



HAL
open science

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

Timothy Foldes, Antony Lesage, Maria Barbi

► **To cite this version:**

Timothy Foldes, Antony Lesage, Maria Barbi. Assessing the polymer coil-globule state from the very first spectral modes. *Physical Review Letters*, 2021, 10.1103/PhysRevLett.127.277801 . hal-03466632

HAL Id: hal-03466632

<https://hal.science/hal-03466632>

Submitted on 6 Dec 2021

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.

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: December 6, 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 and reintroducing well-established spectral methods that have been surprisingly neglected in this field so far. 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 at least three discernible loci on the polymer. Furthermore, the framework we put forward applies to virtually any polymer model, fractal globule, loop extrusion...

The coil-to-globule phase transition, which describes the abrupt compaction of a polymer due to environmental changes, characterizes the physics of polymers in solution, with direct consequences for macroscopic solution properties such as viscoelasticity or transport features and, moreover, successfully models the heterogeneity of chromatin density, which most likely plays a crucial functional role in the regulation of gene expression.

It is usually studied by means of the scaling properties of the radius of gyration as a function of the 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, as it depends on the Kuhn length of the polymer, hence its bending persistence, which can be difficult to determine in the case of real polymers, and a fortiori for complex biological polymers. Moreover, important deviations from the theoretical scaling are induced by the finite size of the polymers [1].

From a different perspective, the dynamics 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 [2, 3]. Now, interestingly, the Rouse model uses a Fourier-like decomposition of the polymer conformation. Noting that this decomposition applies to any given conformation, we used it to determine the average spectral content of the polymer to inquire its equilibrium properties and study the coil-globule transition, rather than its dynamics. This paper aims to investigate this issue and to see how a spectral representation can facilitate the determination of the polymer state (*i.e.* its macrostate). 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 spa-

tial path of the polymer and only relies on *low* spectral modes, *i.e.* long-distance features. The introduction of a new N -independent order parameter provides us with a parsimonious and robust measurement of the state of the polymer on the coil to globule scale. Finally, based on this analysis, we propose an innovative experimental approach to determine the equilibrium state of a polymer from fluorescence imaging, provided that a minimum of three loci, equally spaced along the chain, can be discernibly labeled.

The Rouse model describes an ideal chain (no excluded volume nor attractive interaction) for which the radius of gyration scales as the variance of a random walk, $\langle R^2 \rangle \sim N$. The chain dynamics is described by applying the overdamped Langevin equation to a bead-spring polymer of $N + 1$ monomers, immersed in a fluid of viscosity γ at temperature T [4]. The application of the Langevin equation yields a system of $N + 1$ three-dimensional coupled stochastic equations [2]. To solve the model, this system is decoupled by introducing the new set of variables $\{\vec{X}_p(t), p = 0 \dots N\}$

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

called *Rouse modes* of the polymer. The $p = 0$ mode corresponds to the position of the polymer center of mass. In the following, we will always focus on modes $p = 1 \dots N - 1$ 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, each component of (1) is a Discrete Cosine Transform (DCT), closely related to the Discrete Fourier Transform (DFT), in that it corresponds to the DFT of a symmetrized signal of double length [5]. Hence, the average over a set of equilibrium configurations of the square mode amplitudes $\langle X_p^2 \rangle$ is none other

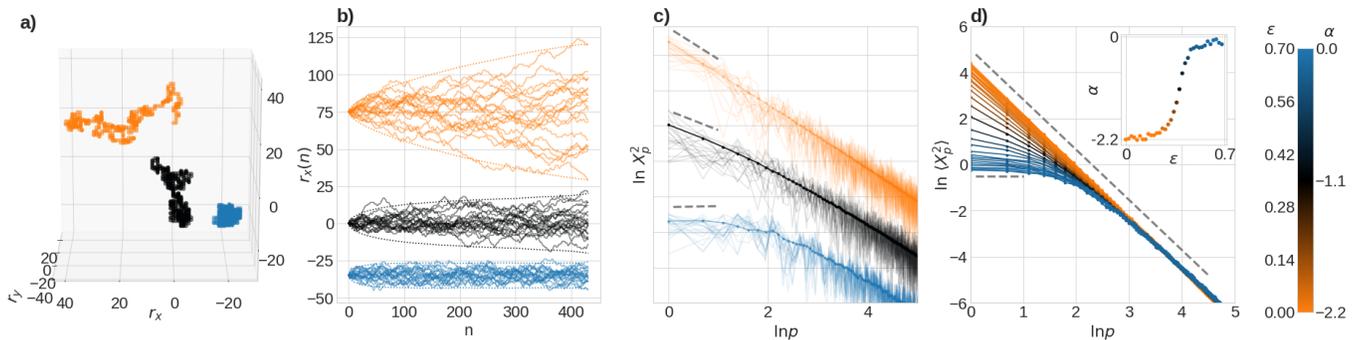


FIG. 1. **(a)** On lattice MC simulation snapshots for $N = 431$ and $\varepsilon = 0$ (orange), $\varepsilon = 0.36$ (black) and $\varepsilon = 0.8$ (blue). **(b)** Projections $r_x(n)$ along the spatial x -axis of a set of 20 configurations, n being the monomer index. Colors refer to the same parameters as in (a). Dotted lines correspond to the $\pm 2.5 \sigma_x(n)$ envelopes. **(c)** Logarithm of the square Discret Cosine Transform of the 20 trajectories of (b), $\ln |X_p|^2$, as functions of the logarithm of the mode number $\ln p$ (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. Dashed lines correspond to slopes $-2.2 \equiv -(1 + 2\nu)$, -1.3 and 0 (from top to bottom). **(d)** Averaged Power Spectral Density (PSD) obtained by applying the same procedure as in (c) to equivalent simulations with ε going from 0 to 0.6 (see colorbar). 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 ε .

than the canonical estimator of the *Power Spectral Density* (PSD), characterizing a polymer macrostate in terms of the relative weight of the different spectral modes. For the Cartesian coordinate x (equivalently y or z), it is well known that [2, 4]

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

yielding an explicit expression for the mean square amplitude of the Rouse modes of an ideal polymer.

The connection between Rouse modes and PSD seems, surprisingly, to have never been noticed before. However, the asymptotic dependence (2) has an immediate interpretation in terms of the equivalence to Brownian motion trajectories [6], for which the PSD has a scaling exponent -2 . Interestingly, exactly the same expression (2) is obtained in a different setting [7].

Now, 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 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 [6]. Including steric interactions adds additional non-linear coupling terms to the Rouse equation system, and, consequently, there is no explicit solution for the PSD [8]. Yet simple arguments provide 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 [6, 9, 10]. Hence, we expect a power law PSD, intuitively related to the presence of long-range correlations induced by steric interactions. The Hurst exponent H provides an estimation of the long-range memory of a signal. In particular, for a centered fractal signal of Hurst index H one has $\langle r_n^2 \rangle \propto n^{2H}$ and a PSD scaling as $p^{-(1+2H)}$. By guessing a long-range correlated process for the SAW, the comparison with the Flory scaling $\langle R^2 \rangle \sim N^{2\nu}$ would lead to the ansatz $H = \nu$ which appoints the PSD

$$\langle \vec{X}_p^2 \rangle \propto p^{-(1+2\nu)} \quad (3)$$

where $-(1 + 2\nu) \approx -2.2$. Panja and Barkema made an equivalent ansatz for long wavelengths Rouse modes of the self-avoiding polymer, and corroborate it by on-lattice simulations [8].

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, $\varepsilon = J/k_B T$. This case is often modeled by the Interacting Self-Avoiding Walk (ISAW), an on-lattice polymer with each nearest neighbor monomer pair contributing $-\varepsilon$ to the energy (the SAW model is retrieved for $\varepsilon = 0$). 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 $\langle R^2 \rangle \propto N^{2/3}$ [6]. However, finite-size polymers undergo a smooth coil-globule transition at a N -dependent critical energy [11–14].

The spatial configuration of the chain can be described by noting that a collapsed polymer is equivalent to an

ideal polymer compressed within a spherical volume of radius R [6]. Over small length scales, the polymer behaves like a free chain, until it reaches the volume boundary, where it is reflected and starts an independent random walk. The chain can thus be described as a series of independent subchains. The positions of two monomers in different sub-chains are uncorrelated [6], resulting in the sequence of positions of monomers sufficiently spaced along the chain being a white noise signal. From the spectral point of view, this implies a constant spectrum for small values of p . We thus predict that the PSD of globules verifies the property

$$\langle \vec{X}_p^2 \rangle \propto p^0 \quad \text{for small } p. \quad (4)$$

To test our hypotheses, we simulated single ISAW on the cubic lattice [15] and sampled conformations thanks to the Metropolis algorithm with reptation moves [16]. We performed simulations for N ranging from 100 to 3000, and ε in the range 0.0–0.99. For each (N, ε) , 32000 statistically independent conformations are recorded. Figure 1-a shows typical conformations for the case $N = 431$ and three different ε . The three spatial components are DCT transformed and their square amplitudes 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. (3), corresponding to a slope of $-(1 + 2\nu) \approx -2.2$ in the log-log representation. Second, in 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 (4).

We aimed to compare the scaling of $\varepsilon_\theta(N)$ to previous results, typically defined either by the vanishing of the second virial coefficient [12], by the condition of divergent specific heat [13], or by a specific scaling of $\langle R^2 \rangle$ [17]. In Ref. [13], Vogel *et al.* show by on-lattice simulations that the boundary between the two phases is best fitted by a function of the form $f(N) = a_1 N^{-1/2} + a_2 (2N)^{-1} + \varepsilon_\theta$. We here 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). A second, more physical definition of a transition line is 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) as shown by the inset of Figure 2. The corresponding

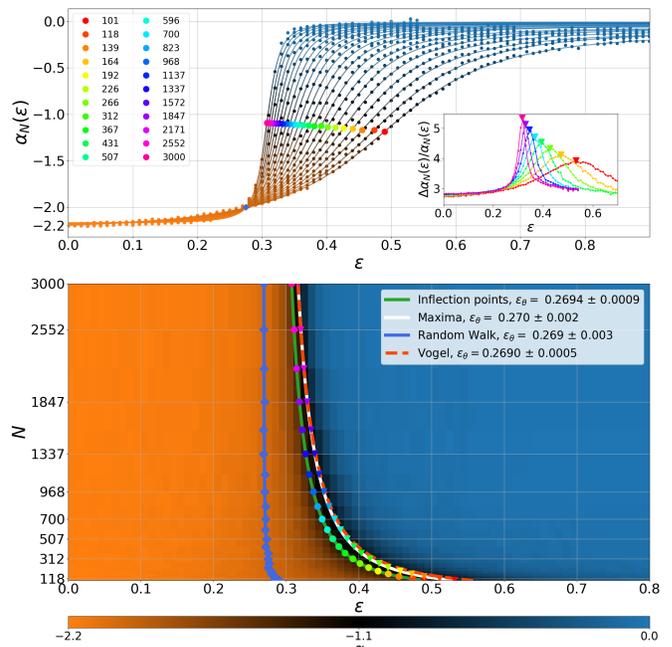


FIG. 2. **Top panel:** Log-log plot slope $\alpha_N(\varepsilon)$ estimated on the first 2 modes of Power Spectral density (dots) for all polymers sizes N from 101 to 3000 (right to left). For each N , $\alpha_N(\varepsilon)$ is fitted by a sigmoid $S(x) = D + A(x - B)/\sqrt{1 + C(x - B)^2}$ (lines). Colored circles are inflection points, defining $\varepsilon^I(N)$. Blue diamonds indicate the $\alpha_N(\varepsilon) = -2$ condition, defining $\varepsilon_\theta^{RW}(N)$. **Inset:** $\alpha_N(\varepsilon)$ relative standard deviation, $\Delta\alpha_N(\varepsilon)/\alpha_N(\varepsilon) = \sigma_{X_1^2}/\langle X_1^2 \rangle + \sigma_{X_2^2}/\langle X_2^2 \rangle$. 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. [13]. Fitting parameters are given in Table S1. *Both panels:* 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.

$\varepsilon^F(N)$ maxima are again best fitted by $f(N)$ (Figure 2, triangles and Table S1).

Interestingly, both $\varepsilon^I(N)$ and $\varepsilon^F(N)$ 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 diagram. More precisely, the condition $\alpha_N(\varepsilon_\theta) = -2$ is met at about $\varepsilon = 0.27$ (blue diamonds in Figure 2) and indeed corresponds to a PSD decreasing as p^{-2} over a wide p -region (Figure S1). The corresponding $\varepsilon_\theta^{RW}(N)$ curve is also fitted by a $f(N)$

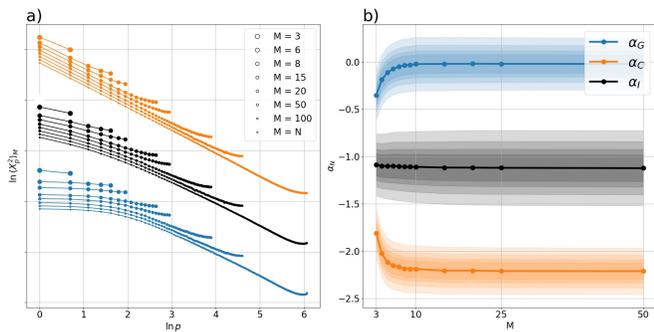


FIG. 3. **(a)** Power Spectral Densities obtained by successive decimation of a $N = 431$ chain for $\varepsilon = 0.0$ (orange, coils), $\varepsilon = 0.32$ (black) and $\varepsilon = 0.46$ (blue, globules). Decimation goes from $M = N$ to $M = 3$ (longer to shorter curves). **(b)** The asymptotic and 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, in orange; $\alpha_I(M) = \alpha_{N,M}(\varepsilon^I)$ for inflection 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, calculated by propagating the statistical variance through Eq. (5). We used a dataset of 16384 configurations for each (N, ε, M) condition.

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 to the well established thermodynamic limit value of $\varepsilon_\theta = 0.2690$, in agreement with previous numerical estimates [12, 13].

The existence of two distinct critical lines was already pointed out by des Cloizeaux and Jannink [18], 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 reducible to a pure random walk. The very accurate phase diagrams obtained by spectral analysis validates it as a new, efficient and parsimonious method in polymer studies, neglected until now despite its very classical basis.

With our general hypotheses confirmed, we address the important issue of characterizing the coil-globule transition. Inspired by our previous observations on the low p limit, we define as a new order parameter the log-log slope at $p = 1$

$$\alpha_N(\varepsilon) = \frac{\ln\langle X_2^2 \rangle - \ln\langle X_1^2 \rangle}{\ln 2 - \ln 1} = \frac{1}{\ln 2} \ln \frac{\langle X_2^2 \rangle}{\langle X_1^2 \rangle}. \quad (5)$$

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 (inset). The width of the transition region depends, however, on N . The $\alpha_N(\varepsilon)$ curves, for all ε and N , have been fitted by a 4-parameter sigmoid function (Figure 2, top panel), and all $\alpha_N(\varepsilon)$ indeed in-

terpolate between the two limiting values $\alpha = -(1 + 2\nu)$ and $\alpha = 0$. 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 typical finite-size crossover region (in black) is clearly visible, allowing the definition of a critical transition line $\varepsilon(N)$.

As a last, but most important point, the fact that our order parameter can be calculated on the basis of the first two spectral modes opens the possibility of experimentally assessing the state of a polymer from a very reduced and accessible information. The argument relies on a well-known property of the PSD: In order to get access to the first M modes, it is indeed in principle sufficient to record the positions of M distinguishable monomers, equally spaced along the polymer and covering the whole chain, meaning that the determination of α would in theory require the knowledge of only 3 positions along the full polymer.

We tested the efficiency of our approach 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 since, in this case, time tracking can also be achieved without much difficulty [19–21] so that time averages can be used instead of ensemble averages. 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 sigmoidal $\alpha_{N,M}(\varepsilon)$ corresponding to "pure" coils (α_C) and globules (α_G) vary, 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. Noteworthy, the asymptotic values for coils and globules remain well apart down to $M = 3$ even for relatively low statistical sampling meaning that the arrangement of monomers on large scales is detectable whatever the sampling.

Once these M -dependent limit values obtained, the order parameter $\alpha_{N,M}(\varepsilon)$ can be normalized as $\tilde{\alpha}_N(\varepsilon) = [2\alpha_{N,M}(\varepsilon) - (\alpha_G(M) + \alpha_C(M))] / (\alpha_G(M) - \alpha_C(M))$ 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 $M = 3$ allows a very accurate reconstruction of the (N, ε) phase diagram (Figure S3). The critical line $\varepsilon^I(N)$ matches that of the complete chain (white line in Figure S3 and Table S3). The robustness of the proposed approach against variations in the size of the statistical sample is explored in Figure S4 showing that the universal character of the $\tilde{\alpha}_N$ parameter potentially outperforms any other method aiming at determining in which phase

a polymer in given conditions falls.

These results have relevant implications in the context of the study of chromatin. The characterization of the multiscale folding state of chromosomes is a crucial, issue since it likely determines its the biological activity. This is made evident from the strong progression of experimental methods to characterize this folding and their impact on the understanding of genetic regulation mechanisms [22]. Very recently, microscopy-based techniques allow for the first time the visualization of this polymer spatial trajectory by sequential labeling and imaging of multiple loci in a collection of *fixed* nuclei [23–26]. This results in configuration sets, sampled with a resolution ranging from 3 to several hundreds of points, which are for the moment difficult to analyze, the most frequent approaches being the reconstruction of parameters that were already obtainable with previous techniques. There is therefore a clear need for efficient methods to process this new data to the fullest and without loss of information. Our method of spectral analysis precisely fills this gap with a physically proven approach. Finally, we underline the very broad scope of this spectral approach, which extends far beyond the pure coil and globule. One can easily predict a specific spectral signature for fractal globules, loops, stretched polymers or a variety of multifractal configurations, potentially also interesting for the study of biopolymers.

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** (2019), 10.1186/s13072-019-0269-6.
- [2] M. Doi and S. Edwards, *The Theory of Polymer Dynamics*, International series of monographs on physics (Clarendon Press, 1988).
- [3] M. Socol, R. Wang, D. Jost, P. Carrivain, C. Vaillant, E. L. Cam, V. Dahirel, C. Normand, K. Bystrycky, J.-M. Victor, O. Gadai, and A. Bancaud, *Nucleic Acids Research* **47**, 6195 (2019).
- [4] P. E. Rouse, *The Journal of Chemical Physics* **21**, 1272 (1953).
- [5] J. Makhoul, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**, 27 (1980).
- [6] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules*, AIP series in polymers and complex materials (AIP Press, 1994).
- [7] D. Krapf, E. Marinari, R. Metzler, G. Oshanin, X. Xu, and A. Squarcini, *New Journal of Physics* **20**, 023029 (2018).
- [8] D. Panja and G. T. Barkema, *The Journal of Chemical Physics* **131**, 154903 (2009).
- [9] G. Slade, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475**, 20180549 (2019).
- [10] L. Peliti, “Random walks with memory,” in *Fractals in physics*, edited by L. Pietronero and E. Tosatti (Elsevier, Amsterdam: North Holland, 1986) p. 73.
- [11] P.-G. de Gennes, *Scaling concepts in polymer physics* (Cornel University Press, 1979).
- [12] P. Grassberger and R. Hegger, *The Journal of Chemical Physics* **102**, 6881 (1995).
- [13] T. Vogel, M. Bachmann, and W. Janke, *Physical Review E* **76** (2007), 10.1103/physreve.76.061803.
- [14] B. R. Caré, P. Carrivain, T. Forné, J.-M. Victor, and A. Lesne, *Communications in Theoretical Physics* **62**, 607 (2014).
- [15] M. C. Tesi, E. J. J. van Rensburg, E. Orlandini, and S. G. Whittington, *Journal of Statistical Physics* **82**, 155 (1996).
- [16] F. T. Wall and F. Mandel, *The Journal of Chemical Physics* **63**, 4592 (1975).
- [17] F. Rampf, K. Binder, and W. Paul, *Journal of Polymer Science Part B: Polymer Physics* **44**, 2542 (2006).
- [18] J. des Cloizeaux and G. Jannink, *Polymers in solution. Their modelling and structure* (Oxford university press, New York, 1991).
- [19] H. Chen, M. Levo, L. Barinov, M. Fujioka, J. B. Jaynes, and T. Gregor, *Nature Genetics* **50**, 1296 (2018).
- [20] H. Sato, S. Das, R. H. Singer, and M. Vera, *Annual Review of Biochemistry* **89**, 159 (2020).
- [21] J. K. Eykelboom and T. U. Tanaka, *Cell Cycle* **19**, 1422 (2020).
- [22] I. Jerkovic´ and G. Cavalli, *Nature Reviews Molecular Cell Biology* **22**, 511–528 (2021).
- [23] A. M. Cardozo Gizzi, D. I. Cattoni, J.-B. Fiche, S. M. Espinola, J. Gurgo, O. Messina, C. Houbbron, Y. Ogiyama, G. L. Papadopoulos, G. Cavalli, M. Lagha, and M. Nollmann, *Molecular Cell* **74**, 212 (2019).
- [24] L. Mateo, S. Murphy, A. Hafner, I. Cinquini, C. Walker, and A. Boettiger, *Nature* **568** (2019).
- [25] H. Q. N. *et al.*, *Nature methods* **17**, 822 (2020).
- [26] J.-H. Su, P. Zheng, S. Kinrot, B. Bintu, and X. Zhuang, *Cell* **182**, 1641 (2020).

Supplementary information

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 [13]	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. [13] (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. [13] 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 $\varepsilon = \varepsilon_{\theta}^{RW}(N)$

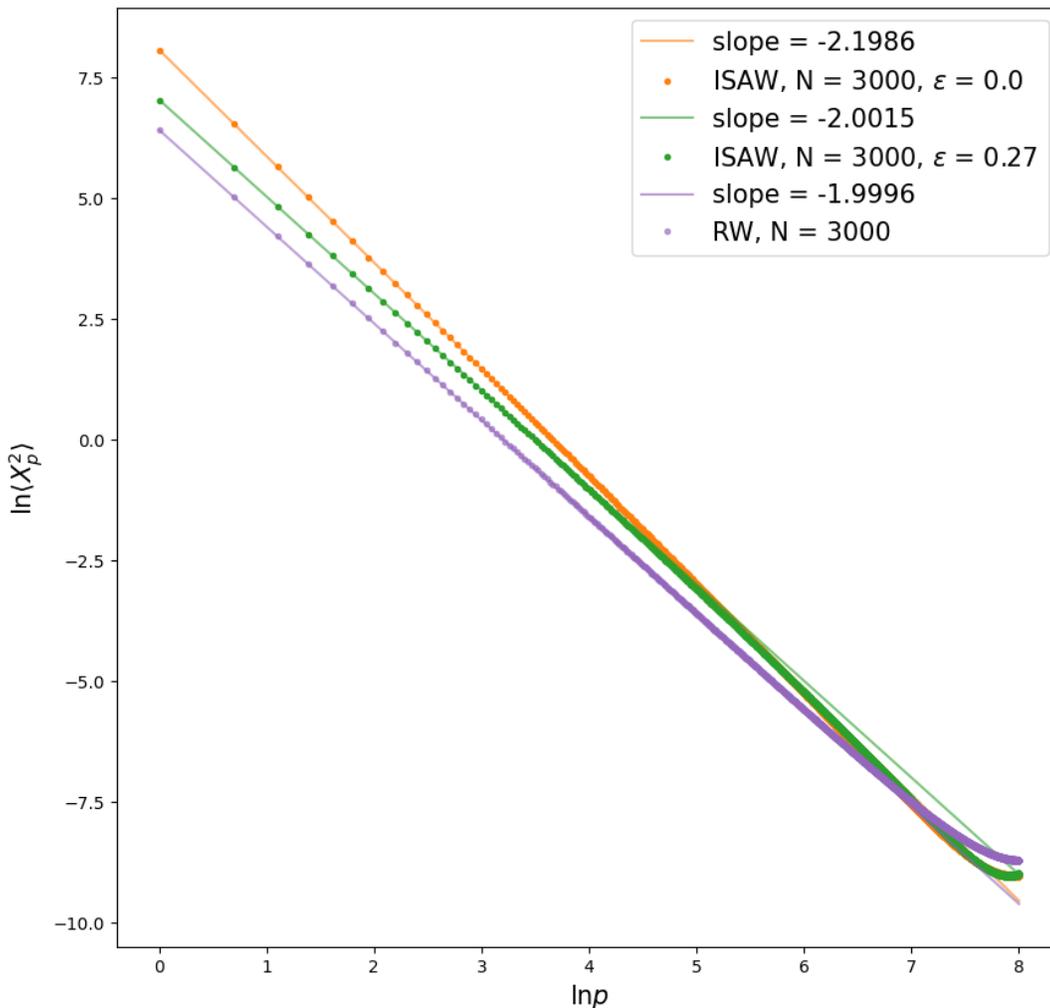


FIG. S1. The plot shows a typical PSD obtained for a value of ε matching the $\alpha = -2$ conditions, i.e for $\varepsilon = \varepsilon_{\theta}^{RW}(N)$ (green points). For comparison, a pure coil spectrum is shown (orange points) with its typical -2.2 slope (orange line) and a pure on-lattice RW (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

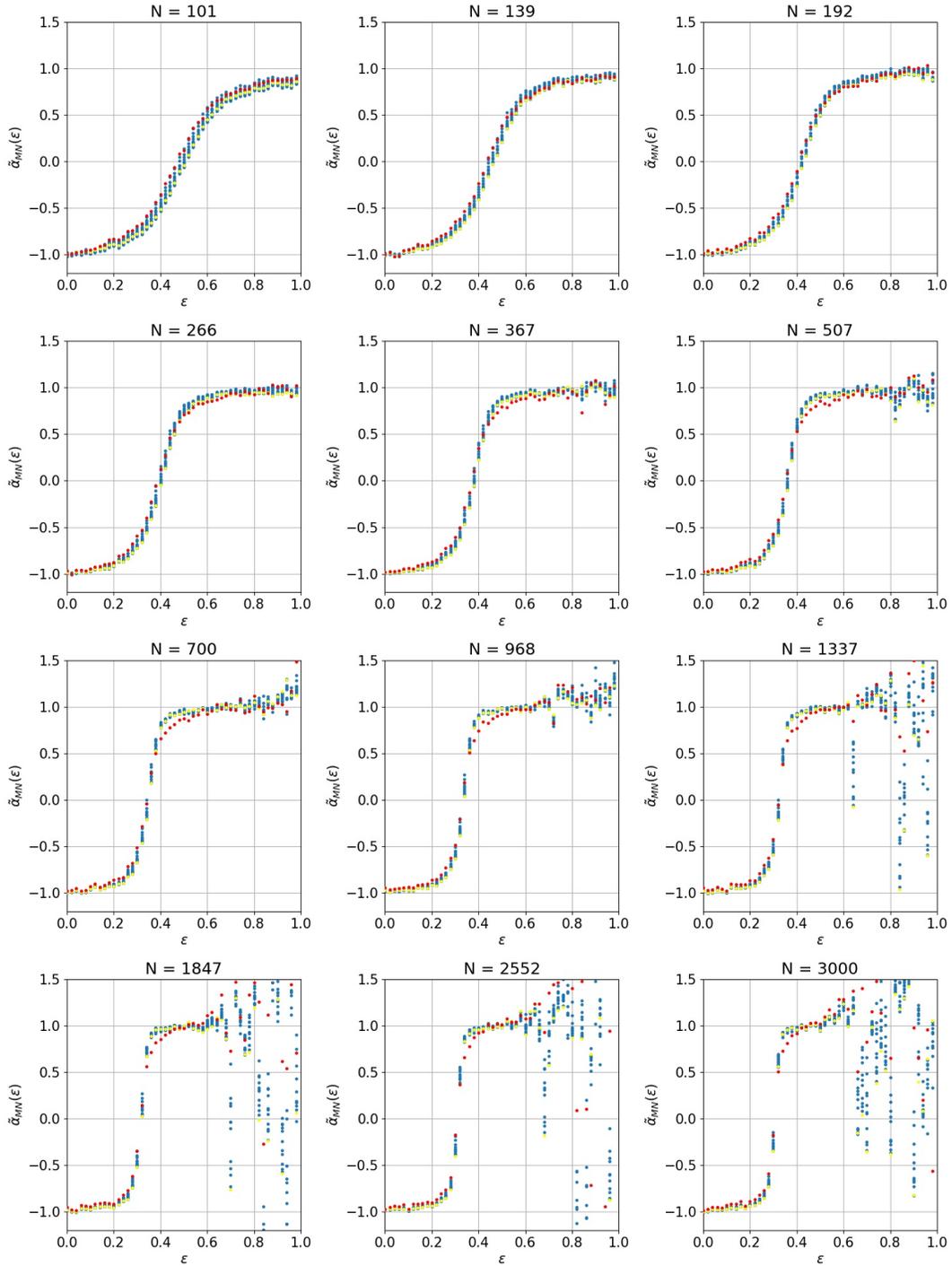


FIG. S2. The different panels show renormalized loglog-slopes $\tilde{\alpha}_N(\varepsilon)$ (Eq. (6)) for different value of N . For each N , we show results obtained by different levels of decimation M (blue points), starting from 3 (red points) to N (yellow points). Comparison of these different cases shows that, for given N , there is essentially no difference in the behavior of the order parameter $\alpha_N(\varepsilon)$, 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 $(N, \varepsilon) \in [700; 3000] \times [0.55; 1.0]$ 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.

Sigmoids and phase portrait from the $M = 3$ decimated configurations

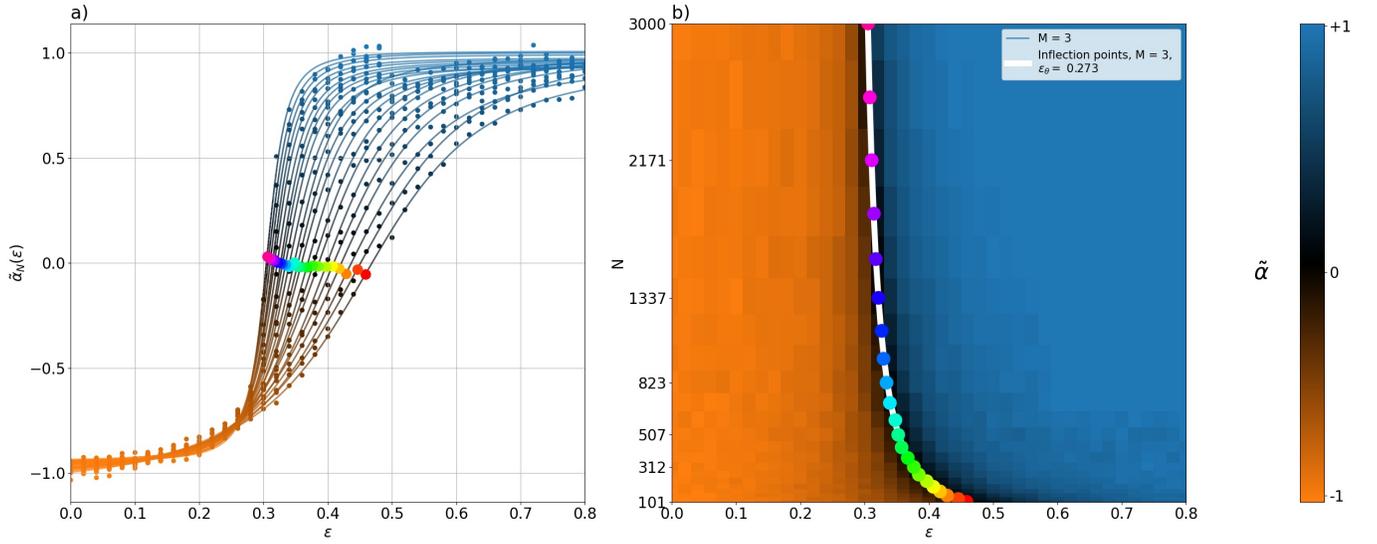


FIG. S3. In Figure 3 of the main article are shown the Power Spectral Densities and critical values of $\alpha_{N,M}$ obtained by successive decimation of a $N = 431$ chain. Here we show, for the same system, (a) the renormalized sigmoids $\tilde{\alpha}_{N,3}(\varepsilon)$ for N from 101 to 3000 with decimation level $M = 3$; and (b) the phase portrait as reconstructed from the $M = 3$ decimated signal. Colored dots are obtained from inflection points as in Figure 2. The white line is a fit yielding $\varepsilon_\theta = 0.268$ (see Table S3). In both graphs we used a dataset of 16384 configurations.

Phase portraits after decimation, with different sample size.

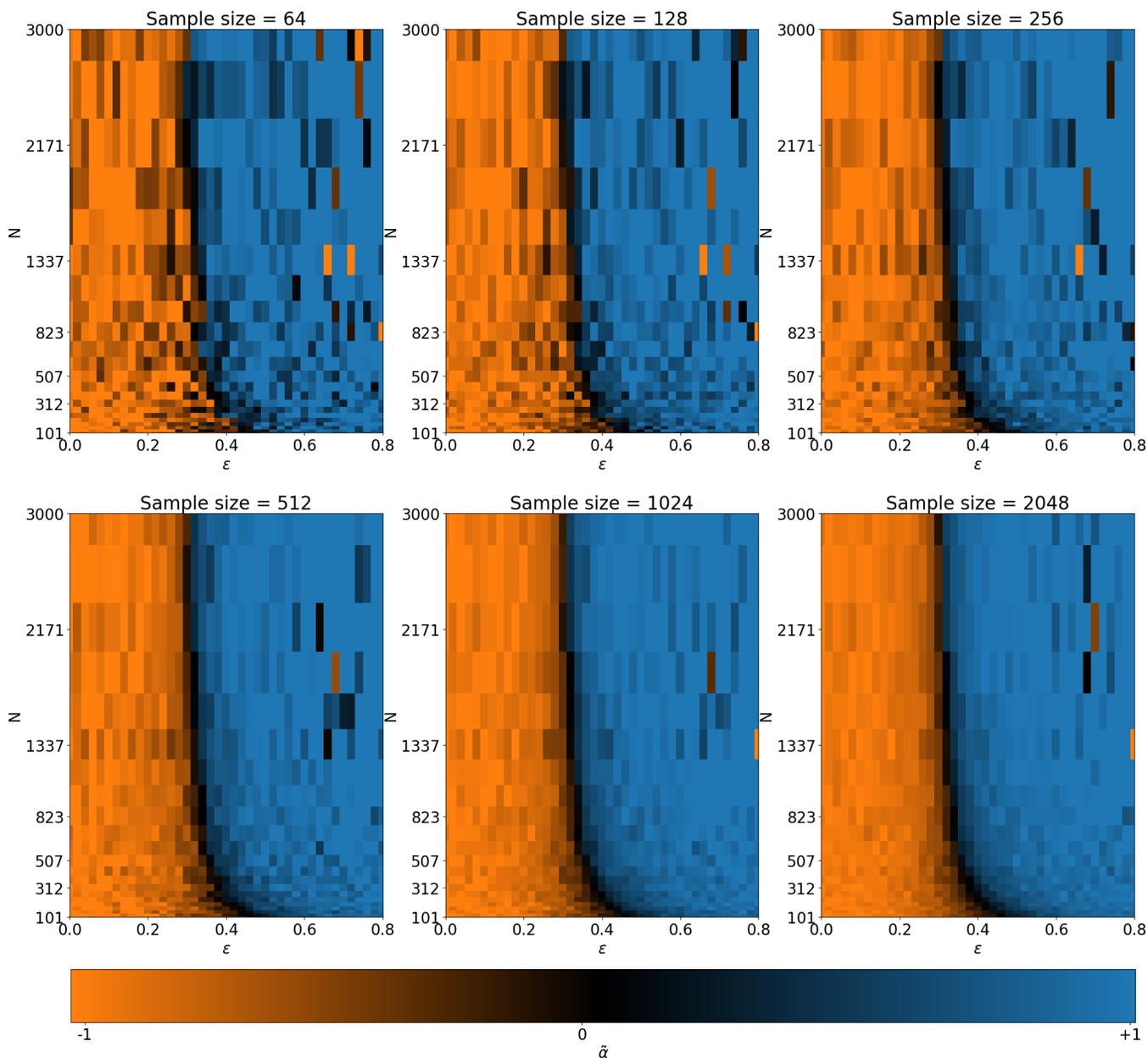


FIG. S4. Example of phase portraits obtained by extreme decimation $M = 3$ starting from a same set of conformations, for different *sample* sizes, from 64 to 2048. We recall that, as in Figure 2, samples from the sectors $(N, \varepsilon) \in [700; 3000] \times [0.55; 1.0]$ and $[300; 700] \times [0.7; 1.0]$ 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.

Simulation method details

We simulated single ISAW on the cubic lattice and sampled conformations thanks to the Metropolis algorithm with reptation moves. 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 shows typical conformations for the case $N = 431$ and three different ε . The three spatial components are DCT transformed, and their square amplitudes 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.

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.