Chromatin structure from high resolution microscopy: scaling laws and microphase separation

Loucif Remini, Midas Segers, John Palmeri, Jean-Charles Walter, Andrea Parmeggiani, Enrico Carlon

Chromatin structure from high resolution microscopy: scaling laws and microphase separation

Loucif Remini,[1] Midas Segers,[2] John Palmeri,[1] Jean-Charles Walter,[1] Andrea Parmeggiani,[1] and Enrico Carlon[2]

[1] *Laboratoire Charles Coulomb (L2C), Univ. Montpellier, CNRS UMR5221, Montpellier, France*
[2] *Soft Matter and Biophysics, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium*

(Dated: November 24, 2023)

Recent advances in experimental fluorescence microscopy allow high accuracy determination (resolution of 50 nm) of the 3D physical location of multiple (up to $\sim 10^2$) tagged regions of the chromosome. We investigate publicly available microscopy data for two loci of the human Chr.21 obtained from multiplexed FISH methods for different cell lines and treatments. Inspired by polymer physics models, our analysis centers around distance distributions between different tags with the aim being to unravel the chromatin conformational arrangements. We show that for any specific genomic site, there are (at least) two different conformational arrangements of chromatin, implying coexisting distinct topologies which we refer to as phase α and phase β. These two phases show different scaling behaviors: the former is consistent with a crumpled globule while the latter indicates a confined, but more extended conformation, such as a looped domain. The identification of these distinct phases sheds light on the coexistence of multiple chromatin topologies and provides insights into the effects of cellular context and/or treatments on chromatin structure.

I. INTRODUCTION

Understanding the organization of eukaryotic chromosomes is an issue of broad interest and has been intensively studied both experimentally [1–8] and computationally [9–21]. This interest stems from the fact that the 3D chromatin structure has a strong influence on several genomic processes such as transcription or replication [22, 23]. Long DNA molecules go through high compaction when chromosomes condense during mitosis. During the interphase, genes are actively transcribed and chromosomes are unpacked and distributed throughout the cell nucleus but still possess a remarkable level of spatial organization. This organization is hierarchical, involving different structures at different genomic length scales. Starting from whole chromosome lengths ($\sim$ 100 Mb) down to $\sim$ 100 kb the DNA organization can be summarized as follows: 1) During the interphase, different chromosomes do not mix but occupy well-defined territories inside the cell nucleus [24]. 2) Each chromosome is split into several stretches of gene-rich domains (euchromatin) alternating with gene-poor domains (heterochromatin). Euchromatin tends to be more open than heterochromatin. Moreover stretches of the same chromatin species aggregate forming so-called A and B compartments. 3) At scales of $\sim$ 1 Mb and below the chromatin is assembled in Topologically Associating Domains (TADs) [25, 26], which are regions where chromatin interacts more frequently. The biological function of TADs and the mechanisms leading to their origin are still debated. A prevalent line of thoughts suggests that TADs originate from a loop extrusion mechanism [27], although it is likely that different mechanisms operate simultaneously to contribute to the complex hierarchical organization of eukaryotic chromatin [28, 29].

Advanced techniques such as Chromosome Conformational Capture (3C) [1], particularly its high-throughput sequencing (Hi-C) version [2], have played a central role in unveiling the chromatin organization. Hi-C provides genome-wide data for contact probabilities between pairs of genomic sites and was instrumental for understanding the folding of chromosomes at different length scales. Although it is by no doubt a very powerful technique, Hi-C has also some limitations. For instance, it provides indirect information on physical distances (via contact probabilities), it typically requires the averaging over a large number of chromosomes ($\sim 10^6$) and may contain some systematic biases [30, 31], therefore raw data require normalization and other pre-processing steps. Fluorescence in situ hybridization (FISH, Fig. 1) has also been used to investigate chromosomal structure [32]. FISH uses fluorescent probes that are complementary to desired chromosomal regions. After hybridization, the 3D location of the probes can be determined from fluorescence microscopy. For a long time FISH remained a

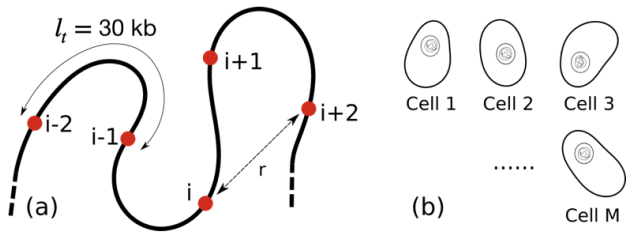

FIG. 1. Setup of multiplexd FISH as developed in Ref. [4]. (a) $1 \le i \le N$ genomic sites equally spaced are targeted by FISH probes during sequential hybridization cycles. In the setup of [4] the genomic distance between consecutive tags was $l_t = 30$ kb. The 3D spatial coordinates of the tagged sites are then read by high-resolution fluorescence microscopy. (b) The experiment is repeated over M distinct copies of chromosomes. The values of N and M for the experimental data used in this paper are given in Table I. We analyzed here the distance distributions for any two given sites i and j.

very low throughput technique, the main issue being the limited number of fluorescent probes with different emission spectra available so that only a few sites could be visualized. However, quite recently, a multiplexed FISH (mFISH) method was developed [4]. Using unique readout sequences and sequential hybridization cycles, the method allows one to determine the positions of $N \sim 50 - 100$ chromosomal sites with high resolution (< 50 nm), see Fig. 1. Measurements are done on samples containing up to $M \sim 10^4$ distinct chromosomes. Although, unlike Hi-C, it is not a genome-wide technique, m-FISH data reproduce well the Hi-C contact probabilities for the same genomic regions [4]. In addition, m-FISH reveals cooperative (i.e. many-body) chromatin interactions [4, 6, 8]. These interactions cannot be inferred from Hi-C pairwise contacts, however polymer physics-based models have recently been developed to describe them [33, 34].

The aim of this paper is to analyze a set of m-FISH data focusing on distance distributions of any two pairs of labels i and j. In particular, we analyze the experimental probability distribution $P_{ij}(r)$ of distance r between tags i and j. The large number of distinct samples ($M \sim 10^4$) implies good statistics, therefore these histograms can be accurately determined. Our analysis reveals that there are (at least) two different modes of chromatin organization at any given genomic site. We report scaling laws characterizing the growth of the chromatin domains when increasing the genomic distance. Overall, our work presents a polymer physics-inspired analysis to characterize chromatin structure. Importantly, this method is straightforward to implement and can be applied to "raw" experimental data.

The paper is organized as follows. Section II presents the method we propose based on the analysis of the experimental distance distribution. Section III focuses on the analysis of the extracted scaling behaviour of typical configuration radii from different cell lines. Section IV discusses, using polymer physics models, the regularities of typical chromatin conformations emerging from the experimental data. Appendices provide technical details on the polymer models considered, as well as a full analysis of the whole distance probability distribution data set.

TABLE I. Summary of experimental data from [4] targeting two different regions on chromosome 21. One region is 2 Mb long, located between the $28 - 30$ Mb position on the chromosome, and was studied in lung fibroblast (IMR90) and erythroleukemia (K562) cell lines. The other region is about 2.5 Mb in size, located between the $34 - 37$ Mb position, and was investigated in a colon cancer cell line (HCT116). Tags in the study are spaced apart by a genomic distance of 30 kb. N is the number of equally spaced genomic sites and M is the number of distinct copies of chromosomes for which the experiments were repeated.

n.	Cell line	Chromosomic region	N	M
1	HCT116 (untreated)	chr21 [34-37] Mb	83	11631
2	HCT116 (+Auxin 6h)	chr21 [34-37] Mb	83	9526
3	IMR90	chr21 [28-30] Mb	65	4871
4	K562	chr21 [28-30] Mb	65	13997

II. DISTANCE PROBABILITY DISTRIBUTIONS

Table I summarizes the main features of the four data sets from [4] analyzed in this paper. These experiments target two different loci of human chr. 21: a 2 Mb region (Chr21:28Mb-30Mb) labeled by $N = 65$ tags and a 2.5 Mb region (Chr21:34.6Mb-37.1Mb) labeled by $N = 83$ tags. These regions contain several TADs across multiple cell lines [4]. More data are available in [4], but these 4 sets were selected because they contain a large number of independent measurements, ranging from $M = 4871$ to $M = 13997$, (Table I) from which accurate distance probability distributions are obtained. For an ideal polymer, the distance probability distribution is a Gaussian that depends only on the length $r = |\vec{r}|$ of the vector distance $\vec{r}$:

$$P_{ij}(r) = g(r; R) \equiv 4\pi r^2 \left(\frac{3}{2\pi R^2}\right)^{3/2} e^{-3r^2/(2R^2)} \quad (1)$$

where the term $4\pi r^2$ comes from the Jacobian transformation of the distribution. The above distribution is normalized and contains the mean-squared radius $R^2 = \langle \vec{r}^2 \rangle$ as a single parameter. Despite its simplicity the Gaussian chain model is a reasonable approximation to complex phases of polymeric matter, such as polymer melts. A well-known result of polymer physics is that a single test chain immersed in a melt of other chains behaves as an ideal polymer [35]. This is because self-avoidance gets screened by the surrounding polymers. The mean square radius of an ideal polymer scales as

$$R \sim |i - j|^{1/2}, \quad (2)$$

where the exponent $1/2$ describes the universal scaling behavior of random walks. It has been recently shown [36] that the introduction of suitable monomer-monomer harmonic pairwise interactions in a bead and spring polymer model leads to an equilibrium distance distribution corresponding to (1), but with a scaling described by

$$R \sim |i - j|^{\nu}. \quad (3)$$

By tuning the monomer-monomer interactions, one can generate any values of the exponent in the interval $1/3 \leq \nu \leq 1/2$ [36]. The statistical properties of the model are those of a fractional Brownian motion (fBm), which was invoked as a model for chromosome arrangement in cells [37, 38], as well as for other problems of polymer dynamics [39–41]. We note that the case $\nu = 1/3$ is the exponent of a crumpled globule [42], a phase discussed in some more detail in Sec. IV. Therefore, the Gaussian

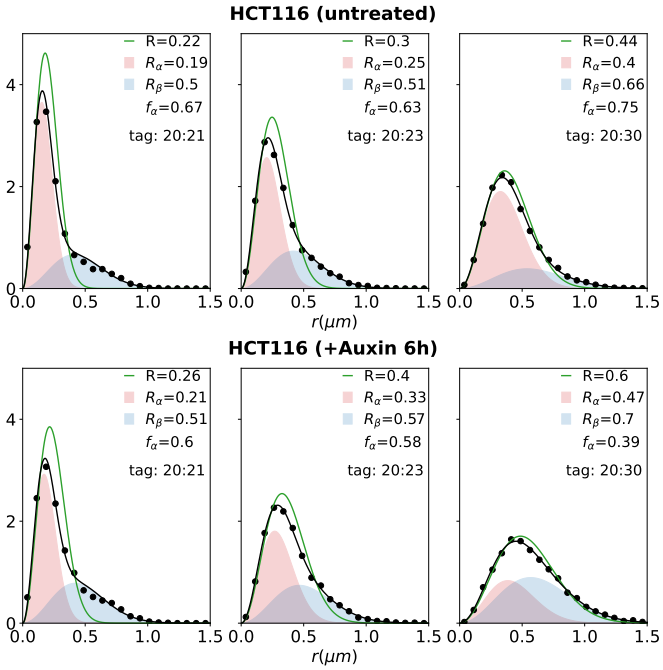

FIG. 2. Distance probability distributions $P_{ij}(r)$ from experimental data (black circles) for different tag locations i, j for experiments of the HCT116 cell line (sets n.1 and n.2 of Table I). The values for i and j are indicated in each graph. In the two sets, we fix $i = 20$ and $j = 21$, 23 and 30 (from left to right). Solid lines are fits to a single gaussian model (Eq. (1), green line) and double gaussian model (Eq. (4), black line). The legends in each plot show the value of R for a single gaussian model (Eq. (1), green) and of R_α, R_β and f_α for the double gaussian model (Eq. (4)). The red and blue filled areas denote the contribution of the two components of the fit. More results for other cell lines are shown in [43].

distribution (1) encompasses a broad range of polymer models.

Figure 2 shows some plots of experimentally measured distance probability distributions $P_{ij}(r)$ (circles), where the tags i and j considered are reported in each graph and the horizontal scale is the spatial distance in micrometers (μm). The green solid lines are a one parameter fit to the Gaussian model Eq. (1), where the mean-squared radius R is the only fitting parameter We note that the model does not fit well the experimental data and the deviations are particularly strong for close tags, i.e. small $|i-j|$ (say within a range of $|i-j| \approx 5$, corresponding to a genomic distance of 150 kb). As the genomic distance increases, i.e. for larger $|i-j|$, the experimental distributions gets closer to (1), as reflected by the trend that the green lines (Gaussian fits) tend to merge with the black circles (experiments) in Fig. 2.

The probability densities for small $|i-j|$ suggest that experiments describe two coexisting populations (phases), therefore we fitted the data using:

$$P_{ij}(r) = f_\alpha\, g(r; R_\alpha) + (1 - f_\alpha)\, g(r; R_\beta) \qquad (4)$$

where the parameter $0 \leq f_\alpha \leq 1$ (respectively $f_\beta = 1 - f_\alpha$) is the fraction of chromosomes with mean distance R_α (respectively R_β) separating tags i and j. We will refer to the two phases as α and β. The two-phase model (4) provides an excellent fit to the experimental data. These fits are shown as solid black lines in Fig. 2. The colored areas are the contributions of the two gaussian components to the final fit.

The two-phase model is fitted to the data using R_α, R_β and f_α as adjustable parameters [43]. These three values are reported in the legends of Fig. 2. The radii change with changing i and j and, as expected, increase for increasing $|i-j|$. The growth rate is, however, different for R_α and R_β, as discussed in the next Section. Interestingly, the two phase model (4) was also considered in a recent analysis of Hi-C chromosomal data [18]. Such a model was introduced to rationalize some observed discrepancies between Hi-C and earlier non-multiplexed FISH data. We examine the relationship between the findings presented in Ref. [18] and our own results, along with other relevant numerical and theoretical studies, in the concluding Section of this paper.

III. SCALING BEHAVIOR OF R_α AND R_β

Figure 3 shows the scaling behavior of the HCT116 cell line. These are log-log plots of R_α and R_β vs. the genomic distance $l_t|i-j|$ for fixed i and varying j, where $l_t = 30$ kb is the spacing in base pairs between two consecutive tags of the experiments (see Fig. 1). The horizontal axis is thus the genomic distance measured in bp, which extends up to about 2.5 Mb. We characterized the scaling behavior by fitting R_α and R_β to (3) for close tags, i.e. $|i-j| \leq 5$ corresponding to a maximal genomic distance of 150 kb. This choice is dictated by two factors: (i) we observe that the scaling of R_α and R_β follows an approximate power-law behavior for a limited distance between tags and (ii) the values of R_α and R_β are more accurately determined from fitting distance probability distributions for small $|i-j|$ as well. This is because there is typically a bigger gap between R_α and R_β for small $|i-j|$, therefore they can be more reliably extracted from the data analysis.

Table II gives a summary of the average exponents ν_α and ν_β for the four different cell lines averaged over all the tags. With the exception of the K562 cell line, we find consistently $\nu_\alpha > \nu_\beta$. Comparing the two HCT116 cell lines we find very little difference between the exponents in the two cases, apart from a possibly slightly higher ν_α in the auxin-treated case, see Table II. This small increase is a signature that the α phase is slightly less compact in the auxin-treated case, consistent with the fact that by degrading cohesin, auxin releases some constraints. The little overall difference in exponents in the two cases is coherent with the conclusions of Bintu et al. [4]. The authors found that the *ensemble averaged* contact matrices from mFISH show no sign of TADs in

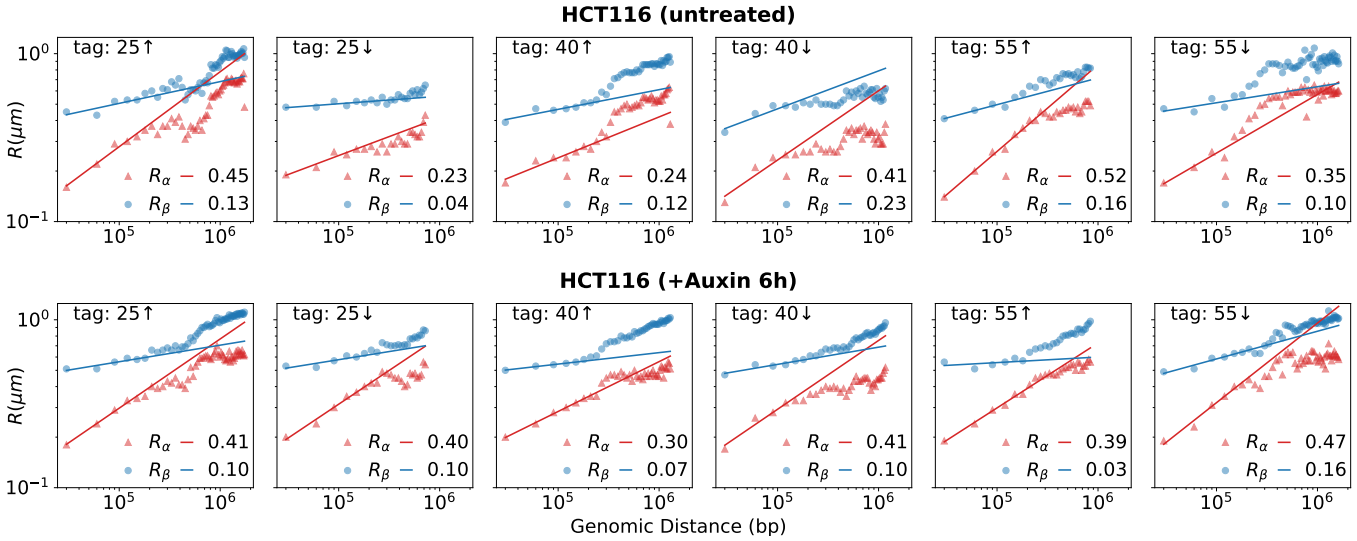

FIG. 3. Log-log plots of scaling behavior R_α (red triangles) and R_β (blue circles) vs. genomic distance, in bases, for some selected tags of the HCT116 cell. Top: untreated and bottom: 6h-auxin treated cells. The reference tag used is given in each plot. The notation "tag: $40\uparrow$" means that R_α and R_β are calculated for $i = 40$ and $j = 41, 42 \ldots N$ and "tag: $40\downarrow$" plots the data for $i = 40$ and $j = 39, 38, \ldots 1$. The solid lines are power-law fits over the first 5 points of each set.

the auxin-treated cells. However, TADs were still visible at the *single cell* level. As a matter of fact, as explained in [4], while TAD boundaries in the untreated HCT116 cells are pinned at specific genomic sites, in the auxin-treated line, they are found in different genomic positions for each cell, thereby washing out any TAD signature in the ensemble averaged Hi-C data. The effect of pinned vs. variable boundaries is also visible in Fig. 3: the scaling of R_α and R_β is smoother and power-laws extend to a broader interval in the auxin-treated cells as compared to the untreated ones. In the latter, several "bumps", which are likely the effect of strongly pinned domains, are visible.

Similar scaling behavior is found for the IMR90 lung fibroblast cell line [43], with values of ν_α and ν_β consistent with those observed in the untreated HCT116 cell lines, see Table II. An exception is the K562 erythroleukemia cell line for which we observe substantially no differences between exponents ν_α and ν_β (Table II). This suggests a different type of chromatin organization for K562 as compared to HCT116 and IMR90. The origin of the peculiar behavior is presently not understood. In view of the numerical value of ν_β it is more logical to characterize the K562 cell line as having two phases of type α, see Fig. A.1.

Having examined the distinct properties of the two phases thus far, we shall now shift our focus to the analysis of a global quantity. Figure 4 shows a plot of the total mean radius $R_{\rm T} = \sqrt{\langle r^2 \rangle}$ for fixed tag i and with either $j > i$ ($\uparrow$) or $j < i$ ($\downarrow$). Using Eq. (4) one can express $R_{\rm T}$ as a function of the radii of the two phases as

$$R_{\rm T}^2 = f_\alpha \, R_\alpha^2 + (1 - f_\alpha) \, R_\beta^2, \tag{5}$$

which suggests that the scaling of $R_{\rm T}$ as function of the genomic distance is a weighted combination of R_α and R_β. The previous relation is valid in the range of genomic distances for which the chromatin actually separates into the α and β phases, and it also assumes that there are no other major components besides these two phases. As we have seen from the analysis of distance distributions, we can identify the two phases up to genomic distances of about 150 kb. Figure 4 shows plots of the scaling of $R_{\rm T}$ vs. s for some selected tags of the two HCT116 cell lines. The data indicate that there is a different scaling behavior at different genomic distances. We fitted therefore a power-law scaling

$$R_T \sim s^\nu \tag{6}$$

in two ranges: $s \leq 150$ kb (which corresponds to $|i-j| \leq 5$) and $s \geq 450$ kb. The fits yield two scaling expo-

TABLE II. The first three columns give the mean values of the critical exponents ν_α and ν_β and of f_α averaged over all data in the range $|i - j| \leq 5$ (corresponding to a genomic distance $s \leq 150$ kb) for the different data sets. The analysis of the HCT116 and IMR90 cell lines gives consistent values of these exponents with $\nu_\alpha > \nu_\beta$. For the leukemia cell line K562 $\nu_\alpha \approx \nu_\beta$. The two last columns give the mean values of the scaling exponent of the total mean radius R_T (Eq. (5), refer to Fig. 4 for details) at short range ($s \leq 150$ kb) and at long range ($s \geq 450$ kb) with uncertainties indicated in parentheses, referring to the last digit.

Data set	ν_α	ν_β	f_α	$\nu_{\rm short}$	$\nu_{\rm long}$
HCT116	0.35(7)	0.12(5)	0.64(5)	0.21(4)	0.33(20)
HCT116 (+Auxin)	0.40(3)	0.10(2)	0.60(5)	0.23(2)	0.29(03)
IMR90	0.29(8)	0.11(4)	0.61(8)	0.14(5)	0.32(13)
K562	0.37(8)	0.33(7)	0.53(8)	0.23(5)	0.22(11)

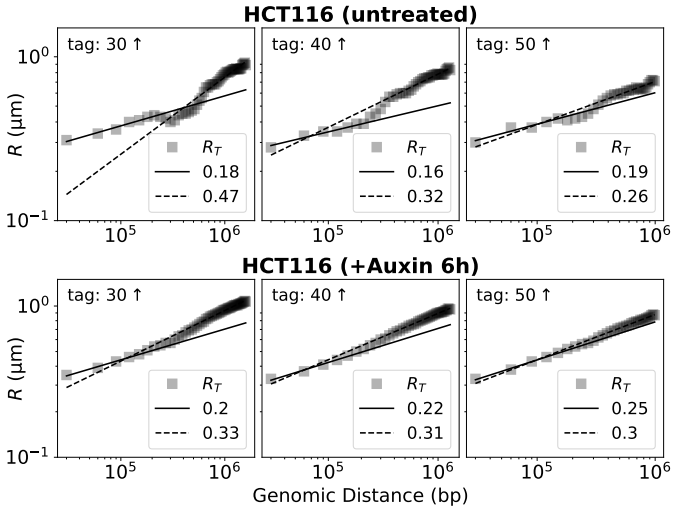

FIG. 4. Log-log plots of scaling behavior of the total mean squared radius $R_{\rm T} = \sqrt{\langle r^2 \rangle}$ from the raw data (black squares). The solid lines are power-law fits (6) over the region $|i-j| \leq 5$, corresponding to genomic distances ≤ 150 kb. The dashed lines are power-law fits of the region ≥ 450 kb. The fits give two distinct exponents and illustrate the short vs. long scale behavior of $R_{\rm T}$. See [43] for more results for other cell lines.

nents, denoted hereafter by $\nu_{\rm short}$ and $\nu_{\rm long}$, which are reported in the last two columns of Table II. Differences in scaling behavior at different genomic distances were also observed and explained in the analysis of Hi-C contact probabilities data [44]. The interpretation of the results of Table II as well as the connections with other studies will be discussed in details in the next Section.

IV. DISCUSSION

In this paper, we have analyzed mFISH data from the experiments reported in Ref. [4] for human chr.21 for different cell lines. Our main result is the evidence of two different coexisting phases of chromatin, which we referred to as the α and β phases. The range of genomic length explored by these experiments goes from 30 kb to 2.5 Mb, but the characteristic radii R_α and R_β can be confidently obtained from data for distances up to approximately 0.5 Mb. Therefore our conclusions are limited to this range of lengths, which is roughly the genomic distance at which TADs are observed. We characterized the two phases using scaling exponents ν_α and ν_β. Interestingly, these exponents are different in the two phases (except for the K562 cell line), indicating substantially different spatial organizations, which we discuss in what follows.

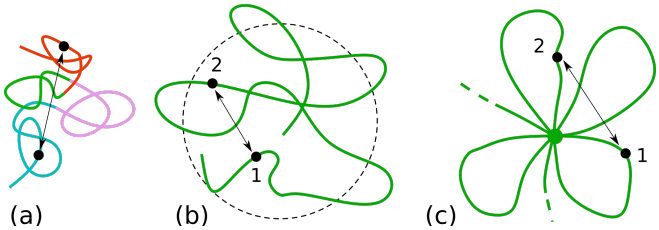

FIG. 5. Possible phases of chromatin: (a) crumpled globule, (b) confined, due to an external potential, and (c) flower. (a) is a candidate for the α phase and (b,c) for the β phase. In (a) different segments of the chromatin fiber are colored differently to illustrate the geometry of the globule. The rapid condensation leads to collapse at the local scale involving vicinal regions. In (b) and (c) strong confinement implies $R_\beta \sim s^0$ for s sufficiently long.

A. α Phase

The exponent we extracted for the α phase is close to that expected for a crumpled globule, which is $\nu = 1/3$ [42, 45, 46]. The crumpled globule is a metastable phase arising from the rapid condensation of a self-attracting polymer following a temperature quench. The polymer condenses, but it has insufficient time to relax to a true equilibrium conformation. The rapid condensation prevents the end points of the long polymer to retract and to form knots, so the crumpled globule remains unentangled [47]. This is different from an equilibrated compact phase, which is strongly entangled and forms knots. The absence of knots makes the unfolding of crumpled globules much more rapid than that of equilibrated conformations [47]. Crumpled globule polymeric phases have been discussed in melts of polymer rings [9] and are believed to be relevant for genome folding structures [47, 48]. Figure 5(a) illustrates the typical conformation of a crumpled globule. Nearby chromatin segments (indicated with different colors in Fig. 5(a)) tend to condense and the process continues hierarchically involving longer and longer length scales, with the end-points playing no specific role in the process.

B. β Phase

The β phase is characterized by a very small exponent, which is an indication of strong confinement. Figure 5(b) and (c) illustrate two possible mechanisms of polymer confinement. In (b) the confinement is due to an external potential forcing the monomers to remain within a certain range from a fixed origin. This could be due to the formation of a droplet by, for instance, liquid-liquid phase separation, with chromatin preferentially absorbed by the droplet. An alternative interpretation is a flower-like conformation (c) where loops, possibly forming via an extrusion mechanism, are bound to a central hub. For both types of conformations (b) and (c) one has $R_\beta \sim s^0$,

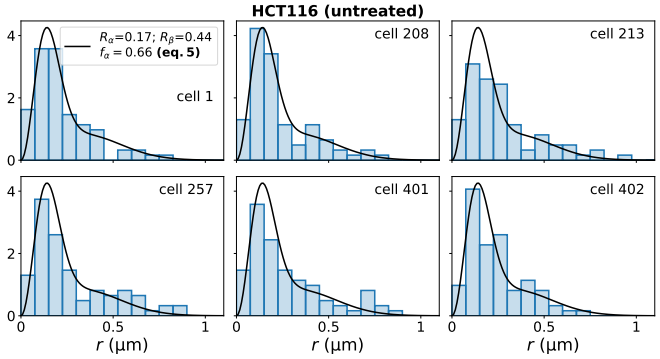

FIG. 6. Distance distributions of consecutive tags ($|i-j|=1$) for individual cells (blue bars) from the HCT116 strain. Similar distributions are obtained for the Auxin treated HCT116 ensemble, as well as other cell-types. Solid black lines correspond to fits of $\langle P_{i,i+1}\rangle_i$ with (4) for the untreated HCT116 cell line. The observed agreement between the single-cell distance distributions and $\langle P_{i,i+1}\rangle_i$ corroborates the existence of at least two distinct topologies of chromatin organization at the single-cell level.

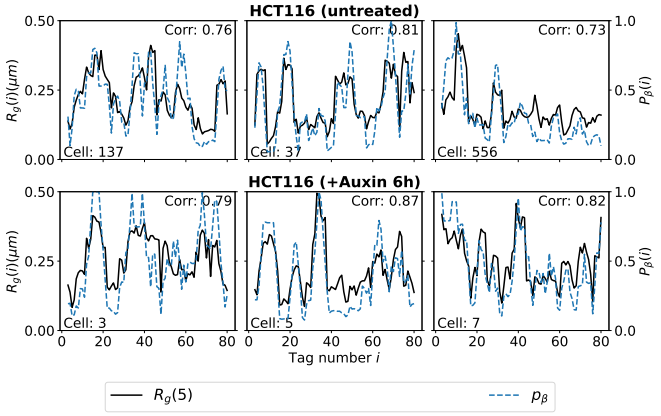

FIG. 7. Radius of gyration R_g (solid black, left y-axis) and β phase probability p_β (dashed blue, right y-axis) as a function of tag number for randomly selected from the HCT116 cell line. R_g is calculated over five subsequent tags, while p_β is defined in Eq. (7). The values in the upper right corner of each panel give the Pearson correlation score between the two quantities.

i.e. the characteristic radius becomes independent of the typical genomic length s beyond some threshold value $s > s_0$. Here s_0 is the typical genomic length of the polymer to reach the droplet surface and return back to the origin in (b) or the typical loop genomic size in (c). A mathematical model for the case (b) is discussed in Appendix B, but the s-independence can also be explained intuitively. Let us consider the point "1" and "2" in Fig. 5(b), which we assume are separated by a genomic distance $s > s_0$. To reach point "2", the polymer chain bounces a few times on the boundaries of the confined region. In doing so any information about the length s is lost and therefore the probability distribution that the two points be separated by a distance $r = |\vec{r}_1 - \vec{r}_2|$ cannot depend on s. For confinement of the type presented in Fig. 5(c), if the points "1" and "2" are on different petals of the flower-like configuration then again the distance distribution of $r = |\vec{r}_1 - \vec{r}_2|$ will not depend on the genomic distance s. Experimental data show no "ideal" confinement ($\nu = 0$), as the average exponent is non-vanishing, although very small $\nu_\beta \approx 0.1$. We note that, although confined, the β phase is characterized by a radius $R_\beta > R_\alpha$ in the range of genomic distances up to ≈ 200 kb.

C. Coexisting topologies in single cells

Although so far we have analyzed experimental probability distributions over the whole set of M distinct cells, here we perform a single-cell analysis. The analysis reveals various coexisting conformations in which the chromosomes of each cell alternate between micro-domains of α and β phases (see also Appendix C).

To illustrate this, we first constructed a global distance probability distribution of consecutive tags by averaging $P_{i,i+1}(r)$ over all tag positions i for cells within a cell line. We denote this averaged distribution by $\langle P_{i,i+1}(r)\rangle_i$. Employing a two Gaussian fit (4) on $\langle P_{i,i+1}(r)\rangle_i$ provides global values for f_α, R_α, and R_β across all tags. Fig. 6 displays plots of single cells distance distributions over consecutive tags ($|i-j|=1$). In contrast to the data in Fig. 2, which are for fixed i and j and cover all $M \sim 10^3 - 10^4$ chromosomes, those in Fig. 6 are limited to just $N \sim 60-80$ samples, corresponding to the number of tags used in mFISH experiments (see Table I). Despite the lower statistics, Fig. 2 demonstrates a strong agreement between the averaged distribution $\langle P_{i,i+1}(r)\rangle_i$ and the distributions of individual cells. This result confirms the existence of at least two populations with distinct spatial distributions within each cell.

One can now define a probability to find a given local chromatin conformation in the β phase as follows

$$p_\beta(i) = \frac{(1-f_\alpha)g(r;R_\beta)}{\langle P_{i,i+1}(r)\rangle_i} \tag{7}$$

with $g(r;R)$ as in (1). This definition applies to a single cell. Note that p_β increases for larger measured distances r, as it is more likely that the measured cell is in the phase β at the tag i. In Fig. 7, we superpose the plots of the local gyration radius $R_g(i)$ (black solid line) and the local probability $p_\beta(i)$ (blue dashed lines), both calculated over a sliding window of 5 subsequent tags centered around the tag i.

This enables the comparison of the spatial variability (expressed in tag number) of these physical quantities. Interestingly, the figure shows that domains with a high β phase content co-localize with maxima in the local radius of gyration, whereas clusters with much lower p_β are centered around the lower values of R_g, thus within

compact chromatin regions. Although the location of these different domains varies strongly from cell to cell, the variation of these two quantities is remarkably correlated.

D. Connection with other studies

Currently, there is a consensus regarding the existence of a markedly heterogeneous chromatin organization that manifests itself in significant cell-to-cell variations. This conclusion is supported by experiments [4, 49, 50] and simulations [29, 44, 51–53]. Conte et al. [29] recently analyzed the same mFISH data discussed in this work [4] and showed that they are consistent with micro-phase separation, which agrees with our conclusions. The advantage of the distance distribution analysis employed here is that one can define scaling exponents for the two phases and extract them in a direct way from the experimental data.

To reconcile discrepancies observed between some Hi-C and (non-multiplexed) FISH data Shi and Thirumalai [18] invoked the two phase model (4). They showed that these discrepancies can be resolved if a heterogeneous model with two or more sub-populations is used. Their analysis was done on Hi-C data and on low-throughput (multicolor) FISH data [50], therefore it could not reveal scaling laws that appear when several labels regularly spaced along the genome are used, as is the case for the multiplexed FISH experiments by Bintu et al. [4] we analyzed here.

Our findings are also in line with recent work modeling chromatin structure using insights obtained from Hi-C data [44, 54, 55]. These papers focus on the decay of Hi-C contact probabilities $P(s)$ as function of the genomic distance s. The crumpled globule model would predict $P(s) \sim R^{-3} \sim s^{-3\nu}$ with $\nu = 1/3$, as obtained from Eq. (1) and (6) by taking the limit $r \to 0$ to obtain the contact probability. This scaling behavior for $P(s)$ is found to be consistent with Hi-C data only at genomic distances $s > 1$ Mb [44]. The deviations observed at shorter scales are explained using a crumpled polymer model with superimposed random loops [44]. This is indeed similar to our description of chromatin as a two phase system: a crumpled globule (α phase) and looped/confined phase (β phase). Our distance distributions analysis indicates phase separation in single cells in domains of the size of $\sim$ 100 kb. From the analysis of the total radius R_T we observed different behaviors at short ($s < 150$ kb) and long ($s > 450$ kb) genomic scales, reminiscent of the different scaling regimes discussed in [44]. The mFISH data indicate an increase of the scaling exponent for large genomic lengths with ν_{long} consistent with, or very close to, the crumpled globule limit, see Table II.

In the future, it would be interesting to extend the distance distribution analysis to more cell lines and also explore chromosomal regions below the 30 kb resolution.

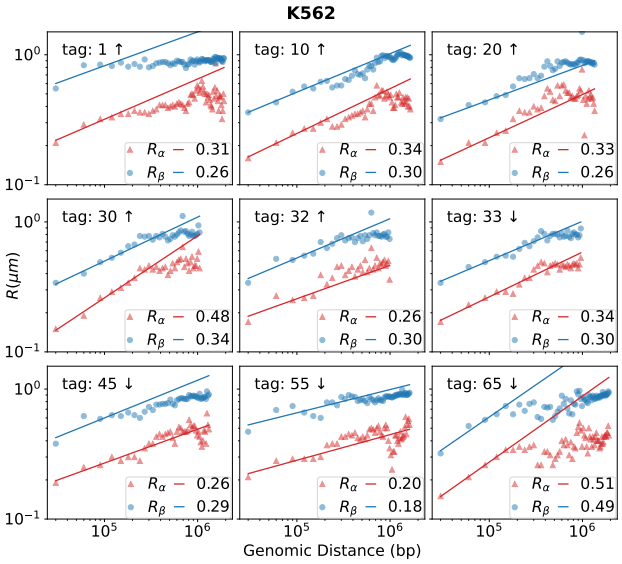

FIG. A.1. Scaling of the characteristic radii of the two phases R_α and R_β for a few selected tags in the leukemia cell line K562. The notation is the same as of Fig. 3 with "tag: 10 ↑" means that R_α and R_β are calculated for $i = 10$ and $j = 11, 12 \ldots N$. The solid lines are power-law fits over the the first five points ($s \leq 150$ kb).

ACKNOWLEDGEMENT

Discussions with M. Barbi, M. Bardou, T. Foldes, M. Liefsoens and M. Nollmann are gratefully acknowledged. LR acknowledges doctoral thesis support from the Lab-MUSE EpiGenMed within the I-Site MUSE (ANR-16-IDEX-0006). MS acknowledges financial support from Fonds Wetenschappelijk Onderzoek (FWO 11O2323N). EC wishes to thank the Laboratoire Charles Coulomb of the University of Montpellier, where part of this work was done, for its kind hospitality. This project was supported by the LabEx NUMEV (ANR-10-LABX-0020) within the I-Site MUSE (ANR-16-IDEX-0006).

Appendix A: Cell lines variability

All distance distributions $P_{ij}(r)$ obtained in the four experiments are similar to those shown in Fig. 2. At genomic distances $s < 150$ kb ($|i - j| < 5$) one can clearly identify two distinct characteristic radii R_α and R_β [43]. While the two HCT116 and the IMR90 cell lines show a markedly different scaling for R_α and R_β, in the leukemia cell line K562 (Fig. A.1) the average exponents ν_α and ν_β are both consistent with the crumpled globule value $\nu = 1/3$, see Table II. The peculiar behavior of the K562 cell line suggests a different type of chromatin organization. To analyze further the cell-to-cell variability we have plotted in Fig. A.2 the values of the exponents ν_α, ν_β and ν_{short} as functions of the tag position i around which they are calculated. For the tags in the interval

$5 < i < N - 5$ the exponents are obtained from averaging over $j > i$ ($\uparrow$) and $j < i$ ($\downarrow$) with $|i - j| \leq 5$ ($s < 150$ kb). For the tags $i < 5$ and $i > N - 5$ we used only the $j > i$ and $j < i$, respectively. Of all cell lines, the Auxin-treated HCT116 shows a remarkable homogeneous behavior across the different tags, with clearly distinct exponents from ν_α and ν_β, and with an intermediate value of $\nu_{\rm short}$, as expected from Eq. (5). While auxin removes some of the topological constraints, the chromatin organization of the auxin-treated cell line as quantified from the values of the exponents is similar to the HCT116 untreated case. The untreated HCT116 line shows much stronger tag-to-tag variability of the scaling exponents, which is attributed to the pinned TAD boundaries, as discussed in [4]. Distinct values of ν_α and ν_β are also obtained for the IMR90 cell line with a tag position variability which is similar to the untreated HCT116 case. Finally the leukemia cell line K562 shows "anomalous" overlapping distributions for ν_α and ν_β with, surprisingly, $\nu_{\rm short}$ smaller than ν_α and ν_β.

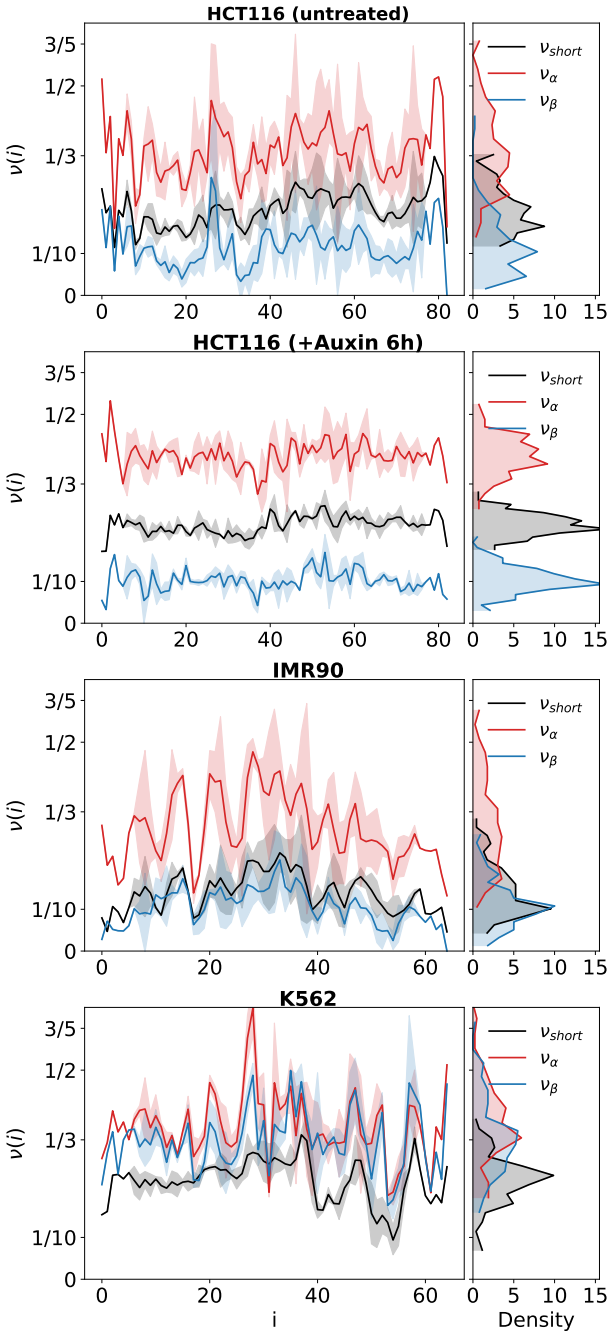

FIG. A.2. Plots of the exponents ν_α, ν_β and ν as function of the tag position i around which they are calculated (see text for details). In the cell lines HCT116 and IMR90 $\nu_\alpha > \nu_\beta$, while $\nu_\alpha \approx \nu_\beta$ for the K562 cell line.

Appendix B: Confined Ideal polymer

We discuss here the effect of confinement in the simpler and analytically tractable case of an ideal polymer. We consider a polymer consisting of N monomers of average bond length b which is subject to an external radial potential $U(r)$ acting on all its monomers, as depicted in Fig. 5(b). The potential is attractive forcing the polymer to remain confined in the vicinity of the origin. One can consider different forms for $U(r)$, such as the infinite well case

$$U_w(r) = \begin{cases} 0 & r < R \\ +\infty & r \geq R, \end{cases} \tag{B1}$$

a harmonic potential

$$U_h(r) = \frac{k}{2} r^2, \tag{B2}$$

of other types of confinement. The probability distribution for the end point vector $\vec{r}$, from the origin of the potential, of a segment of length s satisfies the differential equation [56]

$$\frac{\partial Q(\vec{r}, s)}{\partial s} = \frac{b^2}{6r} \frac{\partial^2 [rQ(\vec{r}, s)]}{\partial r^2} - \beta U(r) Q(\vec{r}, s) \tag{B3}$$

This equation can be solved by separation of variables, i.e. by seeking solutions of the type $Q(\vec{r}, s) = \psi_p(\vec{r}) g_p(s)$, labeled by an index p. The general solution is then given by the linear combination

$$Q(\vec{r}, s) = \sum_p A_p \psi_p(\vec{r}) g_p(s), \tag{B4}$$

where the coefficients A_p are chosen to satisfy the desired boundary conditions (the full series solution of the infinite-well confinement is given in [57]). Separation of variables leads to a differential equation for $g_p(s)$ of the type $dg_p(s)/ds = -\alpha_p g_p(s)$ with exponentially decaying solutions $g_p(s) = e^{-\alpha_p s}$. In the limit $s \to \infty$ one uses the ground state dominance (GSD) approximation which retains the component with the smallest α_p [58]

$$Q(\vec{r}, s) \approx A_0 e^{-\alpha_0 s} \psi_0(\vec{r}) \tag{B5}$$

where $\psi_0(\vec{r})$ is the ground state solution of the associated Schödinger-like equation

$$\frac{b^2}{6r}\frac{\partial^2[r\psi_0(\vec{r})]}{\partial r^2} - \beta U(r)\psi_0(\vec{r}) = -\alpha_0\psi_0(\vec{r}) \qquad \text{(B6)}$$

We note that in the GSD approximation the distribution in $\vec{r}$ is independent of the polymer length s, due to the factorization in (B5). This is different from a free ideal polymer where the probability distribution is a function of the scaled variable $\vec{r}^2/s$.

For a infinite well potential (B1) the solution of (B6) is a spherical Bessel function of order zero

$$\psi_0(\vec{r}) = \frac{\sin(\pi r/R)}{r} \qquad \text{for } r \le R \qquad \text{(B7)}$$

and $\psi_0(r) = 0$ for $r > R$. For a confining harmonic potential (B2) one finds

$$\psi_0(\vec{r}) = \exp(-3r^2/2R_*^2) \qquad \text{(B8)}$$

where we have defined $R_*^2 \equiv b/\sqrt{3\beta K}$. In both cases, we have omitted normalization factors.

We recall that $\psi_0(\vec{r})$, obtained from the ground state dominance (B5), is the probability distribution of the end-point of a very long polymer. The probability that two points along the polymer are separated by a given vector distance $\vec{r} = \vec{r}_2 - \vec{r}_1$ is given by

$$P(\vec{r}) = \int d\vec{r}_1 d\vec{r}_2\, \psi_0(\vec{r}_1)\psi_0(\vec{r}_2)\, \delta(\vec{r}_1 - \vec{r}_2 - \vec{r}) \qquad \text{(B9)}$$

where $\delta()$ denotes the Dirac delta function. Inserting the following Fourier transform representation of the Dirac delta,

$$\delta(\vec{x}) = \frac{1}{(2\pi)^3}\int d\vec{k}\, e^{i\vec{k}\cdot\vec{x}}, \qquad \text{(B10)}$$

we obtain

$$P(\vec{r}) = \int d\vec{k}\ \left[\widetilde{\psi_0}(\vec{k})\right]^2 e^{i\vec{k}\cdot\vec{r}} \qquad \text{(B11)}$$

where $\widetilde{\psi_0}$ is the Fourier transform of $\psi_0(\vec{r})$. The Fourier transform of a Gaussian function is also a Gaussian, hence the harmonic confinement gives

$$P_h(r) = 4\pi r^2 \left(\frac{3}{4\pi R_*^2}\right)^{3/2} e^{-3r^2/4R_*^2} \qquad \text{(B12)}$$

where we have included the jacobian $4\pi r^2$. For the infinite well confinement we solved the problem numerically. We generated independent vector pairs $\vec{r}_1$ and $\vec{r}_2$ within a sphere of radius R distributed according to (B7) and extracted the probability distribution of the distance $r \equiv |\vec{r}_1 - \vec{r}_2|$ shown in Fig. B.1 as solid line. This distribution deviates from a Gaussian one, which is shown as a dashed line. It decays faster than the Gaussian at larger r and vanishes for $r > 2R$ ($r = 2R$ is the maximal distance between two points within a sphere of radius R). For generic confining potentials the distribution is not a Gaussian one.

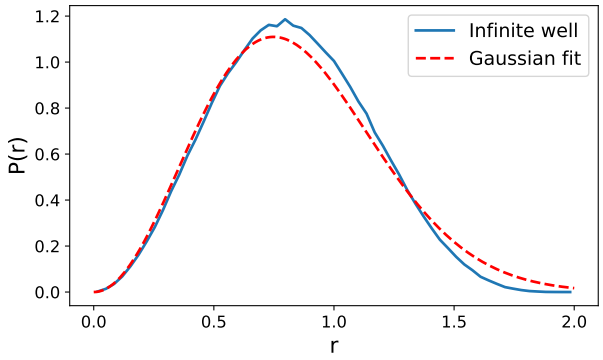

FIG. B.1. Solid line: Numerical solution of the probability distribution $P(r)$ of spatial distance r between two points separated by a long genomic distance s for confinement in the case of infinite well-potential (B1) with $R = 1$. Dashed line: Gaussian fit of $P(r)$.

We consider now the case of confinement of an ideal polymer in a flower-like shape conformation as in Fig. 5(c). For simplicity we will assume that each petal has the same length l and we number them $1, 2, \ldots n$ according to increasing genomic position. Let us again for simplicity consider two points on the flower with positions $\vec{r}_1$, $\vec{r}_2$ and are located in the middle of petals i and j. The genomic distance between these points is therefore $s = l|i - j|$. Given $p_p(\vec{r}^*; l)$, defined as the probability distribution that the mid-point of the petal is at a distance $\vec{r}^*$ from the center of the flower, the probability distribution that the spatial distance between the two points is equal to $\vec{r}$ is then given by

$$P(\vec{r}) = \int d\vec{r}_1\, d\vec{r}_2\, p_p(\vec{r}_1; l)p_p(\vec{r}_2; l)\delta(\vec{r}_1 - \vec{r}_2 - \vec{r}). \qquad \text{(B13)}$$

For an ideal polymer $p_p(\vec{r}^*; l)$ is a Gaussian distribution and so is $P(\vec{r})$ obtained from the previous formula. Most importantly, this result is independent of the genomic distance s, hence it predicts a scaling $R \sim s^\nu$ with $\nu = 0$. This holds for distances $s > l$, i.e. beyond the length of a petal.

Appendix C: Fractions of α and β phases

From fits of the experimental distance distributions $P_{ij}(r)$ with the two phase model (4) one obtains estimates of f_α and $f_\beta = 1 - f_\alpha$, the fraction of the two phases for pairs of labels i and j. Figure C.1 shows the values of f_α (central thick colored lines) averaged over all begin tags i as a function of genomic separation $l_t|i - j|$, where $l_t = 30$ kb is the spacing between two consecutive tags of the experiments. The colored areas estimates the variability (standard deviation) over the

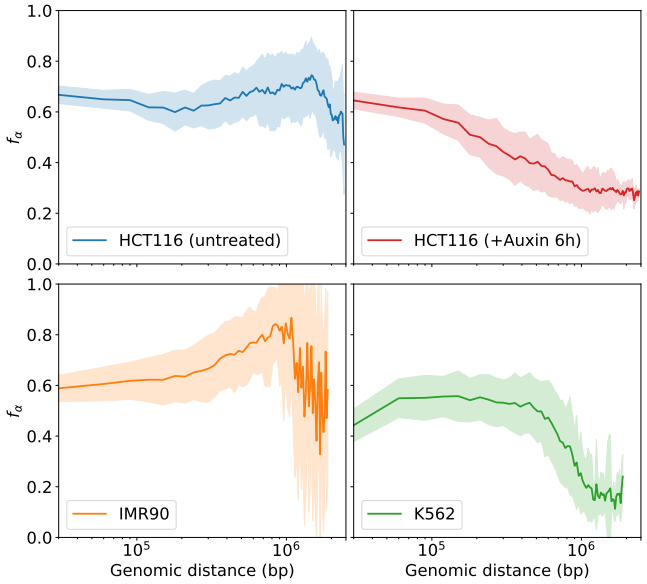

FIG. C.1. Plot of the average f_α, the fraction of the α phase as a function of the genomic distance (solid colored line). The light-colored area indicates the standard deviation.

different data. For a microphase separation we expect a constant value of f_α for a distance range corresponding to the characteristic length of the domains. The data in Fig. C.1 show that $f_\alpha \approx 0.6$ for the shortest tags distance $|i-j| = 1$ (except for K562 which, as discussed previously, displays some anomalous behavior compared to the other cell lines). The data also show that f_α remains approximately constant for some range of genomic distances. This range appears to be somewhat cell-line dependent. As discussed in the main text, the distance distribution analysis is more reliable for tag distances which are not too distant ($|i-j| \lesssim 5$), as the differences between R_α and R_β are larger.

[1] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, Science **295**, 1306 (2002).
[2] 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).
[3] W. A. Bickmore, Annu. Rev. Genomics Hum. Genet. **14**, 67 (2013).
[4] 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**, eaau1783 (2018).
[5] Q. Szabo, F. Bantignies, and G. Cavalli, Sci. Adv. **5**, eaaw1668 (2019).
[6] 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.*, Nature Gen. **53**, 477 (2021).
[7] M. Götz, O. Messina, S. Espinola, J.-B. Fiche, and M. Nollmann, Nature Comm. **13**, 5375 (2022).
[8] J. Gurgo, J.-C. Walter, J.-B. Fiche, C. Houbron, M. Schaeffer, G. Cavalli, F. Bantignies, and M. Nollmann, BiorXiv (2022), 10.1101/2022.05.16.492046.
[9] A. Rosa and R. Everaers, PLoS Comput. Biol. **4**, e1000153 (2008).
[10] M. Emanuel, N. H. Radja, A. Henriksson, and H. Schiessel, Phys. Biol. **6**, 025008 (2009).
[11] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, Proc. Natl. Acad. Sci. **109**, 16173 (2012).
[12] C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo, Proc. Natl. Acad. Sci. **110**, E3605 (2013).
[13] V. F. Scolari and M. C. Lagomarsino, Soft Matter **11**, 1677 (2015).
[14] C. A. Brackley, J. M. Brown, D. Waithe, C. Babbs, J. Davies, J. R. Hughes, V. J. Buckle, and D. Marenduzzo, Genome Biol. **17**, 59 (2016).
[15] A. Goloborodko, J. F. Marko, and L. A. Mirny, Biophys. J. **110**, 2162 (2016).
[16] R. Cortini, M. Barbi, B. R. Caré, C. Lavelle, A. Lesne, J. Mozziconacci, and J.-M. Victor, Rev. Mod. Phys. **88**, 025002 (2016).
[17] J.-C. Walter, N.-O. Walliser, G. David, J. Dorignac, F. Geniet, J. Palmeri, A. Parmeggiani, N. S. Wingreen, and C. P. Broedersz, New J. Phys. **20**, 035002 (2018).
[18] G. Shi and D. Thirumalai, Nature Comm. **10**, 3894 (2019).
[19] A. Lesage, V. Dahirel, J.-M. Victor, and M. Barbi, Epigenetics & Chromatin **12**, 28 (2019).
[20] T. Földes, A. Lesage, and M. Barbi, Phys. Rev. Lett. **127**, 277801 (2021).
[21] M. Liefsoens, T. Földes, and M. Barbi, bioRxiv **2023.10.24.563731**, 2023 (2023).
[22] J. Dekker and L. Mirny, Cell **164**, 1110 (2016).
[23] R. Blossey, *Chromatin: Structure, Dynamics, Regulation* (CRC Press, 2017).
[24] T. Cremer and M. Cremer, Cold Spring Harbor Perspectives in Biology **2**, a003889 (2010).
[25] T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq, Molecular Cell **78**, 539 (2020).
[26] 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).
[27] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell Reports **15**, 2038 (2016).
[28] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucl. Acids Res. **42**, 9553 (2014).

[29] M. Conte, E. Irani, A. M. Chiariello, A. Abraham, S. Bianco, A. Esposito, and M. Nicodemi, Nature Comm. **13**, 4070 (2022).

[30] M. Imakaev, G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, and L. A. Mirny, Nature Methods **9**, 999 (2012).

[31] O. Oluwadare, M. Highsmith, and J. Cheng, Biological Procedures Online **21**, 1 (2019).

[32] C. Cui, W. Shu, and P. Li, Front. Cell. Dev. Biol. **4**, 89 (2016).

[33] K. E. Polovnikov, S. Nechaev, and M. V. Tamm, Phys. Rev. E **99**, 032501 (2019).

[34] S. A. Belan and D. E. Starkov, JETP Letters **115**, 763 (2022).

[35] P.-G. De Gennes, *Scaling concepts in Polymer Physics* (Cornell University Press, 1979).

[36] K. Polovnikov, S. Nechaev, and M. Tamm, Soft Matter **14**, 6561 (2018).

[37] M. V. Tamm, L. I. Nazarov, A. A. Gavrilov, and A. V. Chertovich, Phys. Rev. Lett. **114**, 178102 (2015).

[38] K. E. Polovnikov, M. Gherardi, M. Cosentino-Lagomarsino, and M. V. Tamm, Phys. Rev. Lett. **120**, 088101 (2018).

[39] J. L. A. Dubbeldam, V. G. Rostiashvili, A. Milchev, and T. A. Vilgis, Phys. Rev. E **83**, 011802 (2011).

[40] J.-C. Walter, A. Ferrantini, E. Carlon, and C. Vanderzande, Phys. Rev. E **85**, 031120 (2012).

[41] T. Sakaue, J.-C. Walter, E. Carlon, and C. Vanderzande, Soft Matter **13**, 3174 (2017).

[42] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, J. Phys. France **49**, 2095 (1988).

[43] Python scripts for the data analysis discussed in this paper and for the generation of figures are freely available in https://github.com/Loucif-REMINI/polymer-model-analyses-of-chromatin-mFISH-data.

[44] K. E. Polovnikov, B. Slavov, S. Belan, M. Imakaev, H. B. Brandão, and L. A. Mirny, Phys. Rev. X **13**, 041029 (2023).

[45] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Europhys. Lett. (EPL) **23**, 373 (1993).

[46] A. Rosa and R. Everaers, Phys. Rev. Lett. **112**, 118302 (2014).

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

[48] J. D. Halverson, J. Smrek, K. Kremer, and A. Y. Grosberg, Rep. Prog. Phys. **77**, 022601 (2014).

[49] 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.*, Nature **544**, 59 (2017).

[50] E. H. Finn, G. Pegoraro, H. B. Brandão, A.-L. Valton, M. E. Oomen, J. Dekker, L. Mirny, and T. Misteli, Cell **176**, 1502 (2019).

[51] A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, and N. Gilbert, Mol. Cell **72**, 786 (2018).

[52] M. Conte, L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, and M. Nicodemi, Nature Comm. **11**, 3289 (2020).

[53] M. Chiang, C. A. Brackley, C. Naughton, R.-S. Nozawa, C. Battaglia, D. Marenduzzo, and N. Gilbert, bioRxiv **06.09**, 495447 (2022).

[54] K. Polovnikov and B. Slavov, Phys. Rev. E **107**, 054135 (2023).

[55] B. Slavov and K. Polovnikov, JETP Letters **118**, 208 (2023).

[56] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Vol. 73 (Oxford University Press, 1988).

[57] P. Hahnfeldt, J. E. Hearst, D. J. Brenner, R. K. Sachs, and L. R. Hlatky, Proc. Natl. Acad. Sci. **90**, 7854 (1993).

[58] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (American Institute of Physics, 1994).