



**HAL**  
open science

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

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

► **To cite this version:**

Loucif Remini, Midas Segers, John Palmeri, Jean-Charles Walter, Andrea Parmeggiani, et al.. Chromatin structure from high resolution microscopy: scaling laws and microphase separation. 2023. hal-04177307v1

**HAL Id: hal-04177307**

**<https://hal.science/hal-04177307v1>**

Preprint submitted on 28 Aug 2023 (v1), last revised 18 Mar 2024 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Loucif Remini,<sup>1,2</sup> Midas Segers,<sup>3</sup> John Palmeri,<sup>1</sup> Jean-Charles Walter,<sup>1</sup> Andrea Parmeggiani,<sup>1</sup> and Enrico Carlon<sup>3</sup>

<sup>1</sup>Laboratoire Charles Coulomb (L2C), Univ. Montpellier, CNRS UMR5221, Montpellier, France

<sup>2</sup>Centre de Biologie Structurale (CBS), Univ. Montpellier, CNRS UMR5048, Montpellier, France

<sup>3</sup>Soft Matter and Biophysics, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

(Dated: August 28, 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, aiming 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  $\alpha$  and phase  $\beta$ . 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, 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 [9–19]. This interest stems from the fact that the 3D chromatin structure has a strong influence on several genomic processes such as transcription or replication [20, 21]. 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. 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), 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 [22]. Other models of genome organization have been proposed in the literature such as the strings and binders model [11], the block-copolymer model [23] or the diffusing transcription factor model [12].

It is very likely that different mechanisms operate simultaneously to contribute to the complex hierarchical organization of eukaryotic chromatin.

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 to understand 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 [24], therefore raw data require normalization and other pre-processing steps. Fluorescence in

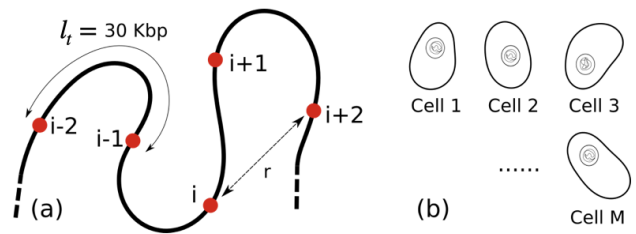


FIG. 1. Setup of multiplexed FISH as developed in Ref. [4]. (a)  $1 \leq i \leq N$  genomic sites equally spaced are targeted by FISH probes during sequential hybridization cycles. In the setup of [4] the genomic distance between consecutive tags was  $l_t = 30$  Kbp. The 3D spatial coordinates of the 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  $j$ .

situ hybridization (FISH, Fig. 1) has also been used to investigate chromosomal structure [25]. 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 very low throughput technique, the main issue being the limited amount of fluorescent probes with different emission spectra available so that only a few sites can be visualized. However, quite recently, a multiplexed FISH (m-FISH) method was developed [4]. Using unique readout sequences and sequential hybridization cycles, the method allows to determine the positions of  $N \sim 50 - 100$  chromosomal sites with high resolution ( $< 50$  nm), 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 which cannot be inferred from Hi-C pairwise contact probabilities [4, 8].

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 histograms  $P_{ij}(r)$  of distances between the 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 be implemented and 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 histograms. 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 emerging regularities from typical chromatin conformations from the experimental data. Appendices provide technical details on the polymer models considered and

TABLE I. Summary of experimental data analyzed from [4]. These target two different regions of chr.21: a 1.2 Mb region in the 28 – 30 Mb genomic position and another one 2.5 Mb in the 34 – 37 Mb position. Consecutive tags are separated by a genomic distance of 30 kbp. HCT116 is a colon cancer cell line, IMR90 a lung fibroblast cell line and K562 an erythroleukemia cell line.

n.	Experiment/Cell line	$N$	$M$
1	HCT116_chr21-34-37Mb_untreated	83	11631
2	HCT116_chr21-34-37Mb_6h-auxin	83	9526
3	IMR90_chr21-28-30Mb	65	4871
4	K562_chr21-28-30Mb	65	13997

a full analysis of the whole dataset histograms.

## II. DISTANCE DISTRIBUTION HISTOGRAMS

Table I summarizes the main features of the four data sets from [4] analyzed in this paper. Each dataset corresponds to a cell line whose properties will be detailed later in Section III. 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 histograms are obtained. The histograms from the  $M$  experimental samples are normalized as follows

$$\int_0^\infty P_{ij}(r) dr = 1 \quad (1)$$

For an ideal polymer, the distance distribution is a Gaussian in the vector distance  $\vec{r}$ , while for its length  $r = |\vec{r}|$  takes the form

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

where the term  $4\pi r^2$  comes from the Jacobian of the distribution. The above distribution is normalized as (1) 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 [26]. This is because self-avoidance gets screened by the surrounding polymers. The mean square radius of an ideal polymer scales as

$$R \sim |i - j|^{1/2} \quad (3)$$

where the exponent 1/2 describes the universal scaling behavior of random walks. It has been recently shown [27] that the introduction of suitable monomer-monomer harmonic pairwise interactions in a bead and spring polymer model leads to an equilibrium distance distribution as (2), but with a scaling described by

$$R \sim |i - j|^\nu \quad (4)$$

By tuning the monomer-monomer interactions, one can generate any values of the exponent in the interval  $1/3 \leq \nu \leq 1/2$  [27]. 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 [27], as well as in other problems of polymer dynamics

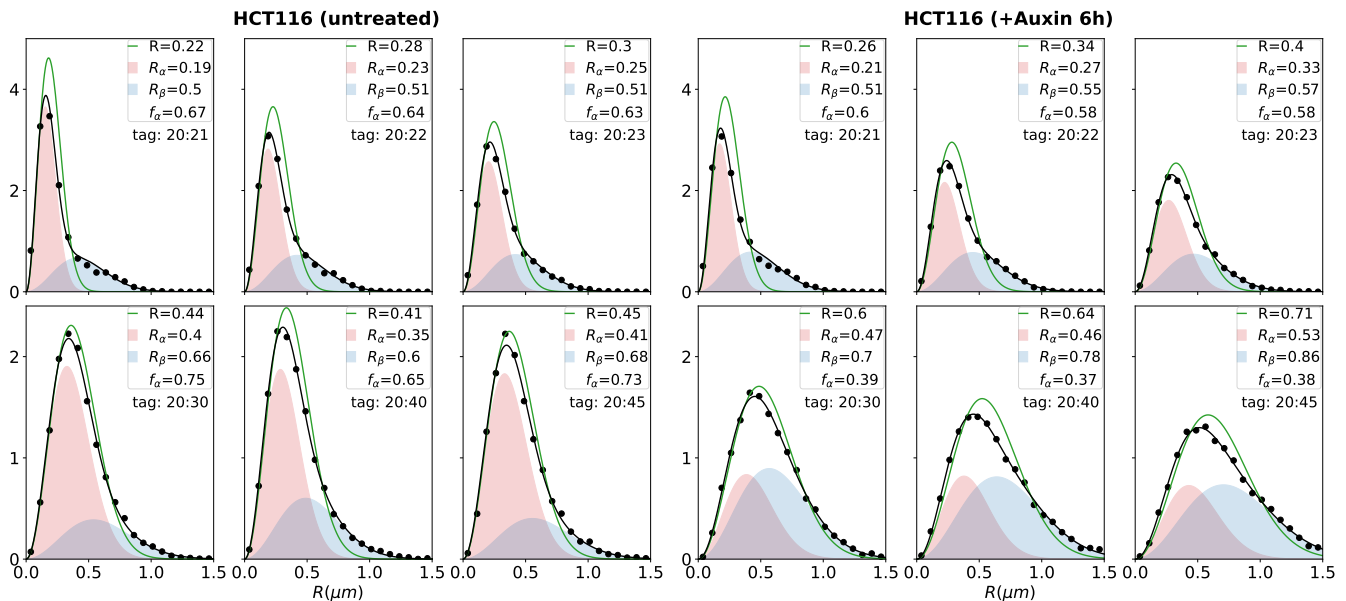


FIG. 2. Histogram of distance distributions  $P_{ij}(r)$  from experimental data (black circles) for different tag locations  $i, j$  for experiments of the HCT116 cell line (sets n.1 and n.2 of Table I). The values for  $i$  and  $j$  are indicated in each graph. In the two sets, we fix  $i = 20$ . The tag  $j$  is increased from  $j = 21$  (top left) to  $j = 45$  (bottom right). Solid lines are fit to a single gaussian model (Eq. (2), green line) and double gaussian model (Eq. (5), black line). The legends in each plot show the value of  $R$  for a single gaussian model (Eq. (2), green) and of  $R_\alpha$ ,  $R_\beta$  and  $f_\alpha$  for the double gaussian model (Eq. (5)). The red and blue filled areas denote the contribution of the two components of the fit. The distance in the horizontal axis is in  $\mu\text{m}$ .

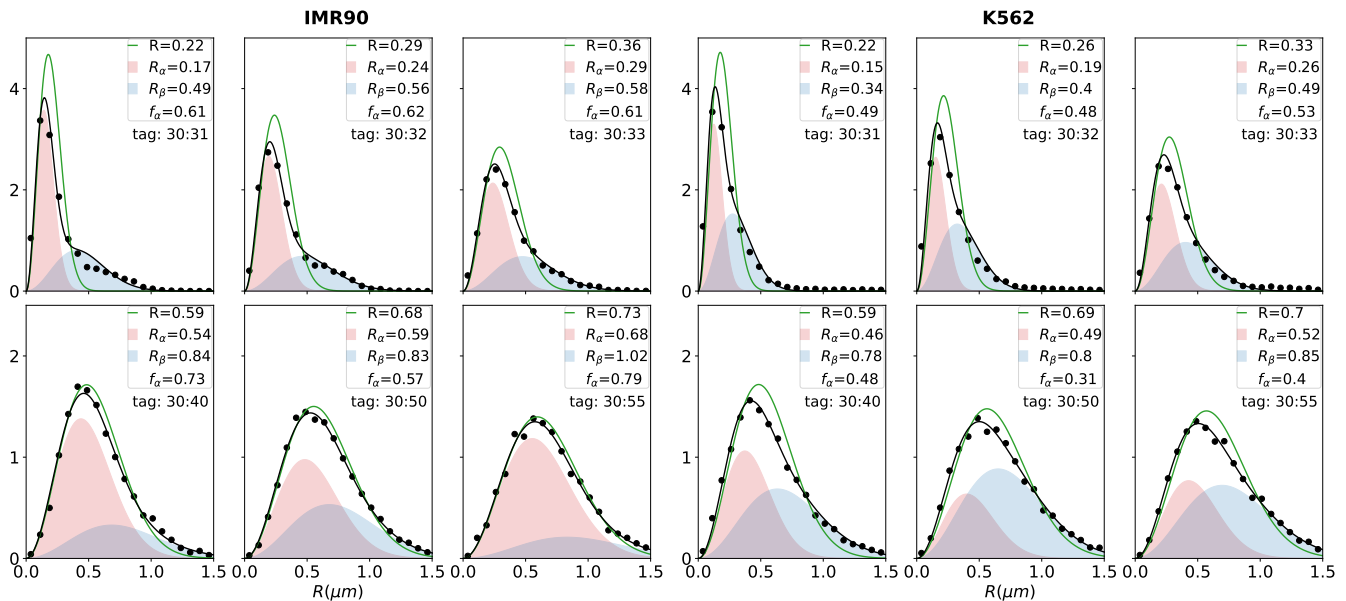


FIG. 3. Same analysis as in Fig. 2 for experiments n.3 (IMR90 cell line) and n.4 (K562 cell line) of Table I.

[28–30]. We note that the case  $\nu = 1/3$  is the exponent of a crumpled globule [31], a phase discussed in some more detail in Sec. IV. Therefore, the Gaussian distribution (2) encompasses a broad range of polymer models.

Figures 2 (HCT116 and auxin-treated HCT116 cell lines) and 3 (IMR90 and K562 cell lines) show several plots of experimentally measured histograms  $P_{ij}(r)$  (cir-

cles), where the tags  $i$  and  $j$  considered are reported in each graph and the horizontal scale is the spatial distance in micrometers ( $\mu\text{m}$ ). The green solid lines are a one parameter fit to Eq. (2), 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 range of  $|i - j| \approx 4$  corresponding to a genomic distance of 120 Kbp). As the genomic distance increases, i.e. larger  $|i - j|$ , the experimental distributions gets closer to (2), seen by the trend of merging of the green line (gaussian fit) to the black circles (experiments) in Figs. 2 and 3. This merging trend is however stronger in the untreated HCT116 and the IMR90 lines, while there remains a substantial gap between the single and double gaussian fits for large genomic distances  $|i - j| = 25$  for the auxin-treated HCT116 and the K562 lines.

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

$$P_{ij}(r) = f_\alpha g(r; R_\alpha) + (1 - f_\alpha) g(r; R_\beta) \quad (5)$$

where the parameter  $0 \leq f_\alpha \leq 1$  (respectively  $f_\beta = 1 - f_\alpha$ ) is the fraction of chromosomes with mean radius  $R_\alpha$  (respectively  $R_\beta$ ) around the tags  $i$  and  $j$ . We will refer to the two phases as  $\alpha$  and  $\beta$ . The two-phase model (5) provides an excellent fit for the experimental data. These fits are shown as solid black lines in Figs. 2 and 3. 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  $R_\alpha$ ,  $R_\beta$  and  $f_\alpha$  as adjustable parameters. These three values are reported in the legends of Figs. 2 and 3. The radii change with changing  $i$  and  $j$  and, as expected, increase for increasing  $|i - j|$ . The growth rate is however different for  $R_\alpha$  and  $R_\beta$ , as discussed in the next Section. Interestingly, the two phase model (5) was considered on a recent analysis of Hi-C chromosomal data [18]. Such model was introduced to solve some observed discrepancies between Hi-C and earlier non-multiplexed FISH data. We discuss the connection between the results of Ref. [18] and our results in the concluding Section of this paper.

### III. SCALING BEHAVIOR OF $R_\alpha$ AND $R_\beta$

Figures 4 and 5 show the extracted scaling behavior for the four different cell lines. These are log-log plots of the mean radii  $R_\alpha$  and  $R_\beta$  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 Mbp. We characterized the scaling behavior of the radii fitting  $R_\alpha$  and  $R_\beta$  to (4) for close tags, i.e.  $|i - j| \leq 6$  corresponding to a maximal genomic distance of 180 Kb. This choice is dictated by two factors: (i) we observe that the scaling of  $R_\alpha$  and  $R_\beta$  follows an approximate power-law behavior for a limited distance between tags and (ii) the values of  $R_\alpha$  and  $R_\beta$  are more accurately determined from fitting histograms for small  $|i - j|$  as well. This is because there is typically a bigger gap between  $R_\alpha$  and  $R_\beta$  for small  $|i - j|$ , therefore the two

TABLE II. Mean values of the geometric exponents  $\nu_\alpha$  and  $\nu_\beta$  for the different data sets. The last column gives the value of  $f_\alpha$  averaged on all data in the range  $|i - j| \leq 5$ .

Data set	$\nu_\alpha$	$\nu_\beta$	$f_\alpha$
HCT116 (untreated)	0.35(10)	0.11(7)	0.64(5)
HCT116 (+Auxin 6h)	0.40(6)	0.10(3)	0.60(5)
IMR90	0.31(11)	0.11(6)	0.61(8)
K562	0.38(8)	0.35(7)	0.53(8)

radii can be more reliably extracted from the data analysis. Table II gives a summary of the average exponents for the four different cell lines averaged over all the tags. The exponents give a measure of the average geometrical properties of the two phases. With the exception of the K562 cell line, we find consistently  $\nu_\alpha > \nu_\beta$ . The properties of the different cell lines will be discussed in the following.

#### A. HCT116 cell line

HCT116 are human colon carcinoma cells. Figure 4 shows the scaling behavior of  $R_\alpha$  and  $R_\beta$  for a few selected tags of the untreated (left) and the auxin-treated (right) cases for HCT116. Auxin degrades cohesin and it is therefore expected to affect the TAD structure. We find, however, very little difference between the exponents in the two cases, apart from a possibly slightly higher  $\nu_\alpha$  in the auxin-treated case, see Table II. This small increase is a signature that the  $\alpha$ -phase is slightly less compact in the auxin-treated case. This is consistent with the fact that by degrading cohesin, auxin releases some constraints. The overall little difference in exponents in the two cases agrees 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 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, therefore washing out any TAD signature in the ensemble average as done in Hi-C data analysis. The effect of pinned vs. variable boundaries is also visible in Fig. 4: the scaling of  $R_\alpha$  and  $R_\beta$  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.

#### B. IMR90 cell line

Figure 5 (left) shows the scaling properties of the IMR90 lung fibroblast cell line. The exponents  $\nu_\alpha$  and  $\nu_\beta$  are consistent with those observed in the untreated HCT116 cell lines, see Table II. As in the untreated

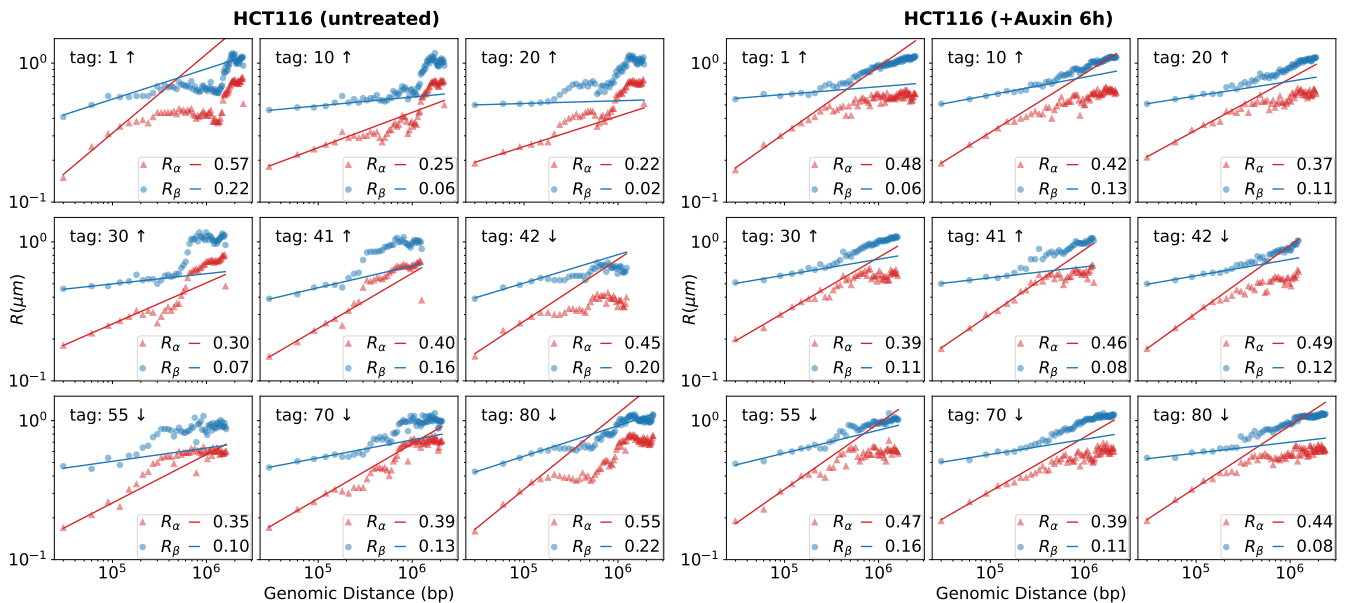


FIG. 4. Log-log plots of scaling behavior  $R_\alpha$  (red triangles) and  $R_\beta$  (blue circles) vs. genomic distance for some selected tags of the HCT116 cell. Left: untreated and Right: 6h-auxin treated cells. The reference tag used is given in each plot. The notation “tag: 1 ↑” means that  $R_\alpha$  and  $R_\beta$  are calculated for  $i = 1$  and  $j = 2, 3, \dots, N$  and “tag: 55 ↓” plots the data for  $i = 55$  and  $j = 54, 53, \dots, 1$ . The solid lines are power-law fits to the first 5 points of each set.

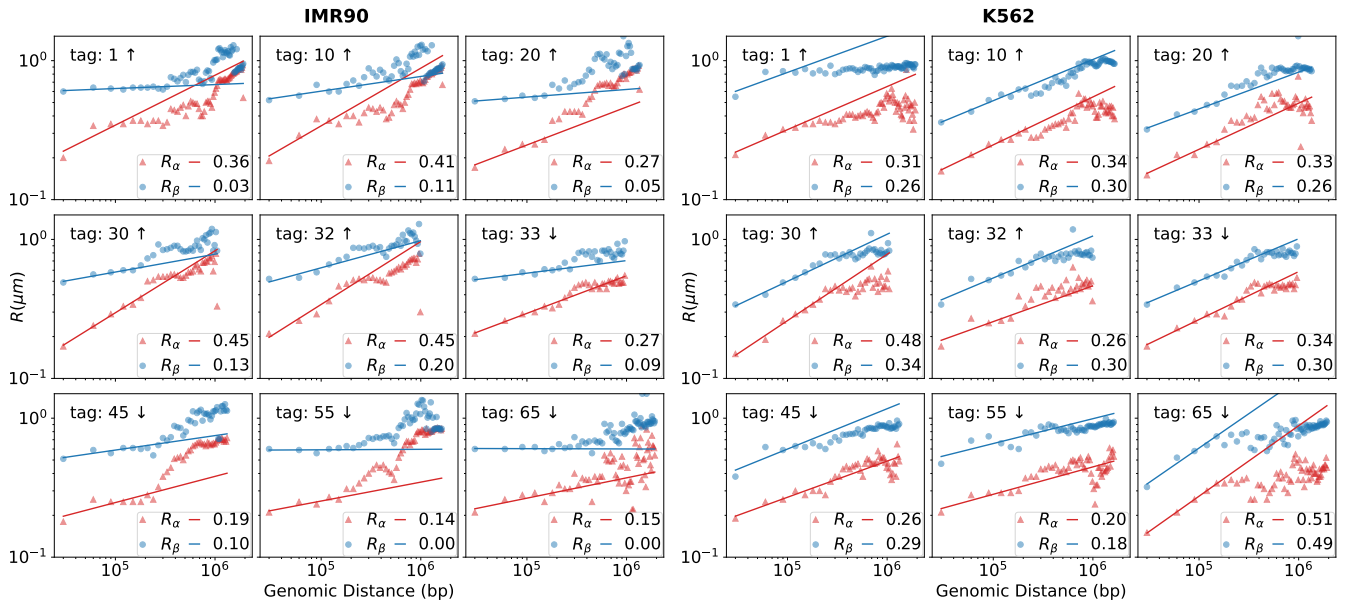


FIG. 5. Same analysis as in Fig. 4 for the cell lines IMR90 (left) and K562 (right).

HCT116 case, one observes from the irregular shape the effects of TADs pinning at specific sites. This is consistent with ensemble-averaged contact map results [4].

### C. K562 cell line

The last case analyzed was the K562 erythroleukemia cell line, shown in Fig. 5 (right). Unlike the previous

cases, we observe for K562 quite close exponents for the two phases  $\nu_\alpha \approx \nu_\beta$ , thus  $\nu_\beta$  is substantially larger than in the other cell lines, see 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. On view of the numerical value of  $\nu_\beta$  it is more logical to characterize this cell line as having two phases of type  $\alpha$  but with two distinct values of  $R_\alpha$ .

## 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 the  $\alpha$  and  $\beta$  phases. The range of genomic length explored by these experiments goes from 30 Kbp to 2.5 Mbp, but the characteristic radii  $R_\alpha$  and  $R_\beta$  can be confidently obtained from data for distances up to approximately 0.5 Mbp. Therefore our claims 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  $\nu_\alpha$  and  $\nu_\beta$ . Interestingly, these exponents are different in the two phases (except for the K562 cell line) indicating substantially different spatial organizations, which we discuss in this Section.

### A. Phase $\alpha$

The exponent we extracted for the phase  $\alpha$  is close to that expected for a crumpled globule, which is  $\nu = 1/3$  [31–33]. 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 no sufficient 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 [33]. 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 [34]. Crumpled globule polymeric phases have been discussed in melts of polymer rings [9] and are believed to be relevant for genome folding structures [34, 35]. Figure 6(a) illustrates the typical conformation of a crumpled globule. Nearby chromatin segments (indicated with dif-

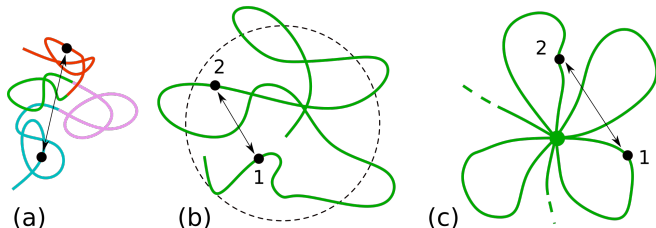


FIG. 6. Possible phases of chromatin: (a) crumpled globule, (b) confining due to external potential, and (c) flower. (a) is a candidate for the phase  $\alpha$  and (b,c) for the phase  $\beta$ . In (a) different segments of the chromatin fiber are colored differently to illustrate the geometry of the globule. The rapid condensation leads to collapses at the local scale involving vicinal regions. In (b) and (c) strong confinement implies  $R_\beta \sim s^0$  for  $s$  sufficiently long.

ferent colors in Fig. 6(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. Phase $\beta$

The phase  $\beta$  is characterized by a very small exponent, which is an indication of strong confinement. Figure 6(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  $R_\beta \sim s^0$ , i.e. the characteristic radius becomes independent on 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 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 A, but the  $s$ -independence can also be explained intuitively. Let us consider the point “1” and “2” in Fig. 6(b), which we assume are separated by a genomic distance  $s > s_0$ . To reach point “2”, the polymer chain bounces a few times with the boundaries of the confined region. In doing that any information about the length  $s$  is lost and the probability distribution that the two points have a distance  $r = |\vec{r}_1 - \vec{r}_2|$  does not depend on  $s$ . For confinement of type of Fig. 6(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 phase  $\beta$  is characterized by a radius  $R_\beta > R_\alpha$  in the range of genomic distances up to  $\approx 2 \cdot 10^5$  bp.

### C. Coexisting topologies

The picture emerging from this analysis reveals different coexisting conformations in which the chromosome of each cell alternates between micro-domains of phases  $\alpha$  and  $\beta$  (see also Appendix B). To illustrate this behavior we plot in Fig. 7 (solid black lines) the local radius of gyration  $R_g(i)$  calculated over five subsequent tags centered around the tag  $i$ . The plots are obtained from the data of six randomly selected cells for the cell lines 1, 2 and 3 of Table I. While so far we have analyzed histograms over the whole set of  $M$  distinct cells, here we perform a single-cell analysis where Fig. 7 illustrates

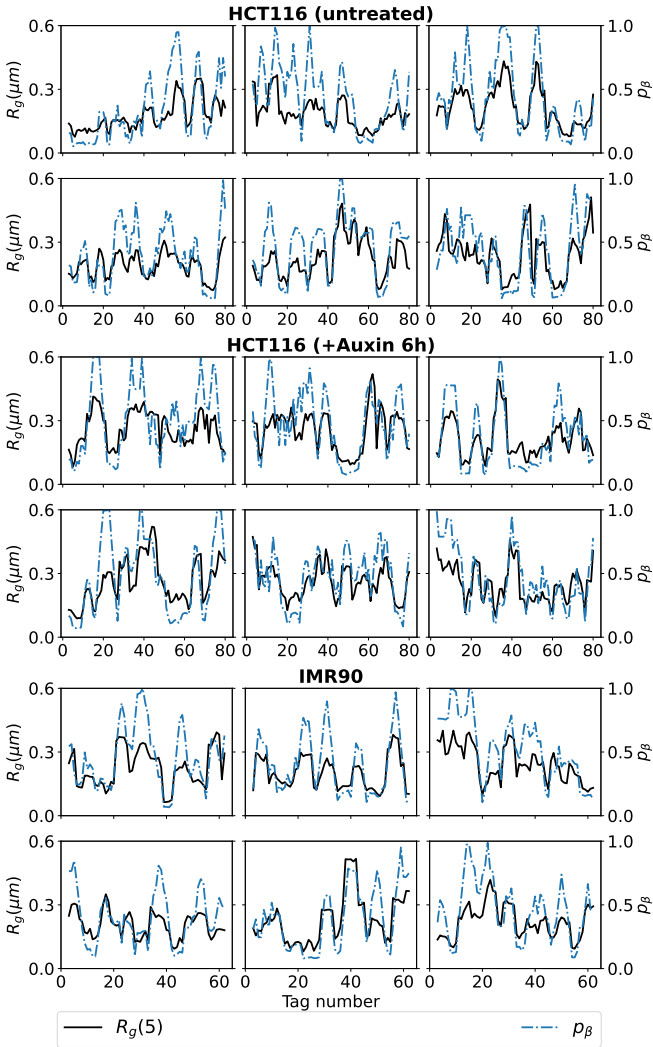


FIG. 7. Radius of gyration  $R_g$  (solid black) and  $\beta$ -phase probability  $p_\beta$  (broken blue) as a function of tag number for different cells and cell lines.  $R_g$  is calculated over five subsequent tags.  $p_\beta$  is defined in Eq. (6).

a marked cell-to-cell variability. The radius of gyration  $R_g(i)$  defined above describes the local structure of the chromatin in a range of 150 Kbp.

In Fig. 7 one remarks large and small values of  $R_g(i)$  indicative of more open and more compact configurations, as expected for the phases  $\beta$  and  $\alpha$ , respectively. We also analyzed the  $\beta$ -phase probability  $p_\beta(i)$  as function of tag position  $i$ , which is defined as follows. We constructed first a global histogram of consecutive tags by averaging  $P_{i,i+1}(r)$  over all tag positions  $i$  for all cells within a cell line. We indicate this averaged histogram with  $\langle P_{i,i+1}(r) \rangle_i$ . From a two gaussian fit (5) of  $\langle P_{i,i+1}(r) \rangle_i$  we obtain global values for  $f_\alpha$ ,  $R_\alpha$  and  $R_\beta$  over all tags. Given now a measured distance  $r = r_{i,i+1}^{(k)}$  between tags  $i$  and  $i+1$  for a cell  $k$  we define the local probability to find the local chromatin conformation in

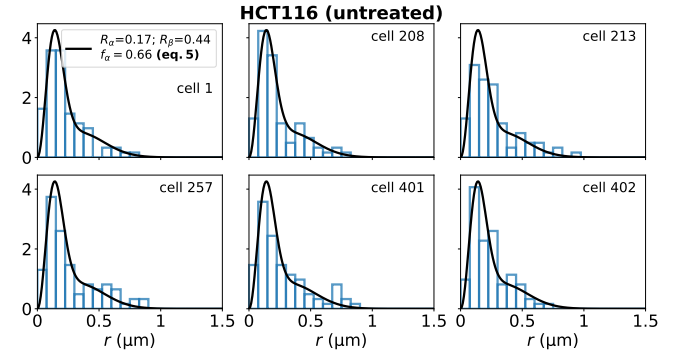


FIG. 8. Histograms show distance distributions of consecutive tags ( $|i-j|=1$ ) for individual cells from the HCT116 strain. Similar distributions are obtained for the Auxin treated HCT116 ensemble as well as other cell-types. Solid black line corresponds to a fit of  $\langle P_{i,i+1} \rangle_i$  with (5) for the untreated HCT116 cell line. Agreement of 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.

the  $\beta$  phase as

$$p_\beta(i) = \frac{(1-f_\alpha)g(r;R_\beta)}{\langle P_{i,i+1}(r) \rangle_i} \quad (6)$$

with  $g(r;R)$  as in (2). 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) averaged over 5 tags centered around the tag  $i$ . We thus compare the spatial variability (expressed in tag number) of these physical quantities. Interestingly, the figure shows that domains with a high  $\beta$ -phase content co-localize with maxima in the local radius of gyration, whereas clusters with much lower  $p_\beta$  are centered around the lower values of  $R_g$  thus with compact regions of the chromatin. The location of these different domains vary strongly from cell to cell, however, the variation of these two quantities is remarkably correlated. A further indication of the heterogeneous cell-to-cell behavior is given in Fig. 8 which shows plots of single cells histograms of spatial distances  $r$  over consecutive tags ( $|i-j|=1$ ). Unlike the histograms shown in Fig. 2 and 3, which are for fixed  $i$  and  $j$  and over all  $M \sim 10^3 - 10^4$  chromosomes, those in Fig. 8 are over just  $N \sim 60 - 80$  samples, which are the number of tags used in mFISH experiments (see Table I). Despite the lower statistics, the data of Fig. 8 show a distribution which is very similar to that of Fig. 2 and 3. These results confirm that there are at least two populations with distinct spatial distributions within each cell.

There is nowadays quite some consensus about the very heterogeneous chromatin organization which manifests itself in strong cell-to-cell variations. This conclusion is supported by experiments [4, 36, 37] and simulations [38, 39]. Our mFISH histogram analysis is in line with this view, with the advantage that it captures the heterogeneous organization in a simple way using general con-



cepts of polymer physics such as scaling laws and exponents. To reconcile discrepancies observed between some Hi-C and (non-multiplexed) FISH data Shi and Thirumalai [18] invoked the two phase model (5). They showed that these discrepancies can be solved 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 [37], therefore it could not reveal scaling laws that appear when several labels regularly spaced along the genome are used. We believe that the multiplexed FISH technology [4] has a lot of potential. Overall, our analysis indicates that the chromatin organization is described by two underlying fractional Brownian motions [27]. In the future, it would be interesting to extend this methodology to more cell lines and also explore chromosomal regions below the 30 Kb resolution.

### ACKNOWLEDGEMENT

Discussions with M. Barbi, T. Foldes, M. Liefsoens, M. Nollmann are gratefully acknowledged. LR acknowledges doctoral thesis support from the LabMUSE Epi-GenMed 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 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: 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. 6(b). The potential is attractive forcing the polymer to remain confined in the vicinity of the origin. One can consider different forms for  $U(r)$  as infinite well case

$$U_w(r) = \begin{cases} 0 & r < R \\ +\infty & r \geq R \end{cases} \quad (\text{A1})$$

a harmonic potential

$$U_h(r) = \frac{k}{2}r^2 \quad (\text{A2})$$

of other types of confinement. The probability distribution for the end point distance  $\vec{r}$  of a segment of length  $s$  satisfies the differential equation [40]

$$\frac{\partial \mathbf{P}(\vec{r}, s)}{\partial s} = \frac{b^2}{6r} \frac{\partial^2 [r \mathbf{P}(\vec{r}, s)]}{\partial r^2} - \beta U(r) \mathbf{P}(\vec{r}, s) \quad (\text{A3})$$

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

$$\mathbf{P}(\vec{r}, s) = \sum_p A_p \psi_p(\vec{r}) g_p(s) \quad (\text{A4})$$

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 [41]). 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  $s \rightarrow \infty$  one uses the ground state dominance (GSD) approximation which retains the component with the smallest  $\alpha_p$  [42]

$$\mathbf{P}(\vec{r}, s) \approx A_0 e^{-\alpha_0 s} \psi_0(\vec{r}) \quad (\text{A5})$$

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

$$\frac{b^2}{6r} \frac{\partial^2 [r \psi_0(\vec{r})]}{\partial r^2} - \beta U(r) \psi_0(\vec{r}) = -\alpha_0 \psi_0(\vec{r}) \quad (\text{A6})$$

We note that in the GSD approximation the distribution in  $\vec{r}$  is independent of the polymer length  $s$ , due to the factorization in (A5). 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 (A1) the solution of (A6) is a spherical Bessel function of order zero

$$\psi_0(\vec{r}) = \frac{\sin(\pi r/R)}{r} \quad \text{for } r \leq R \quad (\text{A7})$$

and  $\psi_0(r) = 0$  for  $r > R$ . For a confining harmonic potential (A2) one finds

$$\psi_0(\vec{r}) = \exp(-3r^2/2R_*^2) \quad (\text{A8})$$

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 (A5), 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

$$P(\vec{r}) = \int d\vec{r}_1 d\vec{r}_2 \psi_0(\vec{r}_1) \psi_0(\vec{r}_2) \delta(\vec{r}_1 - \vec{r}_2 - \vec{r}) \quad (\text{A9})$$

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

$$\delta(\vec{x}) = \frac{1}{(2\pi)^3} \int d\vec{k} e^{i\vec{k}\cdot\vec{x}} \quad (\text{A10})$$

we obtain

$$P(\vec{r}) = \int d\vec{k} \left[ \widetilde{\psi}_0(\vec{k}) \right]^2 e^{i\vec{k}\cdot\vec{r}} \quad (\text{A11})$$

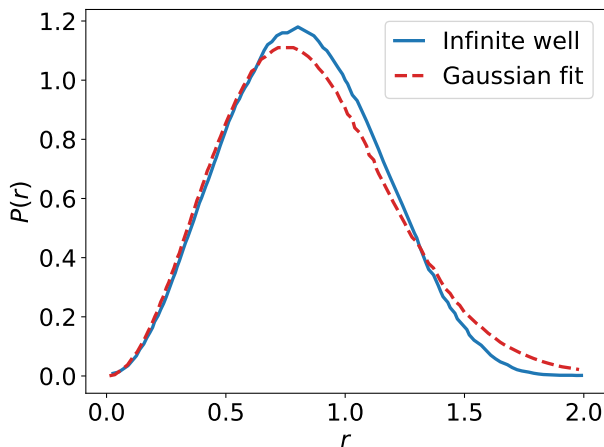


FIG. A.1. Solid line: Numerical solution of the probability distribution  $P(r)$  of spatial distance  $r$  between two points separated by a long genomic distance  $s$  for confinement in the case of infinite well-potential (A1) with  $R = 1$ . Dashed line: Gaussian fit of  $P(r)$ .

where  $\tilde{\psi}_0$  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

$$P_h(r) = 4\pi r^2 \left( \frac{3}{4\pi R_*^2} \right)^{3/2} e^{-3r^2/4R_*^2} \quad (\text{A12})$$

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 (A7) and calculated histograms of the distance  $r \equiv |\vec{r}_1 - \vec{r}_2|$ , which is shown in Fig. A.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 potential the distribution is not a Gaussian one.

We consider now the case of confinement of an ideal polymer in a flower-like shape conformation as in Fig. 6(c). For simplicity we will assume that each petal has the same length  $l$  and we number them  $1, 2, \dots, n$  according to increasing genomic position. Let us, again for simplicity consider two points in 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  $p_p(\vec{r}^*; l)$  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

$$P(\vec{r}) = \int d\vec{r}_1 d\vec{r}_2 p_p(\vec{r}_1; l) p_p(\vec{r}_2; l) \delta(\vec{r}_1 - \vec{r}_2 - \vec{r}) \quad (\text{A13})$$

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 important, this is independent of the genomic distance  $s$ , hence it predicts a scaling  $R \sim s^\nu$  with  $\nu = 0$ . This is for distances  $s > l$ , i.e. beyond the length of a petal.

## Appendix B: Fractions of $\alpha$ and $\beta$ phases

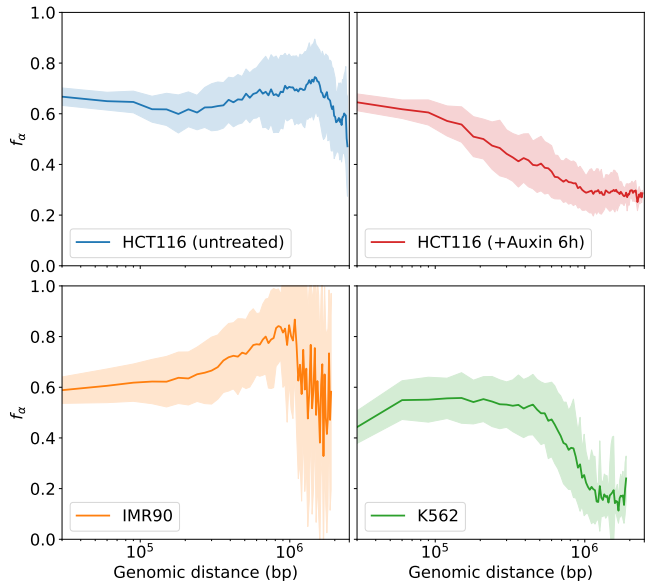


FIG. B.1. Plot of the average  $f_\alpha$ , the fraction of phase  $\alpha$ , as a function of the genomic distance (solid colored line). The light-colored area indicates the error estimate.

From fits of experimental histograms  $P_{ij}(r)$  with the two phase model (5) one obtains estimates of  $f_\alpha$  and  $f_\beta = 1 - f_\alpha$ , the fraction of the two phases for pairs of labels  $i$  and  $j$ . Figure B.1 shows the values of  $f_\alpha$  (central thick colored lines) averaged over all begin tags  $i$  as a function of genomic separation  $l_t|i - j|$ , where  $l_t = 30,000$  is the spacing in base pairs between two consecutive tags of the experiments. The colored areas estimates the variability (standard deviation) over the different data. For a microphase separation we expect a constant value of  $f_\alpha$  for a distance range corresponding to the characteristic length of the domains. The data in Fig. B.1 show that  $f_\alpha \approx 0.6$  for the shortest tags distance  $|i - j| = 1$  (except for K562 that, as discussed previously displays some anomalous behavior compared to the other cell lines). The data also show that  $f_\alpha$  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 histogram analysis is more reliable for tag distances which are not too distant ( $|i - j| \lesssim 5$ ) as the differences between  $R_\alpha$  and  $R_\beta$  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. Ragooczy, 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. Houbbron, 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. Houbbron, 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. Schiesse, *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. Lesage, V. Dahirel, J.-M. Victor, and M. Barbi, *Epigenetics & Chromatin* **12**, 28 (2019).
- [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] T. Földes, A. Lesage, and M. Barbi, *Phys. Rev. Lett.* **127**, 277801 (2021).
- [20] J. Dekker and L. Mirny, *Cell* **164**, 1110 (2016).
- [21] R. Blossey, *Chromatin: Structure, Dynamics, Regulation* (CRC Press, 2017).
- [22] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, *Cell Reports* **15**, 2038 (2016).
- [23] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, *Nucl. Acids Res.* **42**, 9553 (2014).
- [24] O. Oluwadare, M. Highsmith, and J. Cheng, *Biological Procedures Online* **21**, 1 (2019).
- [25] C. Cui, W. Shu, and P. Li, *Front. Cell. Dev. Biol.* **4**, 89 (2016).
- [26] P.-G. De Gennes, *Scaling concepts in Polymer Physics* (Cornell University Press, 1979).
- [27] K. Polovnikov, S. Nechaev, and M. Tamm, *Soft Matter* **14**, 6561 (2018).
- [28] J. L. A. Dubbeldam, V. G. Rostiashvili, A. Milchev, and T. A. Vilgis, *Phys. Rev. E* **83**, 011802 (2011).
- [29] J.-C. Walter, A. Ferrantini, E. Carlon, and C. Vanderzande, *Phys. Rev. E* **85**, 031120 (2012).
- [30] T. Sakaue, J.-C. Walter, E. Carlon, and C. Vanderzande, *Soft Matter* **13**, 3174 (2017).
- [31] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, *J. Phys. France* **49**, 2095 (1988).
- [32] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, *Europhys. Lett. (EPL)* **23**, 373 (1993).
- [33] A. Rosa and R. Everaers, *Phys. Rev. Lett.* **112**, 118302 (2014).
- [34] L. A. Mirny, *Chromosome Res.* **19**, 37 (2011).
- [35] J. D. Halverson, J. Smrek, K. Kremer, and A. Y. Grosberg, *Rep. Prog. Phys.* **77**, 022601 (2014).
- [36] 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).
- [37] 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).
- [38] A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, and N. Gilbert, *Molecular cell* **72**, 786 (2018).
- [39] M. Chiang, C. A. Brackley, C. Naughton, R.-S. Nozawa, C. Battaglia, D. Marenduzzo, and N. Gilbert, *bioRxiv* **06.09**, 495447 (2022).
- [40] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Vol. 73 (Oxford University Press, 1988).
- [41] P. Hahnfeldt, J. E. Hearst, D. J. Brenner, R. K. Sachs, and L. R. Hlatky, *Proc. Natl. Acad. Sci.* **90**, 7854 (1993).
- [42] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (American Institute of Physics, 1994).