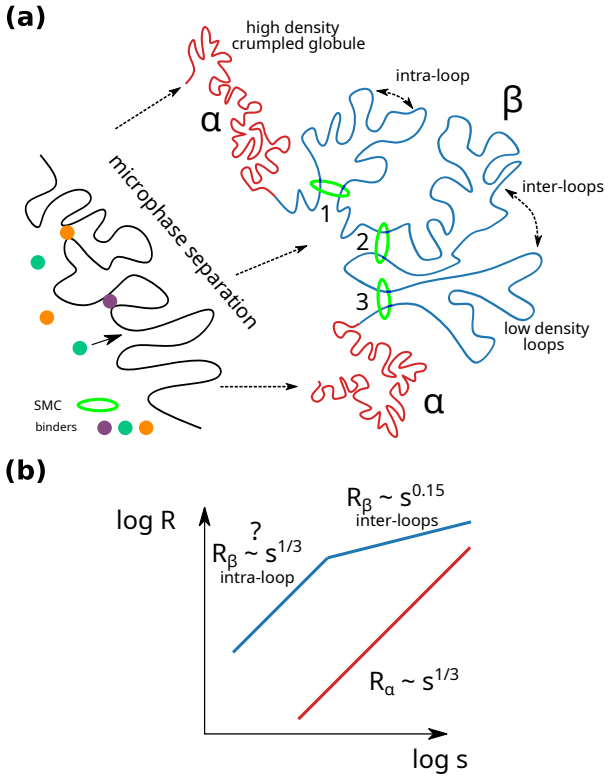

FIG. 6. (a) Représentation schématique de la micro-séparation de phase. La fibre de chromatine, sous l'effet probable de liaisons passives et d'extrudeurs de boucles actifs (complexes protéiques SMC), subit une séparation de phase en microdomaines, avec des phases α et β représentées respectivement en rouge et en bleu. (b) Vue schématique de la mise à l'échelle observée et attendue des rayons caractéristiques R_α et R_β des deux phases en fonction de la distance génomique s. Pour la phase α, les données sont cohérentes avec l'échelle d'un globule enchevêtré, voir Eq. (8). Pour la phase β, nous observons un très faible exposant sur des distances intermédiaires, compatible avec un domaine bouclé. L'échelle de $R_\beta(s)$ varie selon les distances intra-boucles et inter-boucles, comme proposé dans l'Eq. 9.

to diffusing transcription factors[24] or segregation as in block-copolymers[25]. Active loop extrusion[26,27] is known to originate from the action of motor-like protein complexes of the SMC group, such as cohesin and condensin[22]. The emerging picture of chromatin organization as obtained from the analysis of m-FISH data discussed in Sec. II is shown schematically in Fig. 6(a). This figure illustrates the phase separation process resulting in alternating α and β domains shown in red and blue, respectively, where presumably both affinity-mediated and active processes take part. We expect, though, that loop extrusion is a key factor in the shaping of β domains.

Experimental m-FISH data for the α phase indicate a scaling consistent with a crumpled globule[15]

$$R_\alpha(s) \sim A_\alpha s^{1/3} \tag{8}$$

over a broad range of genomic scales, see Fig. 2. The situation is more complex for the phase β. In a phase with multiple loops one expects at least two distinct scaling regimes: an intra-loop regime ($s < s_0$) and an inter-loops regime ($s > s_0$) with s_0 the characteristic loop length, see Fig. 6(a). The effect of confinement should manifest itself at scales beyond the loop length-scale s_0. The mouse data in Fig. 2 show indeed a signature of the onset of a different scaling regime at lengths below $s_0 \approx 10$ kb for the phase β. This value of s_0 could be used as estimated average loop size. Although it is not possible to extract a reliable exponent in the low s regime from the current m-FISH data, one plausible scenario is that loops are formed by crumpled polymer segments. This would suggest the following scaling behavior

$$R_\beta(s) \sim \begin{cases} A_\beta s^{1/3} & s < s_0 \\ s^{\nu_\beta} & s > s_0 \end{cases} \tag{9}$$

as schematically illustrated in Fig. 6(b). The crumpling in the phase β produces polymer conformations with lower density than in the α phase, implying $A_\beta > A_\alpha$. We argue on the basis

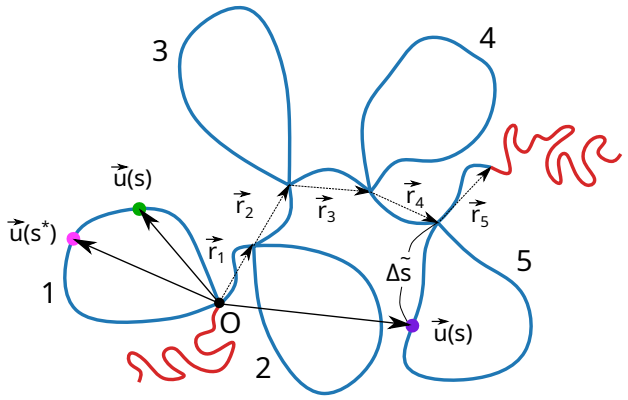

FIG. 7. Ideal polymer (random walk) model of chromatin organization formed by an alternated sequence of α (red) and β (blue) phases. While the α phase is homogeneous, the β phase consist of loops and alternating with short segments. We use different "monomer" sizes for the two phases with $a_\alpha < a_\beta$ to mimick the higher density of the former. Starting from a reference point in the phase β the model predicts three regimes for the mean squared radius (11): intra-loop, inter-loop and multi-domain as described by Eqs. (12), (13) and (14), respectively.

of a random walk model that the scaling $\nu_\beta \approx 0.15$ for $s > s_0$ is due to multiple loops originating from slightly displaced locations indicated as "1", "2", "3" in Fig. 6(a). Such displacement does not produce perfect localization which would result in $\nu_\beta = 0$.

B. A heterogeneous random walk model

It is instructive to consider a simple random walk model of the above chromatin structure. Although such model will not be able to predict the correct scaling exponents (crumpled globule), we believe it can shed some light on the behavior of the different scaling regimes in the two phases. We build such a model using the Gaussian propagator for a segment of chromatin of length l (measured in base-pairs)

$$p_i(\vec{r},l) = \left(\frac{1}{2\pi l a_i^2}\right)^{3/2} e^{-3\vec{r}^2/2la_i^2} \qquad (10)$$

where $i = \{\alpha,\beta\}$ labels either of the two phases and $\vec{r}$ is the vector connecting the two end-points of the segment. We set $a_\alpha < a_\beta$ to generate a denser polymer conformation in the α phase (see Fig. 7). We are interested in the calculation of the average square distance from a given reference location s^*, which is defined as

$$R^2(\Delta s, s^*) = \left\langle (\vec{u}(s) - \vec{u}(s^*))^2 \right\rangle \qquad (11)$$

where $\vec{u}(s)$ denotes the position vector for the genomic site s (where we use the vertex of the first loop as origin, see Fig. 7), $\Delta s \equiv s - s^*$ and s^* can be either located in the α or in the β phase. We note that as the system is not translationally invariant R^2 does not depend only on Δs, but also on the initial point s^*. In what follows, we omit the explicit dependence on s^* in Eq. (11) for brevity.

The phase α is assumed to be of total length L_α, unstructured and thus simply described by the propagator $p_\alpha(\vec{r},l)$ as in (10). To simplify the analysis we consider a β phase composed by loops of fixed length L, whose origins are separated by segments of length n. A sequence of m loops with interdispersed segments generates a phase β of length $L_\beta = m(L+n)$. We start by considering s^* in the mid-point of the first loop and $0 \le \Delta s \le L/2$, i.e. an intra-loop distance. A simple calculation gives

$$R^2(\Delta s) = a_\beta^2 \frac{\Delta s(L-\Delta s)}{L} \qquad (12)$$

as it can be simply obtained from the product of two propagators $p_\beta(\vec{r},\Delta s)p_\beta(\vec{r},L-\Delta s)$ with $\vec{r} \equiv \vec{u}(s) - \vec{u}(s^*)$. For short distances $\Delta s \ll L$ one recovers the Brownian motion limit $R^2(\Delta s) \approx a_\beta^2 \Delta s$. For $\vec{u}(s) = 0$, corresponding to the point O in Fig. 7, we get $R^2(L/2) = \langle \vec{u}^2(s^*)\rangle = a_\beta^2 L/4$. We consider now s^* as in the previous case and s located in the k-th loop ($k > 1$). Being in distinct loops $\vec{u}(s)$ and $\vec{u}(s^*)$ are statistically independent, hence

$$\begin{aligned} R^2(\Delta s) &= \left\langle \vec{u}^2(s)\right\rangle + \left\langle \vec{u}^2(s^*)\right\rangle \\ &= a_\beta^2 \left[\frac{L}{4} + (k-1)n + \frac{\Delta\tilde{s}(L-\Delta\tilde{s})}{L}\right] \end{aligned} \qquad (13)$$

where $\Delta\tilde{s} \le L$ is the length of the segment connecting the point s on the k-th loop with its vertex, see Fig. 7 and where we used $\langle \vec{r}_i^2 \rangle = a_\beta^2 n$. We note that the characteristic radius grows with increasing Δs mainly via the contribution of inter-loop segments $\vec{r}_i$, as the loop contribution which is proportional to $\Delta\tilde{s}(L-\Delta\tilde{s})$ remains bounded. If the length of the inter-loop segments is small compared to the length of the loops (i.e. $n \ll L$), then $R(\Delta s)$ grows very weakly in the inter-loop region. In the analysis of the experimental data we estimated a scaling exponent ≈ 0.15, but this is more likely a crossover region. Finally at even longer genomic length scales Δs comprises several combined α and β domains. In the random walk model the contributions to R^2 for the two domains are independent. The contribution of one α and of one β domain together is

$$R^2(L_\alpha + L_\beta) = a_\alpha^2 L_\alpha + a_\beta^2 mn \qquad (14)$$

where in the phase β only the m inter-loop segments (of length na_β) contribute.

Summarizing, starting from a point in the phase β, three distinct regimes govern the scaling of the mean-square radius with the genomic distance Δs. In the intra-loop regime, described by Eq. (12), the radius grows as $R(\Delta s) \sim \Delta s^{1/2}$ for distances much smaller than the loop length. The inter-loop regime (Eq. (13)) is characterized by an average very weak growth with Δs. Here $R^2(\Delta s)$ has a term which scales linearly with $k-1$ the number of inter-loop segments crossed, however, the constant term proportional to $L/4$ dominates. Finally, for Δs crossing several α and β domains the radius is a multiple of (14). This behavior is shown in Fig. 8 which

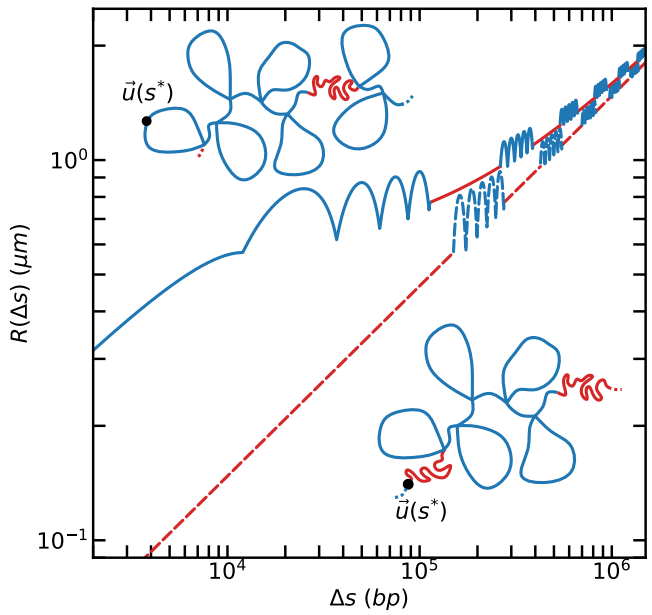

FIG. 8. Root mean squared distance (rmsd) of the heterogeneous random walk model as a function of distance for two distinct starting loci. The dashed curve shows the rmsd obtained starting from the compact α-phase (red). The solid lines show the rmsd starting from an interdispersed looped β domain (blue). For the calculation we assumed a heterogeneous random walk with alternating α and β domains of 150 kb and 125 kb each. Every β domain is comprised of 5 loops of 23 kb and interdispersed unlooped walks of 1 kb. Values of a_α and a_β were chosen to be 0.08 μm and 0.4 μm respectively.

plots $R(\Delta s)$ calculated from a point s^* either located within the α (dashed) or the β phase solid). Despite its simplicity the random walk model captures the main features observed in experiments. We note that a more realistic model of chromatin is the fractional brownian motion (fBm), obtained by a chain with pairwise harmonic interactions with strengths decaying algebraically along the chain[3]. Such models can accommodate a crumpled globule behavior. Moreover, they can be used to compute scaling behavior of different topologies, as for instance in looped states (see Ref. 19 for a generalization of (12) to the fBm case). In principle such models can handle multiple loops, but it seems difficult to incorporate a phase-separating behavior in the fBm formalism of Ref. 3.

C. Concluding remarks

In conclusion, we have discussed how high resolution microscopy m-FISH data reveal a complex and heterogeneous organization of chromatin structure encompassing different scaling laws across various genomic scales. Such complex organization was emphasized by several authors in the past years[19,28–33], but we are not aware of observation of distinct scaling laws in a phase separating system as done in this work. Here we showed how distinct phases follow very naturally from the analysis of the distance distribution of m-FISH tagged sites. We showed that Human[9] an Mouse[16] cell lines provide consistent and complementary results spanning over almost three order of magnitudes across the genomes and interpreted the data using a simple heterogeneous random walk model. The m-FISH technique, relying on high resolution fluorescence imaging, is only relatively new and only limited fractions of genomes have been so far analyzed. It would be interesting to correlate scaling behavior with biological details of the tagged genomic region (actively transcribed genes vs. silent ones). Some studies have started to explore these correlations[5], but more data are needed to establish robust correlations. Another aspect not yet understood is the role of dynamics in the system. The m-FISH method (as the Hi-C) provides a snapshot of the chromatin conformation at a given time. We have seen that a micro-phase separation takes place in the system. One possibility is that the α and β domains are rather dynamical so that the cell constantly swaps between these two in the course of time, rather than being a frozen structure. Future studies will likely elucidate these issues.

SUPPLEMENTARY MATERIAL

The Supplementary Material contains a description of the experimental data and additional figures.

ACKNOWLEDGMENTS

Discussions with M. Bardou, M. Nollmann, J. Palmeri and J-C. Walter are gratefully acknowledged. LR acknowledges doctoral thesis support from the LabMUSE EpiGenMed within the I-Site MUSE (ANR-16-IDEX-0006) and from the *Agence Nationale de la Recherche* (TRANSLAxon project ANR-20-CE16-0025). MS acknowledges financial support from Fonds Wetenschappelijk Onderzoek (FWO 11O2323N). This project was supported by the LabEx NUMEV (ANR-10-LABX-0020) within the I-Site MUSE (ANR-16-IDEX-0006).

DATA AVAILABILITY STATEMENT

Microscopy data of human cell lines studied in this paper have been downloaded from one of Bintu et al[9] authors' Github repository. Microscopy data of mouse liver cells of Liu et al[16] have been downloaded from Siyuan Wang's lab webpage..

[1] W. A. Bickmore, "The spatial organization of the human genome," Annu. Rev. Genomics Hum. Genet. **14**, 67–84 (2013).
[2] J. Dekker and L. Mirny, "The 3d genome as moderator of chromosomal communication," Cell **164**, 1110–1121 (2016).
[3] K. Polovnikov, S. Nechaev, and M. Tamm, "Effective hamiltonian of topologically stabilized polymer states," Soft Matter **14**, 6561–6570 (2018).
[4] A. Lesage, V. Dahirel, J.-M. Victor, and M. Barbi, "Polymer coil-globule phase transition is a universal folding principle of drosophila epigenetic domains," Epigenetics & Chromatin **12**, 28 (2019).
[5] S. M. Espinola, M. Götz, M. Bellec, O. Messina, J.-B. Fiche, C. Houbron, M. Dejean, I. Reim, A. M. Cardozo Gizzi, M. Lagha, *et al.*, "Cis-regulatory chromatin loops arise before tads and gene activation, and are independent

of cell fate during early drosophila development," Nature Gen. **53**, 477–486 (2021).

[6]M. Conte, E. Irani, A. M. Chiariello, A. Abraham, S. Bianco, A. Esposito, and M. Nicodemi, "Loop-extrusion and polymer phase-separation can co-exist at the single-molecule level to shape chromatin folding," Nature Comm. **13**, 4070 (2022).

[7]S. A. Belan and D. E. Starkov, "Influence of active loop extrusion on the statistics of triple contacts in the model of interphase chromosomes," JETP Letters **115**, 763–769 (2022).

[8]M. Liefsoens, T. Földes, and M. Barbi, "Spectral-based detection of chromatin loops in multiplexed super-resolution FISH data," bioRxiv **2023.10.24.563731**, 2023–10 (2023).

[9]B. Bintu, L. J. Mateo, J.-H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang, "Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells," Science **362**, eaau1783 (2018).

[10]L. J. Mateo, S. E. Murphy, A. Hafner, I. S. Cinquini, C. A. Walker, and A. N. Boettiger, "Visualizing DNA folding and RNA in embryos at single-cell resolution," Nature **568**, 49–54 (2019).

[11]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.*, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," Science **326**, 289–293 (2009).

[12]A. M. C. Gizzi, D. I. Cattoni, J.-B. Fiche, S. M. Espinola, J. Gurgo, O. Messina, C. Houbron, Y. Ogiyama, G. L. Papadopoulos, G. Cavalli, *et al.*, "Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms," Mol. Cell **74**, 212–222 (2019).

[13]J. Gurgo, J.-C. Walter, J.-B. Fiche, C. Houbron, M. Schaeffer, G. Cavalli, F. Bantignies, and M. Nollmann, "Multiplexed chromatin imaging reveals predominantly pairwise long-range coordination between drosophila polycomb genes," Cell Reports **43** (2024), 10.1016/j.celrep.2024.114167.

[14]L. Remini, M. Segers, J. Palmeri, J.-C. Walter, A. Parmeggiani, and E. Carlon, "Chromatin structure from high resolution microscopy: scaling laws and microphase separation," Phys. Rev. E **109**, 024408 (2024).

[15]A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, "The role of topological constraints in the kinetics of collapse of macromolecules," J. Phys. France **49**, 2095–2100 (1988).

[16]M. Liu, Y. Lu, B. Yang, Y. Chen, J. S. Radda, M. Hu, S. G. Katz, and S. Wang, "Multiplexed imaging of nucleome architectures in single cells of mammalian tissue," Nature Comm. **11**, 2907 (2020).

[17]L. A. Mirny, "The fractal globule as a model of chromatin architecture in the cell," Chromosome Res. **19**, 37–51 (2011).

[18]F. Grubert, R. Srivas, D. Spacek, M. Kasowski, M. Ruiz-Velasco, N. Sinnott-Armstrong, P. Greenside, A. Narasimha, Q. Liu, B. Geller, A. Sanghi, M. Kulik, S. Sa, M. Rabinovitch, A. Kundaje, S. Dalton, J. B. Zaugg, and M. Snyder, "Landscape of cohesin-mediated chromatin loops in the human genome," Nature **583**, 737–743 (2020).

[19]K. E. Polovnikov, B. Slavov, S. Belan, M. Imakaev, H. B. Brandão, and L. A. Mirny, "Crumpled polymer with loops recapitulates key features of chromosome organization," Phys. Rev. X **13**, 041029 (2023).

[20]D. Starkov and S. Belan, "Effect of active loop extrusion on the two-contact correlations in the interphase chromosome," The Journal of Chemical Physics **161** (2024), 10.1063/5.0221933.

[21]S. Belan and V. Parfenyev, "Footprints of loop extrusion in statistics of intra-chromosomal distances: An analytically solvable model," The Journal of Chemical Physics **160** (2024), 10.1063/5.0199573.

[22]L. A. Mirny, "Chromosome and protein folding: In search for unified principles," Curr. Opin. Struct. Biol. **81**, 102610 (2023).

[23]M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, "Complexity of chromatin folding is captured by the strings and binders switch model," Proc. Natl. Acad. Sci. **109**, 16173 (2012).

[24]C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo, "Nonspecific bridging-induced attraction drives clustering of dna-binding proteins and genome organization," Proc. Natl. Acad. Sci. **110**, E3605 (2013).

[25]D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, "Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains," Nucl. Acids Res. **42**, 9553 (2014).

[26]E. Alipour and J. F. Marko, "Self-organization of domain structures by DNA-loop-extruding enzymes," Nucl. Acids Res. **40**, 11202–11212 (2012).

[27]A. Goloborodko, J. F. Marko, and L. A. Mirny, "Chromosome compaction by active loop extrusion," Biophys. J. **110**, 2162–2168 (2016).

[28]V. F. Scolari and M. C. Lagomarsino, "Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid," Soft Matter **11**, 1677–1687 (2015).

[29]T. J. Stevens, D. Lando, S. Basu, L. P. Atkinson, Y. Cao, S. F. Lee, M. Leeb, K. J. Wohlfahrt, W. Boucher, A. O'Shaughnessy-Kirwan, *et al.*, "3d structures of individual mammalian genomes studied by single-cell hi-c," Nature **544**, 59–64 (2017).

[30]A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, and N. Gilbert, "Polymer simulations of heteromorphic chromatin predict the 3d folding of complex genomic loci," Mol. Cell **72**, 786–797 (2018).

[31]E. H. Finn, G. Pegoraro, H. B. Brandão, A.-L. Valton, M. E. Oomen, J. Dekker, L. Mirny, and T. Misteli, "Extensive heterogeneity and intrinsic variation in spatial genome organization," Cell **176**, 1502–1515 (2019).

[32]G. Shi and D. Thirumalai, "Conformational heterogeneity in human interphase chromosome organization reconciles the fish and hi-c paradox," Nature Comm. **10**, 3894 (2019).

[33]M. Conte, L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, and M. Nicodemi, "Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation," Nature Comm. **11**, 3289 (2020).