Assessing the polymer coil-globule state from the very first spectral modes

Timothy Földes^{1,*}, Antony Lesage², and Maria Barbi^{1†}
¹ Sorbonne Université, CNRS, LPTMC, F-75005 Paris, France and
² Sorbonne Université, CNRS, Phenix, F-75005 Paris, France.

(Dated: July 17, 2021)

The determination of the coil-globule transition of a polymer is generally based on the reconstruction of scaling laws, implying the need for samples from a rather wide range of different polymer lengths N . The spectral point of view developed in this work allows for a very parsimonious description of all the aspects of the finite-size coil-globule transition on the basis of the first two Rouse (cosine) modes only, shedding new light on polymer theory. Capturing the relevant configuration path features, the proposed approach enables to determine the state of a polymer without the need of any information about the polymer length or interaction strength. Importantly, we propose an experimental implementation of our analysis that can be easily performed with modern fluorescent imaging techniques, and would allow differentiation of coil or globule conformations by simply recording the positions of three discernible loci on the polymer.

The coil-globule phase transition is a characteristic of the physics of polymers in solution, with direct consequences on macroscopic solution properties as viscoelasticity and transport features. This framework also successfully applies to several biological processes, from protein denaturation to local chromatin density, in which biopolymers compactness abruptly changes under environmental changes. Its widespread applications and deep theoretical relevance justify the relentless effort of both theorists and experimentalists in its understanding and observation. The coil-globule transition is usually studied by means of the scaling properties of the radius of gyration as a function of the polymer length, or number of monomers N . However, this approach requires comparing polymers of different lengths, which is not always possible. Even the determination of the number of monomers N is not always easy: typically, this number depends on the Kuhn length of the polymer, hence its bending persistence, which can be challenging in the case of real polymers, and a fortiori for complex biological polymers. Moreover, important deviations from the theoretical scaling are induced due to polymer finite-size [1]. An alternative but less direct approach is the study of polymer structure factor and its scaling law as a function of the wave vector in the large q limit [2]. We note however that both the radius of gyration and the structure factor are averages over the inter-monomer distances and, therefore, lose information about the precise configuration of the chain in space, which suggests the interest of having another approach that would preserve this information.

From a different perspective, the dynamical behavior of the polymer can be studied. Typically, the dynamics of single monomers can be compared to theoretical predictions such as those of the Rouse or Zimm models [3, 4]. Now, interestingly, the Rouse model uses a spectral mode decomposition. By noticing that this decomposition applies to any given conformation of the chain, we used it to determine the power spectral density as to inquire

equilibrium properties of the polymer, rather than its dynamics. Since it contains in principle all the information on the folding state of the polymer, this representation in the reciprocal space can moreover been exploited in the study of the coil-globule transition. The aim of this paper is to investigate this issue and to see how a spectral representation can facilitate the determination of the polymer state (i.e. its macrostate).

By looking in detail at the spectral behavior of polymers close to the coil-globule transition from on-lattice Monte Carlo simulations, we developed a method to discriminate the folding state of a polymer of a given size N , without the need to study the scaling law or to have access to the energy parameter. Our method exploits the information about the spatial path of the polymer, and only relies on *low* spectral modes, i.e. long distance features. As we will show, indeed, the very first modes contain the main information concerning the polymer state. This allows us to propose a parsimonious and efficient method, introducing a new N -independent order parameter that provides us with a robust measurement to identify the state of the polymer on the coil to globule scale. Most importantly, based on this analysis, we are able to propose an innovative *experimental* approach to determine the equilibrium state of a polymer from data obtained by fluorescence imaging, provided that a minimum of three equally spaced loci can be discernibly labeled, and we provide instructions on how to use these data.

The Rouse model describes an ideal chain, i.e. a chain with no excluded volume nor attractive interaction. In this case, the radius of gyration scales as the variance of a random walk, $\langle R^2 \rangle \sim N$. If hydrodynamic interactions can be neglected, the chain dynamics can be described by applying the overdamped Langevin equation to a polymer of $N + 1$ monomers, whose positions are denoted by $\{\vec{r}_n, n = 0 \dots N\}$, immersed in a fluid of viscosity γ at temperature T [5]. Subsequent monomers n and $n + 1$ are linked by a bonding harmonic potential

FIG. 1. (a) On lattice MC simulation snapshots for $N = 431$ and $\varepsilon = 0$ (orange), $\varepsilon = 0.36$ (black) and $\varepsilon = 0.8$ (blue). (b) Projections $r_x(n)$ along the spatial x -axis of a set of 20 configurations, n being the monomer index. Colors refer to the same parameters as in (a). Dotted lines correspond to the $\pm 2.5 \sigma_x(n)$ envelopes. An arbitrary vertical shift is introduced for clarity. (c) Logarithm of the square DCT-transform of the 20 trajectories of (b), $\ln |X_p|^2$, as functions of $\ln p$ (thin lines) and their average calculated over the ensemble of 16384 simulated configuration and over the three spatial directions (thick lines and dots). Standard deviations are smaller than the symbol size. Groups of curves are vertically shifted for aim of clarity. The dashed lines correspond to slopes $-2.2 \equiv -(1+2\nu)$, -1.3 and 0 (from top to bottom). (d) Averaged PSD obtained by applying the same procedure as in (c) to equivalent sets of simulated polymers with ε going from 0 to 0.6 (see colorbar). The dashed lines correspond to slopes $-(1+2\nu)$ (top) and 0 (bottom). **Inset:** The slope $\alpha_N(\varepsilon)$ of the log-log representation of the PSD, estimated from the first 2 modes, as a function of the energy parameter ε .

$U_{n,n+1} = -\frac{1}{2}k(\vec{r}_n - \vec{r}_{n+1})^2$, where the spring constant $k = 3k_B T/b^2$ sets the mean bond length to b . The application of the Langevin equation yields a system of $N+1$ three-dimensional coupled equations of the form

$$\frac{d\vec{r}_n}{dt} = \frac{3k_B T}{\gamma b^2} (-\vec{r}_{n-1} + 2\vec{r}_n - \vec{r}_{n+1}) + \vec{\xi}_n(t), \quad (1)$$

where the random forces $\xi_n(t)$, mimicking the effect of thermal fluctuations, satisfy the usual conditions [6]. To solve the model, the system is diagonalized (decoupled) by introducing the new set of variables $\{\vec{X}_p, p = 0 \dots N\}$

$$\vec{X}_p = \frac{1}{N+1} \sum_{n=0}^N \vec{r}_n \cos\left(\frac{p\pi}{N+1} \left(n + \frac{1}{2}\right)\right) \quad (2)$$

called *Rouse modes* of the polymer and equivalent to the standing modes of an oscillating string with one fixed end. Note that the $p = 0$ mode corresponds to the position of the polymer center of mass. In the following, we will always focus on modes $p = 1 \dots N$ only. Usually, one uses the auto-correlation function of $\vec{X}_p(t)$ to derive single monomer dynamical scaling laws, overlooking the potentialities of the Rouse decomposition in itself. Indeed, in the language of signal processing, the projection (2) is a Discrete Cosine Transform (DCT), often used for its high performances in image compression. It is closely related to the Discrete Fourier Transform (DFT), in that it corresponds to the DFT of a symmetrized signal of double length [7]. Hence, the mean square mode amplitudes give (up to a multiplicative constant) the *power spectral density* (PSD) of the polymer configuration. For

the Cartesian coordinate x (and equivalently y or z) [5]

$$\langle X_{p,x}^2 \rangle = \frac{b^2}{24(N+1) \sin^2\left(\frac{p\pi}{2(N+1)}\right)} \underset{p \ll N}{\sim} \frac{1}{p^2}, \quad (3)$$

yielding an explicit expression for the PSD of an ideal polymer, and giving an asymptotic power law behavior with exponent -2 .

The connection between Rouse modes and power spectral density seems, surprisingly, to have never been noticed before. However, the asymptotic dependence (3) has an immediate interpretation. Ideal chain configurations are indeed equivalent to Brownian motion trajectories [8], and it can be shown that the corresponding PSD has scaling exponent -2 . It is moreover noteworthy that *exactly* the same expression (3) is obtained in a different setting by Krapf *et al.* for the PSD of a single discrete Brownian trajectory [9].

Now, if the above description holds for an ideal polymer, a real polymer is characterized by interactions between monomers: repulsive excluded volume and (solvent dependent) attractive interactions. The case where only excluded volume is present defines the (non-interacting) *self-avoiding polymer* model, i.e. the prototypical coil polymer. Due to steric hindrance, the polymer occupies more space resulting in swollen, elongated conformations analogous to self-avoiding walks (SAW). The polymer's radius of gyration then scales as $\langle R^2 \rangle \sim N^{2\nu}$, where $\nu \approx 0.588$ is the Flory exponent [8]. Including steric interactions adds non-linear terms to the set of equations (1). In this case, the Rouse modes no longer uncouple the equations of motion and, consequently, there is no explicit solution for the PSD [10]. Yet simple arguments provide an insight into the general trend of the

spectrum. A usual (yet unproven) assumption about SAWs is their fractal nature, meaning that their statistical properties are scale invariant [8, 11, 12]. Hence, we expect a power law PSD, intuitively related to the presence of long range correlations induced by steric interactions [13]. The Hurst exponent provides an estimation of the long-range memory of a signal. By guessing a long range correlated process with a Hurst exponent H such that $\langle r_n^2 \rangle \propto n^{2H}$, the comparison with the Flory's behavior would lead to the ansatz $H = \nu$ which appoints the PSD

$$\langle \tilde{X}_p^2 \rangle \propto p^{-(1+2\nu)} \quad (4)$$

where $-(1+2\nu) \approx -2.2$. This is the first prediction to be tested. Interestingly, Panja and Barkema made an equivalent ansatz for the variance of the long wavelengths Rouse modes of the self-avoiding polymer, and corroborate it by extensive on-lattice simulations [10].

Depending on the physico-chemical properties of the polymer and the solvent, an effective attraction J between monomers may also exist. The state of the polymer thus depends on the relative strength of inter-monomer forces and thermal energy, $\varepsilon = J/k_B T$. In the thermodynamic limit $N \rightarrow \infty$, above a critical value ε_θ (θ -point), the strong attraction induces a second order phase transition to curled up conformations of uniform density called globules, whose typical scaling is $\langle R^2 \rangle \propto N^{2/3}$ [8]. However, *finite-size* polymers undergo a smooth coil-globule transition at a N -dependent critical energy. Due to the fact that there are fewer monomers in *finite-size*, then more energy per contact is required to have globular conformations. Thus, the transition energy ε_{θ_N} is all the higher that the polymer is small: $\varepsilon_{\theta}(N) > \varepsilon_\theta$. [2, 14–16].

Coils and globules can therefore in principle be distinguished by their radius of gyration but, as mentioned above, this average quantity does not provide information on the spatial configuration of the polymer, to which the PSD is intrinsically related. The latter can be described by noting with Grosberg and Khokhlov that a collapsed polymer is equivalent to an ideal polymer compressed within a spherical volume of radius R [8]. Over small length scales, the polymer behave like a free chain, until it reaches the volume boundary, where it is reflected back and starts an independent random walk. The chain can thus be described as a series of independent subchains. The position of two monomers in different sub-chains is then uncorrelated [8], so that the sequence of positions of monomers sufficiently spaced along the chain is a white noise signal. From the spectral point of view, this implies a constant spectrum for small values of p . We can thus predict that the PSD of globules verifies the property

$$\langle \tilde{X}_p^2 \rangle \propto p^0 \quad \text{for small } p. \quad (5)$$

This is the second hypothesis which we now wish to verify.

FIG. 2. Top panel: Log-log plot slope $\alpha_N(\varepsilon)$ estimated on the first 2 modes of PSD (dots) for all polymers sizes N from 101 to 3000 (right to left). For each N , $\alpha_N(\varepsilon)$ is fitted by a sigmoid $S(x) = D + A(x - B)/\sqrt{1 + C(x - B)^2}$ (lines). Colored circles correspond to inflection points, defining $\varepsilon^I(N)$. Blue diamonds indicates the $\alpha_N(\varepsilon) = -2$ condition, defining $\varepsilon_\theta^{RW}(N)$. **Inset:** $\alpha_N(\varepsilon)$ relative standard deviation, $\Delta\alpha_N(\varepsilon)/\alpha_N(\varepsilon) = \sigma_{X_1^2}/\langle X_1^2 \rangle + \sigma_{X_2^2}/\langle X_2^2 \rangle$, used to define the large fluctuation transition curve. Triangles indicate the position of maxima $\varepsilon^F(N)$, determined by fitting the top of the pics by a Gaussian function. **Bottom panel:** Corresponding phase diagram in the (ε, N) phase space. The orange to blue colorscale correspond to values of $\alpha_N(\varepsilon)$ from -2.2 to 0 (see colorbar). Inflection points (colored circles) and fluctuation maxima (colored triangles) are shown and fitted by curves of expression $\varepsilon_\theta(N) = a_1 N^{-1/2} + a_2 (2N)^{-1} + \varepsilon_\theta$ (green and white lines, respectively). The red dashed line is the critical curve of same expression, with the parameters obtained by Vogel *et al.* in Ref. [15]. Fitting parameters are given in Table S1. Both panels: Since the deep globular phase is dominated by low energy, entropically suppressed conformations, a statistical analysis in the high N , low ε region requires sophisticated simulation methods that weren't undertaken in this study. Samples from the sectors $(N, \varepsilon) \in [700; 3000] \times [0.55; 1.0]$ and $[300; 700] \times [0.7; 1.0]$ didn't meet statistical relevance and were omitted from the results.

To test our hypotheses, we simulated single ISAWs cubic lattice [17] and sampled conformations thanks to the Metropolis algorithm with reptation moves [18]. We performed simulations for 22 values of N ranging from 100 to 3000, and 100 values of ε in the range 0.0–0.99. For each (N, ε) , 32000 statistically independent conformations are recorded. Figure 1-a show typical conformations for the case $N = 431$ and three different ε . The three spatial components are DCT transformed and their square am-

plitudes are summed to obtain their DCT spectra, as shown in Figure 1-c. For any given (N, ε) , we average over the set of configurations to obtain the corresponding PSD (thick lines in Figure 1-c). Figure 1-d shows the resulting PSDs for different values of ε and for the given N .

The results confirm both our predictions. First, for small values of the interaction parameter ε , the PSDs display the expected scaling of Eq. (4), corresponding to a slope of $-(1+2\nu) \approx -2.2$ in the log-log representation. Moreover, the spatial trajectories of Figure 1-b display a typical SAW behavior, with mean square displacement $\langle r_x^2(n) \rangle = n^{2\nu}$.

Second, on the large ε region, the spatial trajectories clearly show the effect of confinement into the globule volume (Figure 1-b). Correspondingly, the low p modes are strongly attenuated, leading to a rather flat spectrum, in agreement with our prediction (5).

Once confirmed our general hypotheses on the spectral profiles for the coil and globule configurations, we address the important issue of how to characterize the coil-globule transition. Inspired by previous observations on the low p limit, we defined as a new *order parameter* the log-log slope at $p = 1$ (that is proportional to the logarithm of the ratio of the first two squared modes),

$$\alpha_N(\varepsilon) = \frac{\ln \langle X_2^2 \rangle - \ln \langle X_1^2 \rangle}{\ln 2 - \ln 1} = \frac{1}{\ln 2} \ln \frac{\langle X_2^2 \rangle}{\langle X_1^2 \rangle}. \quad (6)$$

For the case of Figure 1-d, $\alpha_N(\varepsilon)$ clearly performs a *continuous* transition from the typical SAW exponent $\alpha = -(1+2\nu)$ to 0, as evidenced in the inset. The width of the transition region depends, however, on N . Figure 2 (top panel) shows the $\alpha_N(\varepsilon)$ curves with all the ε values and for all simulated polymer sizes N . These curves have been fitted by a general, 4-parameter sigmoid function, and the fit confirms that all $\alpha_N(\varepsilon)$ interpolates between the two limiting values $\alpha = -(1+2\nu)$ (coils) and $\alpha = 0$ (globules). The fact that these limiting values are independent of N and ε makes of $\alpha_N(\varepsilon)$ an excellent order parameter for assessing the polymer state, as it can be applied to each *single* (ε, N) polymer configuration set and *without the need of any information about these two parameters*. The bottom panel of Figure 2 shows a color plot of α as a function of N and ε where colors span from orange, for $\alpha = -2.2$, to blue for $\alpha = 0$. The two regions corresponding to coil and globule states are easily identified, with the typical finite-size crossover region (in black), clearly allowing the definition of a critical transition line $\varepsilon(N)$.

We aimed to compare the scaling of this critical line $\varepsilon_\theta(N)$ to previous results on finite-size crossover. Typically, the latter is defined either by the vanishing of the second virial coefficient [14], by the condition of divergent specific heat [15], or by a specific scaling of $\langle R^2 \rangle$ [19]. In Ref. [15], Vogel *et al.* show by on-lattice simulations that the boundary between the two phases is best fitted by a

function of the form $f(N) = a_1 N^{-1/2} + a_2 (2N)^{-1} + \varepsilon_\theta$. In the framework of our approach based on the first modes, we propose two different definitions for the phase transition critical line. The first one is simply given by the inflection points of the sigmoid functions. The corresponding $\varepsilon^I(N)$ points are very well-fitted by $f(N)$ (Figure 2, circles, and Table S1). Inspired by the diverging fluctuations typically observed in second order phase transition, we introduce a second, more physical definition of a transition line given by the $\varepsilon^F(N)$ that maximizes the fluctuations of our order parameter $\alpha_N(\varepsilon)$. Indeed, the relative standard deviation $\Delta \alpha_N(\varepsilon)/\alpha_N(\varepsilon)$ displays a sharp peak at the transition (meaning that the conformation of the polymer itself largely fluctuates at the crossover) as shown by the inset of Figure 2. The corresponding $\varepsilon^F(N)$ maxima are again best fitted by $f(N)$ (Figure 2, triangles and Table S1).

Interestingly, both $\varepsilon^I(N)$ and $\varepsilon^F(N)$ curves correspond to values of α close to ~ -1.1 (black region) and are therefore far from $\alpha = -2$, associated with RW-like configurations usually identified with the θ -point conditions. Our results show, instead, that these RW-like conditions do not appear at the crossover, but are rather located in the coil region of the phase portrait. More precisely, the condition $\alpha_N(\varepsilon_\theta) = -2$ is met at about $\varepsilon = 0.27$, close to the point where all the sigmoids roughly cross each other (the blue diamonds in Figure 2) and indeed corresponds to a PSD decreasing as p^{-2} over a wide region (Figure S1). The corresponding $\varepsilon_\theta^{RW}(N)$ curve is also fitted by a $f(N)$ function (although with negligible a_1 , see Table S1; blue line in Figure 2). As expected, however, the asymptotic ε_θ is virtually the same for the three critical curves $\varepsilon_\theta^I(N)$, $\varepsilon_\theta^F(N)$ and $\varepsilon_\theta^{RW}(N)$ and converge, within the error bar, to the well established thermodynamic limit value of $\varepsilon_\theta = 0.2690$, in agreement with previous numerical estimates [14, 15].

These results point out the richness of the coil-globule transition in finite-size: we were able to distinguish conditions where the chain configurations are RW-like from the thermodynamic (fluctuation based) transition in finite-size. The existence of two distinct critical lines were already pointed out by des Cloizeaux and Jannink [20], but not clearly illustrated by numerical experiments, to the best of our knowledge. As a corollary message of our work, we confirm therefore that the conformation of a polymer of finite-size at the coil-globule transition is not assimilable to a pure random walk. The detailed study of the chain configuration in between these limits, and more generally of their distribution, remains an open question that we will address in an incoming work.

As a last, but most important point, we now come back to the fact that, thanks to the universal values of the order parameter α , previous findings allows to assess the state of a single polymer from a sampling of its conformations. Moreover, since this order parameter can be calculated on the basis of the first two spectral modes

FIG. 3. (a) PSD obtained by successive decimations for a $N = 431$ chain starting from conformation sets at $\varepsilon = 0.0$ (orange, coils), $\varepsilon = 0.32$ (black) and $\varepsilon = 0.46$ (blue, globules). Decimation goes from $M = N$ to $M = 3$ (longer to shorter curves). (b) The asymptotic and critical values of $\alpha_{N,M}$ obtained from the sigmoid fit of the decimated spectra, as a function of M (dots and lines): $\alpha_C(M) = \alpha_{N,M}(0)$ for coils, orange; $\alpha_I(M) = \alpha_{N,M}(\varepsilon^I)$ for inflexion points, in black; $\alpha_G(M) = \alpha_{N,M}(\varepsilon \rightarrow \infty)$ for globules, in blue. These values are averaged over N . Shaded regions correspond to $2\sigma_\alpha$ confidence intervals for different sample size from 256 (larger intervals) to 2048. They are calculated by propagating the statistical variance of the first two modes through Eq. (6). (c) Renormalized sigmoids $\tilde{\alpha}_{N,3}(\varepsilon)$ (Equation 7) obtained for N from 101 to 3000 with a maximal decimation level $M = 3$. (d) Phase portrait as reconstructed from a $M = 3$ decimated signal. Red dots are obtained from inflexion points as in Figure 2. The blue line is a fit with the same $\varepsilon_\theta(N)$ expression, yielding $\varepsilon_\theta = 0.268$ (see Table S3 for further details). In all graphs we used a dataset of 16384 configurations for each (N, ε, M) condition.

only, it opens important perspectives on the possibility of *experimentally* assessing the state of a polymer from a very reduced and rather accessible information. In order to get access to the first M modes, indeed, it is in principle sufficient to record configurations of M *distinguishable, equally spaced monomers* covering the whole chain, either by following their dynamics or by averaging over a collection of identical polymers to ensure enough statistics. The design of multicolor fluorescent imaging of DNA sites with high spatial and temporal resolution now makes it possible to record the trajectories and relative positions of multiple loci simultaneously [21–23]. If the DCT-transform of a set of M -point configurations contain the equivalent of the first modes for the whole chain as we predict, this yields a new experimental approach for the determination of the state of single polymers and biopolymers.

We tested this idea on *decimated* polymer chains, i.e. reduced signals where the positions $\vec{r}_n$ of only M equally spaced monomers are retained. We decimated down to $M = 3$, an extreme condition of particular interest from the experimental point of view. Figure 3-a shows examples of PSD obtained for different choices of M . Extreme decimation ($M < 10$) alters the first modes. As a consequence, the asymptotic values of the sigmoid $\alpha_{N,M}(\varepsilon)$ corresponding to “pure” coil (α_C) and globule (α_G) states vary (while, interestingly, the inflexion point α_I seems to stay roughly constant), so that it is necessary to provide reference values for these limit slopes as a function of M . We give these values in Figure 3-b and in the Table S2. Despite this effect, the asymptotic values for coils and globules remain well apart down to $M = 3$ (Figure 3-b), even for relatively low statistical sampling

as shown by the shaded areas, giving a measure of the expected confidence interval on α for different sample sizes.

Once these M -dependent limit values obtained, the order parameter $\alpha_{N,M}(\varepsilon)$ can be normalized as

$$\tilde{\alpha}_N(\varepsilon) = \frac{2\alpha_{N,M}(\varepsilon) - (\alpha_G(M) + \alpha_C(M))}{(\alpha_G(M) - \alpha_C(M))} \quad (7)$$

so to span from -1 to 1. In this way, equivalent sigmoids $\tilde{\alpha}_N(\varepsilon)$ are independent of M (Figure S2). For sufficiently large samples, even an extreme decimation with only 3 equally spaced points allows a very accurate reconstruction of the (N, ε) phase diagram, as shown in Figure 3-d. The critical line $\varepsilon^I(N)$ matches that of the complete chain (white line in Figure 3-d and Table S3).

We finally checked the robustness of the proposed approach against variations in the size of the statistical sample. The obtained phase portraits are shown in Figure S3. We have been able to reconstruct the entire phase diagram from the configurations of only three points, by exploring the (N, ε) parameter space. Obviously, trying to determine the state of a polymer under given conditions, without being able to vary any parameter, is further complicated by the fact that finite-size effects generate a relatively large crossover region in which the state of the polymer smoothly change from coil to globule. However, the great robustness of the method introduced in this work and the universal character of the $\tilde{\alpha}_N$ parameter allow us to conclude that the proposed approach potentially outperforms any other method aiming at determining in which phase a polymer in given conditions falls. An important additional advantage comes from the fact that our approach does not require knowledge of the length of the polymer in terms of Kuhn lengths, nor of

the magnitude of the attraction between monomers, a parameter otherwise difficult to characterize.

The three-point measurement is very easy to perform experimentally, especially since, in this case, it is sufficient to discern the central point from the two extremities, which can be done for example by two-color fluorescence. Finally, preliminary results indicate that the method is also extremely robust to variations in the position of the three selected monomers. To go even further, we will implement an upgraded method that allows to explicitly take into account their position, in order to obtain results as good as for perfectly equidistant monomers. This improvement will be presented in a future work.

The authors would like to acknowledge networking support by the EUTOPIA COST Action CA17139. This work has been partially supported by the ANR project ANR-19-CE45-0016. Vincent Dahirel and Jean-Marc Victor are warmly acknowledged for their constant support and insightful advice.

* timothy.foldes@sorbonne-universite.fr

† maria.barbi@sorbonne-universite.fr

- [1] A. Lesage, V. Dahirel, J.-M. Victor, and M. Barbi, *Epigenetics & Chromatin* **12**, 10.1186/s13072-019-0269-6 (2019).
- [2] P.-G. de Gennes, *Scaling concepts in polymer physics* (Cornell University Press, 1979).
- [3] M. Doi and S. Edwards, *The Theory of Polymer Dynamics*, International series of monographs on physics (Clarendon Press, 1988).
- [4] M. Socol, R. Wang, D. Jost, P. Carrivain, C. Vaillant, E. L. Cam, V. Dahirel, C. Normand, K. Bystricky, J.-M. Victor, O. Gadal, and A. Bancaud, *Nucleic Acids Research* **47**, 6195 (2019).
- [5] P. E. Rouse, *The Journal of Chemical Physics* **21**, 1272 (1953).
- [6] The random forces $\xi_n(t)$ satisfy the conditions $\langle \xi_n(t) \rangle = 0$ and $\langle \xi_n(t') \cdot \xi_m(t) \rangle = 2D \delta_{nm} \delta(t - t')$ with $D = k_B T / \gamma$ the diffusion coefficient of a single monomer.
- [7] J. Makhoul, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**, 27 (1980).
- [8] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules*, AIP series in polymers and complex materials (AIP Press, 1994).
- [9] D. Krapf, E. Marinari, R. Metzler, G. Oshanin, X. Xu, and A. Squarcini, *New Journal of Physics* **20**, 023029 (2018).
- [10] D. Panja and G. T. Barkema, *The Journal of Chemical Physics* **131**, 154903 (2009).
- [11] G. Slade, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475**, 20180549 (2019).
- [12] L. Peliti, Random walks with memory, in *Fractals in physics*, edited by L. Pietronero and E. Tosatti (Elsevier, Amsterdam: North Holland, 1986) p. 73.
- [13] The PSD of a self similar process has asymptotic power law behavior provided that it has stationary (*i.e.* n -independent) increments $(\vec{r}_{n+1} - \vec{r}_n)$, a property that holds for our system as we verified on simulations (data not shown).
- [14] P. Grassberger and R. Hegger, *The Journal of Chemical Physics* **102**, 6881 (1995).
- [15] T. Vogel, M. Bachmann, and W. Janke, *Physical Review E* **76**, 10.1103/physreve.76.061803 (2007).
- [16] B. R. Caré, P. Carrivain, T. Forné, J.-M. Victor, and A. Lesne, *Communications in Theoretical Physics* **62**, 607 (2014).
- [17] M. C. Tesi, E. J. J. van Rensburg, E. Orlandini, and S. G. Whittington, *Journal of Statistical Physics* **82**, 155 (1996).
- [18] F. T. Wall and F. Mandel, *The Journal of Chemical Physics* **63**, 4592 (1975).
- [19] F. Rampf, K. Binder, and W. Paul, *Journal of Polymer Science Part B: Polymer Physics* **44**, 2542 (2006).
- [20] J. des Cloizeaux and G. Jannink, *Polymers in solution. Their modelling and structure* (Oxford university press, New York, 1991).
- [21] H. Chen, M. Levo, L. Barinov, M. Fujioka, J. B. Jaynes, and T. Gregor, *Nature Genetics* **50**, 1296 (2018).
- [22] H. Sato, S. Das, R. H. Singer, and M. Vera, *Annual Review of Biochemistry* **89**, 159 (2020).
- [23] J. K. Eykelenboom and T. U. Tanaka, *Cell Cycle* **19**, 1422 (2020).

Supplementary informations

Coefficients a_1 , a_2 and ε_θ for the three curves

Data	a_1	a_2	ε_θ
Inflection points $\varepsilon^I(N)$	2.1 ± 0.04	3.3 ± 0.9	0.2694 ± 0.0005
Fluctuation maxima $\varepsilon^F(N)$	2.5 ± 0.09	2.7 ± 1.8	0.270 ± 0.002
Random Walk $\varepsilon^{RW}(N)$	0.0 ± 0.1	4.6 ± 1.8	0.269 ± 0.003
Vogel <i>et al.</i> dC_v/dt [15]	2.5	8.0	0.2690 ± 0.0005

TABLE S1. In Figure 2 we have fitted inflection points (colored circles) and $\alpha_N(\varepsilon)$ fluctuation maxima (colored triangles) by the expression $\varepsilon_\theta(N) = a_1 N^{-1/2} + a_2 (2N)^{-1} + \varepsilon_\theta$ from Ref. [15] (green and white lines, respectively). Here we list the corresponding fitting parameters a_1 , a_2 and ε_θ , together to those obtained by Vogel *et al.* in Ref. [15] by fitting a different order parameter.

Reference values α_C , α_I and α_G obtained for different M

M	α_C^a	α_I^a	α_G^a
3	-1.81	-1.09	-0.36
4	-2.03	-1.10	-0.19
5	-2.11	-1.10	-0.11
6	-2.15	-1.10	-0.07
7	-2.16	-1.10	-0.05
8	-2.19	-1.11	-0.03
10	-2.19	-1.11	-0.03
20	-2.21	-1.12	-0.02

^a Significant digits.

TABLE S2. Numerical values of the loglog-slopes $\alpha_{N,M}(\varepsilon)$ obtained from the spectra in Figure 3-a and displayed in Figure 3-b (points), as a function of M . The parameter α_C , for coils, is obtained at the limit $\varepsilon = 0$; α_I is measured at the fitting sigmoid inflexion point; α_G , for globules, is calculated as $\varepsilon \rightarrow \infty$ limit of the fitting sigmoid.

Critical $\varepsilon^I(N)$ inflection point fit parameters upon decimation

M	a_1	a_2	ε_θ
3	1.7 ± 0.1	2.7 ± 1.8	0.273 ± 0.002
5	1.8 ± 0.05	5.0 ± 0.9	2.270 ± 0.001
8	1.9 ± 0.04	4.4 ± 0.7	2.270 ± 0.001
10	2.0 ± 0.05	4.2 ± 0.8	0.270 ± 0.001
$M = N$	2.1 ± 0.04	3.3 ± 0.9	0.2694 ± 0.0005

TABLE S3. As shown in Figure 3, decimated configurations retain the main information about the polymer state and allows us to reconstruct the phase diagram in the (ε, N) phase space. Moreover, inflection points (colored circles in Figure 3) can be used to define a critical line also in decimated conditions, and fitted by the function $\varepsilon_\theta(N) = a_1 N^{-1/2} + a_2 (2N)^{-1} + \varepsilon_\theta$. Here we give corresponding values of the three fitting parameters for $M = 3, 5, 8, 10$ and, for comparison, for the non decimated case $M = N$. Note that the asymptotic value ε_θ is essentially independent on decimation, ensuring a correct definition of the phase transition in the thermodynamic limit. The a_1 and a_2 parameters, that gives the N dependence of the transition in finite-size, are instead slightly modified by the decimation procedure, but converge toward the original values as soon as $M \gtrsim 10$.

Typical PSD at the theta conditions $\varepsilon = \varepsilon_{\theta}^{RW}(N)$

FIG. S1. The plot shows a typical PSD obtained for a value of ε matching the $\alpha = -2$ conditions, i.e for $\varepsilon = \varepsilon_{\theta}^{RW}(N)$ (green points). For comparison, a pure coil spectrum is shown (orange points) with its typical -2.2 slope (orange line) and a pure on-lattice RW (purple points) of slope -2 (purple line) are shown. The first modes slope α for $\varepsilon = \varepsilon_{\theta}^{RW}(N)$ is as expected equal to -2 (green line); Moreover, the plot shows how the same scaling law holds over a wide range of modes. This ensures that the polymer configurations can be described as random walks up to short scales (where a SAW behavior is recovered).

Effect of different levels of decimation on the phase transition

FIG. S2. The different panels show renormalized loglog-slopes $\tilde{\alpha}_N(\epsilon)$ (Eq. (7)) for different value of N . For each N , we show results obtained by different levels of decimation M (blue points), starting from 3 (red points) to N (yellow points). Comparison of these different cases shows that, for given N , there is essentially no difference in the behavior of the order parameter $\alpha_N(\epsilon)$, once rescaled with the maximum and minimum values, $\alpha_G(M)$ and $\alpha_C(M)$, respectively. Data sample includes here 16384 configurations per point. We recall that, as in Figure 2, samples from the sectors $(N, \epsilon) \in [700; 3000] \times [0.55; 1.0]$ and $[300; 700] \times [0.7; 1.0]$ didn't meet statistical relevance. Here we have kept the raw results from these sectors, which also gives a measure of the variability of the corresponding values.

Phase portraits after decimation, with different sample size.

FIG. S3. Example of phase portraits obtained by extreme decimation $M = 3$ starting from a same set of conformations, for different *sample* sizes, from 64 to 2048. We recall that, as in Figure 2, samples from the sectors $(N, \epsilon) \in [700; 3000] \times [0.55; 1.0]$ and $[300; 700] \times [0.7; 1.0]$ didn't meet statistical relevance. Here we have kept the raw results from these sectors, which also gives a measure of the variability of the corresponding values. Overall, we observe that if a sample of only 64 configurations is not sufficient to assign with certainty a system to the coil or globule states, it becomes rather good starting from 512.