bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hierarchical domain model explains multifractal scaling of chromosome contact maps

S. Pigolotti[1], M. H. Jensen[2] and G. Tiana[3]
[1] *Biological Complexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan.*[*]
[2] *The Niels Bohr Institute, Blegdamsvej 16, 2100 Copenhagen Ø, Denmark. and*
[3] *Department of Physics and Center for Complexity and Biosystems, Università degli Studi di Milano and INFN, via Celoria 16, 20133 Milano, Italy.*
(Dated: June 28, 2019)

Modern experimental techniques such as Hi–C make it possible to measure the probability that different chromosomal regions are close in space. Usually, these measurements are characterized by the scaling of the contact probability as a function of the genomic distance between regions. In this work, we introduce a multifractal analysis of chromosomal contact maps. Our analysis shows that Hi–C maps display a non-trivial multifractal spectrum. We introduce a simple analytical model that describes the structure of chromosomes as a hierarchical set of domains nested in each other and we solve it exactly. The predicted multifractal spectrum is characterized by a phase transition between two phases with different fractal dimension, in excellent agreement with experimental data. These results support the view that there is no privileged level in the hierarchy of conformational domains. Within the same model, the scaling exponent of the contact probability can also be calculated. We argue that such procedure leads to a much more precise estimate of the contact probability exponent than previous approaches. By applying this method to experimental data, we demonstrate that it can capture subtle conformational differences among chromosomes that are robust in different realizations of the same experiment.

I. INTRODUCTION

During cellular interphase, mammalian chromosomes assume a globular structure in the nucleus [1, 2]. Their conformational properties can be studied *in vivo* with a set of techniques called chromosome conformation capture [3], most notably their version called Hi–C [4]. Results of Hi–C experiments can be represented as matrices whose elements are proportional to the contact probability in space between pairs of chromosome regions [5].

A way to characterize the information contained in Hi–C matrices is by studying the decay of the contact probability p_{ij} of two chromosomal regions i and j with respect to their genomic distance $|i-j|$. Empirically, such decay is well described by a scaling law

$$p_{ij} \sim |i-j|^{-\beta}. \tag{1}$$

The fitted contact probability exponent for mouse chromosomes is $\beta \approx 0.75$ [6]. Such low value is incompatible with simple equilibrium homopolymeric models. An explanation proposed for such low exponent is that the chromosome fiber is not at thermodynamic equilibrium, but populates a 'crumpled globule' with $\beta \approx 1$ [4, 7]. Similarly low scaling exponents can also be obtained if the interaction in the polymer are mediated by other molecules [8] or by an active loop-extrusion mechanism [9]. Low exponents can also emerge in heteropolymers as consequences of finite–size effects, amplified by the heterogeneity of monomer interactions [10].

Another distinct feature of Hi–C matrices is the presence of blocks along the diagonal with large values of the contact probability. Physically, these blocks corresponding to structured regions of the chromosome with density larger than the surrounding. Historically, structures on the multi–megabase scale are called 'compartments', sub–megabase structures are called 'topological associating domains' (TADs) and those at the kilobase scale are called 'contact domains'. It was suggested that the physical interaction that stabilizes compartments is different than that which stabilizes smaller structures [11].

Despite this distinction, a quantitative analysis of Hi–C matrices points to a continuous hierarchy of structures nested into each other. Although the level of the hierarchy corresponding to TADs is biologically more important than the others because of its relevance for gene regulation [12], it does not seem to be privileged from the conformational point of view [13]. It has been recently suggested that, because of this hierarchical structure, the properties of Hi–C matrices can be studied by an iterative coarse-graining procedure [14].

In this paper, we perform a multifractal analysis of Hi–C matrices of mouse embryonic stem cells [5, 15]. The theory of multifractals has been developed to study heterogeneous systems characterized by scale invariance such as chaotic systems and turbulent flows [16–18]. The theory of multifractals describes the scaling of moments of the probability distribution of a system as a function of a coarse-graining resolution. This procedure leads to a family of scaling exponents (the *multifractal spectrum*) as a function of the moments. A linear multifractal spectrum indicates that the system is homogeneous, i.e. all of its parts are characterized by the same scaling behavior. Our analysis shows that the multifractal spectrum of mouse Hi–C data presents two different linear regimes, one at low moments with the properties of a two-dimensional object and one at high moments char-

[*] Electronic address: simone.pigolotti@oist.jp

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

acterized by a fractal dimension between one and two. In the theory of multifractals, such scenario is analogous to a phase transition in statistical physics, with the moments playing the role of an inverse temperature [19].

To understand this observation, we introduce a hierarchical domain model of Hi–C matrices, characterized by a fractal hierarchy of block domains along the diagonal. We calculate analytically the multifractal spectrum of the model. Strikingly, the analytical solution explains the phase transition and provides an excellent fit to the experimental multifractal spectrum. This suggests that the multifractal spectrum of Hi–C matrix directly reflects the nested organization of domains within the chromosomes. Within the model, we also calculate exactly the scaling exponent β. We find that the determination of β from the multifractal spectrum more robust than with conventional methods.

II. SCALING PROPERTIES OF HI–C MAPS

We illustrate the main idea of our work within the example of a Hi–C map of chromosome 1 in mouse embryonic stem cells [15], Fig. 1A. Darker red pixels in the figure represent higher contact probabilities. We seek to statistically characterize the information contained in this map.

A common way to approach this problem is to study the decay of the contact probability with the genomic distance, i.e. the distance from the diagonal [4, 7]. This approach is sketched in Fig. 1B for the same data. The decay of contact probability at distance $x \lesssim 10^6$ bp is well described by the power law of Eq. (1). The fitted value of the exponent β is in general smaller than one; in this example $\beta \approx 0.78$.

By carefully observing Fig. 1A, it is possible to discern domains of different sizes, sometimes nested into each other, corresponding to a hierarchy of domains in the chromosome. In principle, it is unclear whether this complex structure relates with the power-law decay of the contact probability in Fig.1B. To shed light on this issue, we take the original Hi–C map and construct a hierarchy of coarse-grained versions of it, Fig.1C. Such coarser maps are obtained by considering a grid of resolution ϵ and making a two-dimensional histogram of the Hi–C map. Geometrical structures of linear size smaller than ϵ are not resolved in these maps.

The key idea of a multifractal [17, 18] is to analyze scaling of moments of the coarse-grained Hi–C maps as a function of the mesh size ϵ. To this aim, we introduce a partition function

$$Z(q,\epsilon) = \sum_{ij} p_{ij}(\epsilon)^q \qquad (2)$$

where $p_i(\epsilon)$ is the density in bin i at resolution ϵ. For convenience, we always consider normalized maps, so that $Z(1,\epsilon) = \sum_{ij} p_{ij}(\epsilon) = 1$ for all choices of ϵ. In

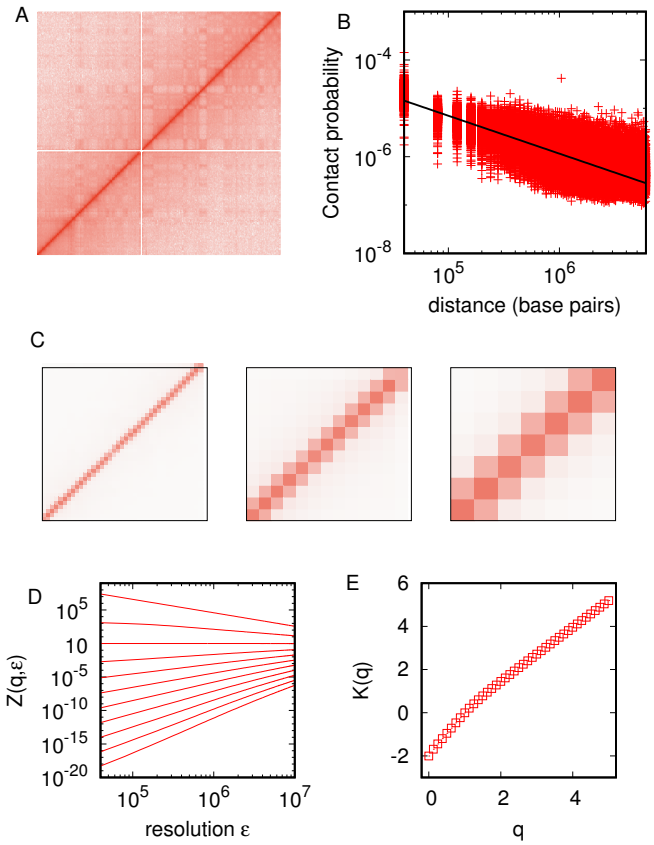

FIG. 1: A) Hi–C map for mouse chromosome 1 [15]. Darker shades of red correspond to higher values of the contact probability p_{ij} B) Scaling of the contact probability p_{ij} as a function of the distance $d = |i - j|$. The continuous line is a best-fit of a power law, Eq. (1), in the range of distances $d \in [10^5, 3\ 10^6]$, yielding $\beta \approx 0.78$. C) Sequential coarse-graining of the Hi–C map at increasing values of the resolution $\epsilon = 4\ 10^5, 1.2\ 10^6, 2\ 10^6$ bp. D) Scaling of $Z(q)$ as a function of resolution, see Eqs. (2) and (3). Different lines correspond to different values of q, increasing from top to bottom from $q = 0$ to $q = 5$ at intervals of $\Delta q = 0.5$. E) Corresponding multifractal spectrum $K(q)$. Here and throughout the paper, spectra are obtained by a fit in log-log scale of $Z(q,\epsilon)$ versus ϵ in the range $\epsilon \in [10^5, 3\ 10^6]$ unless specified otherwise.

practice, the smallest possible value of ϵ is the resolution of the experimental map, in our case $\epsilon = 4\ 10^4$ bp . The name "partition function" for $Z(q,\epsilon)$ originates from an analogy with equilibrium statistical mechanics, where the exponent q plays the role of an inverse temperature. Indeed in an inhomogeneous system, for $q \to 0$ (high temperature), all states give similar contributions to the sum in Eq. (2), whereas for large q (low temperature) the sum is dominated by relatively few terms characterized by highest values of the measure p_{ij}.

For a scale-invariant system, one expects a power-law scaling of the partition function as a function of the mesh size ϵ, at least for ϵ small enough

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$Z(q,\epsilon) \sim \epsilon^{K(q)} \qquad \epsilon \to 0. \tag{3}$$

We observed a beautiful power-law behavior as predicted by Eq. (3) in chromosomes, Fig.1D. The function $K(q)$ is the *multifractal spectrum* and characterizes the scaling behavior with resolution of the partition function [18]. To gain intuition about the multifractal spectrum, it is useful to first consider the simple case of a homogeneous, possibly fractal object of physical dimension D. Such object can be filled a number $N \sim \epsilon^{-D}$ of bins of size ϵ, where $\sim$ denotes the leading order for small ϵ. Since the object is homogeneous, such bins would have comparable weight. Normalization then implies that in the filled bins $p_{ij} \sim \epsilon^{D}$. Substituting these estimates into Eq. (3) we obtain

$$K(q) = D(q-1). \tag{4}$$

This means that homogeneous objects are characterized by a linear multifractal spectrum, with a slope equal to the fractal dimension. Conversely, a non-linear multifractal spectrum signals that the geometric set behaves in a different way depending on the weight attributed to the measure of each element. The multifractal spectrum of chromosome 1 is markedly nonlinear, Fig. 1E. In the following, we shall see that an exactly solvable model can account for all these observations in terms of a single free parameter.

III. HIERARCHICAL DOMAIN MODEL

To rationalize the multifractal spectrum of chromosomes, we introduce a hierarchical domain model able to generate synthetic contact matrices similar to the experimental ones. The model is defined in terms of an iterative transformation of a measure on the unit square $[0,1]\times[0,1]$. At each iteration n, the unit square is divided into smaller squares of size $\epsilon = 2^{-n}$. We use ϵ for the side of the square since this length plays the same role as the coarse graining resolution for experimental contact maps. Inside each of these squares the measure is uniform, so that the measure can be represented by a $2^n \times 2^n$ matrix whose entries are the probabilities p_{ij} in each square. The transformation is constructed in such a way that the measure remains normalized at each iteration. The first iteration reads

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix} \tag{5}$$

with $a > b > 0$. Normalization requires $b = 1/2 - a$. This means that, effectively, the model is defined by a single free parameter a. Because of the normalization and the requirement that the measure should be concentrated along the diagonal, such parameter falls in the range

$$\frac{1}{4} \le a \le \frac{1}{2}. \tag{6}$$

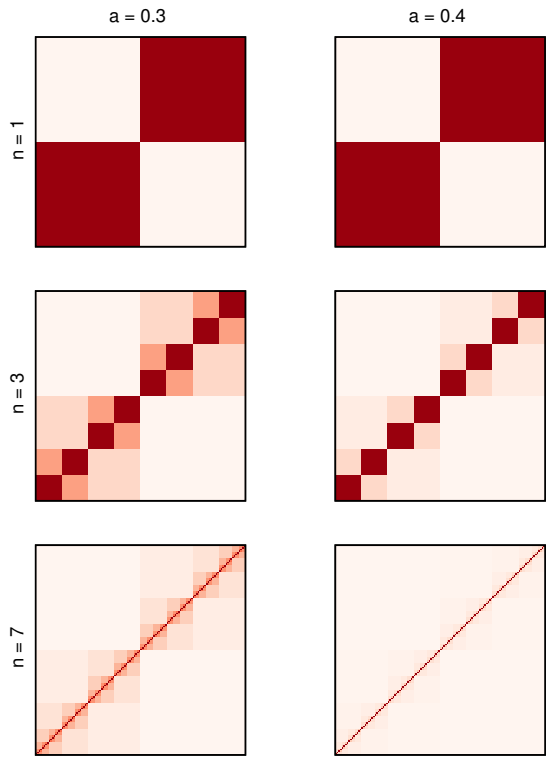

FIG. 2: Fractal model at different iterations n and for two different values of parameter a.

In each of the following iterations, the blocks along the diagonal are divided in 4 sub-blocks by multiplying by the same matrix of Eq. (5), whereas the off-diagonal blocks are divided in 4 identical sub-blocks. Following this procedure, the measure at the second iteration reads

$$\begin{pmatrix} a^2 & ab & b/4 & b/4 \\ ab & a^2 & b/4 & b/4 \\ b/4 & b/4 & a^2 & ab \\ b/4 & b/4 & ab & a^2 \end{pmatrix} \tag{7}$$

and the third iteration reads

$$\begin{pmatrix} a^3 & a^2b & ab/4 & ab/4 & b/16 & b/16 & b/16 & b/16 \\ a^2b & a^3 & ab/4 & ab/4 & b/16 & b/16 & b/16 & b/16 \\ ab/4 & ab/4 & a^3 & a^2b & b/16 & b/16 & b/16 & b/16 \\ ab/4 & ab/4 & a^2b & a^3 & b/16 & b/16 & b/16 & b/16 \\ b/16 & b/16 & b/16 & b/16 & a^3 & a^2b & ab/4 & ab/4 \\ b/16 & b/16 & b/16 & b/16 & a^2b & a^3 & ab/4 & ab/4 \\ b/16 & b/16 & b/16 & b/16 & ab/4 & ab/4 & a^3 & a^2b \\ b/16 & b/16 & b/16 & b/16 & ab/4 & ab/4 & a^2b & a^3 \end{pmatrix} \tag{8}$$

Physically, the parameter a represents the "weight" of TADs compared to the rest of the Hi–C map. Matrices with larger values of a have a measure more concentrated along the diagonal, whereas matrices with lower a are characterized by a more uniform measure. In particular,

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in the limiting case $a = 1/4$ the matrix is completely uniform at all iterations, whereas for $a = 1/2$ the matrix is characterized by a uniform measure on the diagonal and all the off-diagonal elements are zero. Examples of measures after different numbers of iterations are shown in Fig. 2.

A. Multifractal scaling

We now write the partition function $Z(q, n)$, corresponding to the definition of Eq. (2) with $\epsilon = 2^{-n}$. For example, we have

$$\begin{aligned} Z(q,1) &= 2a^q + 2b^q \\ Z(q,2) &= 4(a^2)^q + 4(ab)^q + 8(b/4)^q \\ Z(q,3) &= 8(a^3)^q + 8(a^2b)^q + 16(ab/4)^q + 32(b/16)^q \\ Z(q,4) &= 16(a^4)^q + 16(a^3b)^q + 32(a^2b/4)^q + \\ &+ 64(ab/16) + 128(b/64)^q. \end{aligned} \tag{9}$$

The general expression of the partition function is

$$Z(q,n) = \sum_{k=0}^{n} 2^{[n+\max(n-k-1,0)]} \left[\frac{a^k b^{(1-\delta_{kn})}}{4^{\max(n-k-1,0)}} \right]^q \tag{10}$$

where δ_{kn} is the Kronecker delta. We express the partition function in the form

$$Z(q,n) = \sum_{k=0}^{n} \exp[\xi(k)] \tag{11}$$

where we defined the exponent

$$\begin{aligned} \xi(k) &= [n + \max(n-k-1,0)] \ln 2 + kq \ln a + \\ &+ (1-\delta_{kn}) q \ln b - \max(n-k-1,0) q \ln 4. \end{aligned} \tag{12}$$

We now seek for the maximum of the exponent. Assuming $n - k - 1 \geq 0$ and deriving the exponent with respect to k we obtain

$$\frac{d\xi}{dk} = -\ln 2 + q \ln 4a \tag{13}$$

which is positive for $q > q_c = (\ln 2)/(\ln 4a)$. This means that, for $q \geq q_c$, the leading term is either $\xi(n)$ or $\xi(n-1)$. Since $\xi(n) - \xi(n-1) = q \ln(a/b)$, the maximum of the exponent is attained at $k = n$.

Fr large n, the partition function of Eq. (11) is dominated by the term corresponding to the maximum of the exponent. We therefore substitute, for large n, $Z(q,n) \sim \exp[\xi(n)] \sim \epsilon^{K(q)}$ with the length scale $\epsilon = 2^{-n}$. We directly obtain

$$2^n a^{nq} \sim 2^{-nK(q)} \tag{14}$$

from which

$$K(q) = -q \frac{\ln a}{\ln 2} - 1. \tag{15}$$

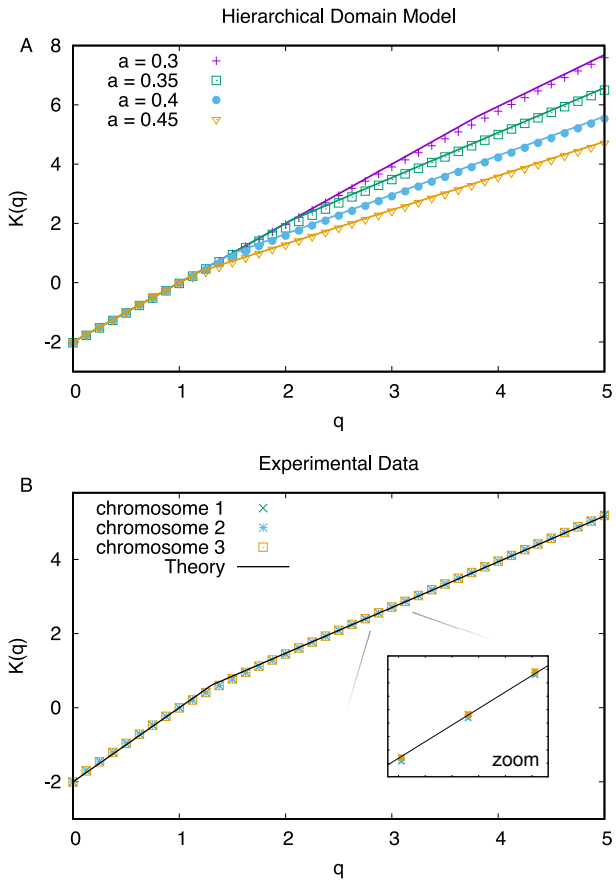

FIG. 3: Multifractal spectra of the hierarchical domain model and experimental Hi–C are both quantitatively predicted by our theory. A) Multifractal spectra of the model, numerically calculated for $n = 10$ iterations of the model, for different values of a shown in the caption. The lines are theoretical predictions of Eqs. (15) and (17). Small discrepancies for low values of a and high moments are due to finite-size effects. B) Multifractal spectra of the first three chromosomes. The spectra of the three sets of experimental data are quantitatively very similar and in excellent agreement with the theory (continuous line) with a value of $a \approx 0.425$ The agreement can be better appreciated in a zoom of the data (inset).

For $a = 1/2$ the distribution is linear and we have $D = 1$. For $a = 1/4$ the distribution is uniform on the square, so that $D = 2$. Between these two limiting cases, Eq. (15) predicts that the distribution is monofractal with a dimension $D = -\ln a / \ln 2$ between 1 and 2.

Let us now figure out the scaling in the "high temperature phase" where $-\ln 2 + q \ln 4a < 0$. In this case, the term dominating the scaling is $k = 0$, so that

$$2^{2n-1} b^q 4^{(1-n)q} \sim 2^{-nK(q)} \tag{16}$$

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which implies

$$K(q) = 2(q-1). \tag{17}$$

This means that, since the scaling is determined by the terms far from the diagonal in the high temperature phase, the spectrum is that of a regular two-dimensional set.

Summarizing, the saddle-point calculation predicts a multifractal spectrum characterized by two linear regimes: one with slope $-\log(a)/\log(2)$ for $q < q_c = (\ln 2)/(\ln 4a)$, Eq. (15), and one with slope 2 for $q < q_c$, Eq. (17). Such predictions are very well confirmed by numerical simulations, Fig. 3A, with very small discrepancies for high values of q and low values of a arising from finite size effects. These results confirm the validity of our saddle point approximation.

Strikingly, our theory predicts extremely well also the multifractal spectra of real chromosomes, as illustrated in Fig. 3B for the first three chromosomes, with a value of $a \approx 0.425$ and very little variability among the three chromosomes. We will examine in more details the differences among chromosomes in the next section.

To verify whether the multifractal spectrum in Fig. 3 is truly a signature of a hierarchical mechanism and not a more generic statistical feature, we numerically computed the spectra of two null models: one characterized by a power-law decay plus blocks at a single scale $p_{ij} = |i-j|^{-0.8} + b_{ij}$ (where b_{ij} is 1 if i and j belongs to the same block of linear size 64 and zero otherwise), and another characterized by the same power law plus white noise with zero mean and standard deviation 0.1. In the latter case, the noise is truncated to ensure that probabilities are always non-negative. The solution of Eqs. (15) and (17) provides a much better fit to the hierarchical model (sum of residuals ≈ 0.04) than either block model (sum of residuals ≈ 0.8) or the power law plus noise model (sum of residuals ≈ 0.3), data not shown. The sums of residuals for the fits to the first three chromosomes are in order 0.049, 0.051 and 0.043. Taken together, these observations support the idea that the specific multifractal spectrum of the hierarchical model is compatible with that observed in chromosomes and significantly different than that of other simple models that are lacking a hierarchy of domains.

B. Scaling of the contact probability

We now study the decay of the contact probability with the genomic distance $d = |i-j|$. In our framework, the contact probability $P(d;n)$ at the n-th iteration can be expressed as the total probability of blocks at a distance from the diagonal equal to d

$$P(d;n) = \sum_{ij} p_{ij}(n)\delta_{|i-j|,d}. \tag{18}$$

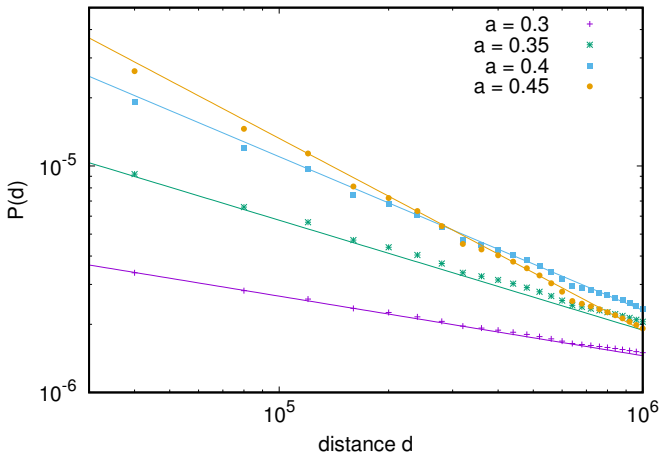

FIG. 4: Decay of contact probability $P(d;n)$ as a function of the genomic distance d in the hierarchical model for $n = 10$ and different values of the parameter a, shown in the figure legend. Continuous lines are the exponent predictions of Eq. (25).

To find an explicit form for $P(d;n)$ we use our expression of the partition function, Eq. (11), for $q = 1$. The only difference between $P(d;n)$ and $Z(1,n)$ is the Kronecker delta in (18), that selects a subset of terms contributing to the partition function. In particular, we note from the structure of the matrix that each term with a given power of a first increases linearly with d up to a maximum, then decreases linearly to zero. Following this observation, we introduce the triangular function

$$g(x) = \begin{cases} x & \text{if} \quad 0 \le x \le \frac{1}{2} \\ 1-x & \text{if} \quad \frac{1}{2} \le x \le 1 \end{cases}. \tag{19}$$

Using this definition permits to write $P(d;n)$ in the form

$$P(d;n) = \sum_{k=0}^{n-1} 2^n g(d\ 2^{k-n}) \frac{a^k b^{(1-\delta_{kn})}}{4^{\max(n-k-1,0)}}. \tag{20}$$

We now approximate the sum with an integral

$$\begin{aligned} P(d;n) \approx & \int_{-\infty}^{n-\frac{\ln 2d}{\ln 2}} dk\ d \exp\left[k \ln 2 + k \ln a + \right. \\ & + \left. (1-\delta_{kn}) \ln b - \max(n-k-1,0) \ln 4\right] \\ & + \int_{n-\frac{\ln d}{\ln 2}}^{n-\frac{\ln 2d}{\ln 2}} dk\ \exp\left[n \ln 2 + \ln\left(1 - d\ e^{(k-n)\ln 2}\right)\right. \\ & + \left. k \ln a + (1-\delta_{kn}) \ln b - \max(n-k-1,0) \ln 4\right] \end{aligned} \tag{21}$$

In the first integral, the derivative of the exponent with respect to k is equal to $\ln(8a)$ and is therefore always positive due to Eq. (6). The derivative of the exponent in the second integral is equal to

$$\ln(4a) - \frac{d2^{(k-n)}\ln(2)}{1 - d2^{k-n}}. \tag{22}$$

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which vanishes at

$$d\ 2^{(k^*-n)} = \frac{\ln(4a)}{\ln(4a) + \ln 2}. \tag{23}$$

One has $1/2 < d\ 2^{(k^*-n)} < 1$, so that the maximum of the exponent falls within the integration domain. Approximating the integral with the maximum of the integrand yields

$$P(d;n) \sim d^{-\frac{\ln(4a)}{\ln(2)}} \tag{24}$$

for large n. This implies that the contact probability scales with the exponent

$$\beta = \frac{\ln(4a)}{\ln(2)}. \tag{25}$$

In the allowed range of a given by Eq. (6), the contact probability exponent predicted by Eq. (25) falls in the range $\beta \in [0, 1]$. The prediction of Eq. (25) is supported by numerical simulations, see Fig. 4. This confirms the validity of our saddle-point approximation.

IV. ESTIMATES OF CONTACT PROBABILITY EXPONENT

An important practical consequence of our theory is that it provides an alternative method to estimate the contact probability exponent β via Eq. (25). This requires estimating the scaling exponents of the partition function. In practice, the partition function exhibits a much cleaner scaling than the contact probability, as seen by comparing Fig. 1D with Fig. 1B. In this Section we scrutinize in more details the robustness of fits of the exponent β with the two methods. We also compare the experimental results in [15] with two different realization of a more recent experiment [5].

For both methods, the value of β is affected by the choice of the power-law fitting range. We fixed a lower cutoff for the power law fits equal to 10^5 bp, slightly higher than the minimum resolution of the data, and plotted the prediction of the exponent β as a function of the upper cutoff, Fig. 5A. We found that exponent displays a stronger dependence on the upper cutoff for the direct power law fit than for the multifractal method. Moreover, comparing different experiments [5] provides quite similar values of the exponents β when the multifractal method is used, whereas a large variability is observed with the power law fit.

Such more precise inference of the value of β can be appreciated when comparing results of different measures of Hi–C exponents. The multifractal method predicts much less variability of the exponent β and much more consistent values across different experiments, with an average $\beta \approx 0.77$, Fig. 5B. Comparing different replicates of the same experiment, the exponents predicted by the multifractal method present a higher correlation (Pearson $r \approx 0.998$) compared with those obtained by a direct power law fit ($r \approx 0.994$), Fig. 5C.

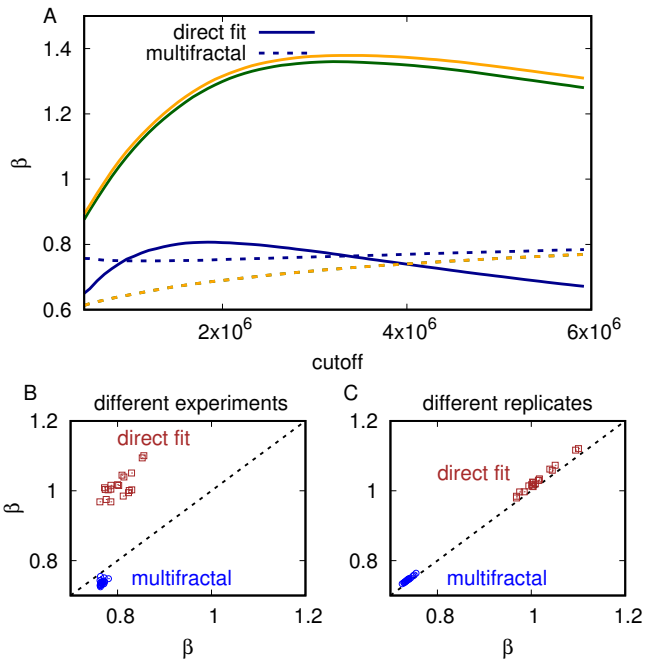

FIG. 5: A) (continuous lines) Power law fit of the first chromosome as a function of the upper cutoff of the power law fit, shown in the x axis. The three different colors correspond to different experiments (brown: data from [15]; green and orange: two different experimental realizations from [5]). Dashed lines denote the same exponents obtained with the multifractal method. In this case, the exponent β is obtained from the parameter a via Eq. (25) and the cutoff refers to the range of ϵ for the power-law fit of $Z(q, \epsilon)$. The lower cutoff of all fits is equal to 10^5. The dashed green curve is barely visible since the difference with the yellow curve is negligible at this scale. (B) Comparison of the estimated values of the exponent β in mouse chromosomes from the experiment in [15] with those from the experiment in [5]. Each point is a different chromosome. (C) Comparison of the estimated values of the exponent β in mouse chromosomes from two independent replicates of the same experiments [5]. In panels B and C, the upper cutoff equal to 3×10^6 bp for the multifractal method and 5×10^5 for the power-law fit.

V. CONCLUSIONS

In this paper, we introduced a multifractal analysis of Hi–C data as a powerful statistical tool to characterize the scaling properties of chromosomes. We have shown that the multifractal spectrum computed for chromosomes of mouse embryonic stem cells has a broken–line shape. To understand this observation we introduced a hierarchical domain model, whose analytical solution is

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in striking quantitative with the experimental spectrum of in terms of a single free parameter.

Our solution directly implicates a power–law scaling of the contact probability with the genomic distance with exponents β lower than one, also in agreement with experimental observations. A comparison with other null models characterized by similar contact probability exponents but without a hierarchy of domains shows that the multifractal method is sensitive enough to discard them. This implies that the predicted form of the multifractal spectrum provides a more stringent prediction than the contact probability exponent alone. The analysis proposed here constitutes an excellent benchmark to select among different polymer models that provide similar values of β [7–10].

Our results provide a strong indication that scaling properties of chromosomes are a direct consequence of the hierarchical structure of chromosome domains [13, 22]. Recent work has suggested that such domain structure is generated by a "hierarchical folding" mechanism, mediated by different proteins such as cohesin and CCCTC-binding factors [23]. The precise mechanism underlying folding of these domains has been subject of debate [24]. It will be interesting to test whether the activity of these factors can produce self-similar structures compatible with our observations.

The shape of the multifractal spectrum is controlled by a single parameter, that also controls the contact probability exponent β via Eq. (25). We demonstrated that the determination of β from the fit of the multifractal spectrum is a much more robust way of characterizing the Hi–C map than the direct fit of β, highlighting structural difference among chromosomes that we consistently observed in different realizations of the same experiment. The analysis of the multifractal spectrum is simple and robust enough to become a routinely tool to analyze contact maps, both from polymer models and experiments.

Acknowledgments

We thank Yinxiu Zhan and Luca Giorgetti for help with the data.

[1] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam *et al.*, *Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription.*, Cell **157**, 950 (2014).

[2] G. Tiana, A. Amitai, T. Pollex, T. Piolot, D. Holcman *et al. Structural Fluctuations of the Chromatin Fiber within Topologically Associating* , Biophys J. **110**, 1234 (2016)

[3] J. Dekker, K. Rippe, M. Dekker and N. Kleckner, *Capturing chromosome conformation*, Science, **295**, 1306 (2002).

[4] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling *et al.*, *Comprehensive mapping of long-range interactions reveals folding principles of the human genome*, Science **326**, 289 (2009).

[5] J. Redolfi, Y. Zhan, C. Valdes, M. Kryzhanovska, I. Misteli Guerreiro *et al.*, *Modeling of DNA methylation in cis reveals principles of chromatin folding in vivo in the absence of crosslinking and ligation*, bioRxiv https://doi.org/10.1101/407031 .

[6] B. Bonev, N. M. Cohen Q. Szabo, L. Fritsch, G. L. Papadopoulos *et al.*, *Multiscale 3D Genome Rewiring during Mouse Neural Development* Cell **171**, 557.e1 (2017).

[7] L. Mirny, *The fractal globule as a model of chromatin architecture in the cell*, Chrom. Res. **19**, 37 (2011).

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

[9] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur and L. Mirny, *Formation of Chromosomal Domains by Loop Extrusion*, Cell Rep. **15**, 2038 (2016).

[10] Y. Zhan, L. Giorgetti and G. Tiana, *Looping probability of random heteropolymers helps to understand the scaling properties of biopolymers.*, Phys. Rev. E **94**, 032402 (2016).

[11] J. Nuebler, G. Fudenberg, M. Imakaev, N Abdennur and L. A. Mirny, *Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation*, Proc. Natl. Acad. Sci. USA **115**, E6697 (2018).

[12] D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki *et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions*, Cell **161** 1012 (2015) http://doi.org/10.1016/j.cell.2015.04.004

[13] Y. Zhan, L. Mariani, I Barozzi, E. G. Schulz, N. Blüthgen *et al.*, *Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes*, Genome Research **27**, 479 (2017).

[14] A. M. Chiariello, S. Bianco, C. Annunziatella, A. Esposito and M. Nicodemi, *The scaling features of the 3D organization of chromosomes are highlighted by a transformation la Kadanoff of Hi-C data*, Europhys. Lett. **120** 40004 (2017).

[15] L. Giorgetti, B. R. Lajoie, A. C. Carter, M. Attia, *et al.*, *Structural organization of the inactive X chromosome in the mouse.*, Nature **535**, 575 (2016).

[16] G. Parisi and U. Frisch in "Turbulence and Predictability in Geophysical Fluid Dynamics", p. 8487, eds. M. Ghil et al (1985).

[17] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, *On the multifractal nature of fully developed turbulence and chaotic systems*, Journal of Physics A **17**, 3521 (1984).

[18] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B. Shraiman, *Fractal measures and their singularities: The characterization of strange sets*, Phys. Rev. A **33**, 1141 (1986).

[19] T. Bohr and M.H. Jensen, *Order parameter, symmetry breaking, and phase transitions in the description of multifractal sets* Phys. Rev. A **36**, 4904 (1987).

bioRxiv preprint doi: https://doi.org/10.1101/686279; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[20] G. Tiana and L. Giorgetti, *Integrating experiment, theory and simulation to determine the structure and dynamics of mammalian chromosomes*, Curr. Opin. Struct. Biol. **49**, 11 (2018).

[21] Y. Zhan, L. Giorgetti and G. Tiana, *Modelling genome-wide topological associating domains in mouse embryonic stem cells*, Chromosome Res. **25** 5 (2017).

[22] S. S. P. Rao, M. H. Huntley, N. C. Durand, *et al*, *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping* Cell, **159** 1665 (2014).

[23] B. Bonev and G. Cavalli, *Organization and function of the 3D genome*, Nat. Rev. Gen. **17**. 661 (2016).

[24] M. J. Rowley and V. G. Corces, *Organizational principles of 3D genome architecture*, Nat. Rev. Gen. **19**, 789 (2018).