

Finite-Size Conformational Transitions: A Unifying Concept Underlying Chromosome Dynamics*

Bertrand R. Caré,^{1,5} Pascal Carrivain,^{2,5} Thierry Forné,^{3,5} Jean-Marc Victor,^{1,3,5} and Annick Lesne^{1,3,4,5,†}

¹Laboratoire de Physique Théorique de la Matière Condensée, CNRS UMR 7600, UPMC University Paris 06, F-75005, Paris, France

²Institut de Génétique Humaine, CNRS UPR 1142 Montpellier, France

³Institut de Génétique Moléculaire de Montpellier, CNRS UMR 5535, University of Montpellier, Montpellier, France

⁴Kavli Institute for Theoretical Physics China, CAS, Beijing 100190, China

⁵CNRS GDR 3536, UPMC University Paris 06, F-75005, Paris, France

(Received February 27, 2014; revised manuscript received June 25, 2014)

Abstract Investigating average thermodynamic quantities is not sufficient to understand conformational transitions of a finite-size polymer. We propose that such transitions are better described in terms of the probability distribution of some finite-size order parameter, and the evolution of this distribution as a control parameter varies. We demonstrate this claim for the coil-globule transition of a linear polymer and its mapping onto a two-state model. In a biological context, polymer models delineate the physical constraints experienced by the genome at different levels of organization, from DNA to chromatin to chromosome. We apply our finite-size approach to the formation of plectonemes in a DNA segment submitted to an applied torque and the ensuing helix-coil transition that can be numerically observed, with a coexistence of the helix and coil states in a range of parameters. Polymer models are also essential to analyze recent *in vivo* experiments providing the frequency of pairwise contacts between genomic loci. The probability distribution of these contacts yields quantitative information on the conformational fluctuations of chromosome regions. The changes observed in the shape of the distribution when the cell type or the physiological conditions vary may reveal an epigenetic modulation of the conformational constraints experienced by the chromosomes.

PACS numbers: 64.60.an, 87.15.hp

Key words: coil-globule transition, helix-coil transition, chromosome conformational capture, two-state model

1 Introduction

Recent advances of nanotechnologies and molecular-scale experiments, for instance single-molecule micro-manipulations, FRET (fluorescence resonance exchange transfer) or chromosome conformational capture, give a direct access to the microscopic conformation of macromolecules.^[1–2] Data on the features of finite-size complexes are now available, and the notions of conformation, conformational fluctuations and conformational transitions of biopolymers are becoming central for unravelling their biological functions. While opening a way to a mechanistic understanding at the scale of individual molecules, the interpretation of such experimental data requires to stick to a finite-size framework. A straightforward application of the standard thermodynamic framework, in terms of statistical averages, to the conformational transitions of isolated macromolecules is highly questionable. The relevant transitions mostly involve a finite chain (e.g. the chain of amino-acids making a pro-

tein) or a finite region of a linear filament (e.g. the DNA stretch embedding a gene and the associated regulatory sequences), and strong finite-size effects are expected.

Linear polymer chains differ from standard many-body systems explored in statistical physics: connectivity constraints make the polymers live in a very complex conformational space. The number N of monomers is not only a measure of the linear size of the polymer, it also controls their spatial extension in a non extensive way. The limit $N \rightarrow \infty$ does not have the same rigorous status as a thermodynamic limit, defined e.g. in Ref. [3]. Lanford–Dobrushin–Ruelle conditions expressing the insensitivity to surroundings of the state of a large enough subsystem, in turn ensuring the existence of a well-defined thermodynamic limit and the extensivity of global thermodynamic quantities like free energy or susceptibilities, fail to be satisfied.

The occurrence of conformational changes in a localized region of a filament is in particular encountered in a

*This work has been initiated during the program “Small Systems Nonequilibrium Fluctuations, Dynamics and Stochastics, and Anomalous Behavior” of the Kavli Institute for Theoretical Physics China at the Chinese Academy of Sciences (KITPC), held in Beijing in July 2013. It has been funded by the French Institut de la Recherche Médicale, under Grant MICROMEGAS PC201104, and the French Institut National du Cancer, under Grant INCa-5960, and the French Agence Nationale de la Recherche under Grant No. ANR-13-BSV5-0010-03. UPMC belongs to Sorbonne Universités

†Corresponding author, E-mail: lesne@lptmc.jussieu.fr

biological context within chromatin loops, a higher-level structure displayed by genomic DNA. Actually, in higher organisms (specifically eukaryotic organisms, whose cells possess a nucleus), the DNA molecule is first wrapped around protein cores, forming basic units called nucleosomes. The bead-on-string structure thus formed is then folded into a fiber, the chromatin fiber, itself organized in loops, up to the whole chromosome. Chromatin is thus the intermediary level of genomic organization between DNA and chromosome. Each level can be described in an effective way as a filament, within the framework of polymer physics. The spontaneous conformational fluctuations at each level provide the dynamic setting in which specific biological events occur. An open question is to unravel up to what point the conformational constraints condition the occurrence, coordination and regulation of the various functional processes involving the genome, for instance gene transcription.

In this paper, we demonstrate that an operational framework is to investigate the size-dependent distribution of an appropriate order parameter, which allows to describe quantitatively conformational transitions and their nature. We will see that it explains the experimentally or numerically observed coexistence of two distinct sets of conformational states. We apply this framework to complex conformational transitions of biomolecules, for instance the formation of plectonemes on DNA and the epigenetic regulation of the constrained conformational fluctuations of chromosome regions.

2 Conformational Transitions of a Linear Polymer

2.1 A Basic Example: the Coil-Globule Transition

In a polymer solution dilute enough to neglect interactions between the different polymers, their conformation can be investigated as if each one were isolated. A first important parameter is the number N of monomers forming the polymer, directly related to its length. However, we will see that $N \rightarrow \infty$ has not the status of a thermodynamic limit, and N plays a more complex role in the physical behavior of the polymers.

The solvent also has an essential influence on the polymer conformation. If the affinity between the solvent and the monomers is high (good solvent), the polymer adopts a coil conformation. When the affinity between the solvent and the monomers decreases below the mutual affinity between the monomers (bad solvent), the polymer collapses onto itself and adopts a compact globule conformation. Actually, under the action of thermal noise, the polymer conformation fluctuates. Accordingly, it is usually described by its statistical features, for instance the root-mean-square end-to-end distance R or the root-mean-square radius of gyration R_g (averages at thermal equilibrium over the whole conformational space).^[4–5] In a good solvent, these quantities scale with the chain length N according to $R(N) \sim R_g(N) \sim N^\nu$ where $\nu \approx 3/5$ is the Flory exponent. In contrast, they scale

as $R(N) \sim R_g(N) \sim N^{1/3}$ in a bad solvent. The scaling is different at the transition point (the so-called θ -point): $R(N) \sim R_g(N) \sim N^{\nu_\theta}$.^[6–7]

However, these average behaviors do not reflect the instantaneous conformation adopted by the finite-size polymer nor the range of its fluctuations, that are observed in a single-molecule experiment or a numerical simulation. Neither do they give insights on what occurs at the coil-globule transition in terms of conformational space exploration. We propose that a better understanding of the finite-size conformational transition is provided by investigating the *distributions* of various quantities describing the individual polymer conformation. Such distributions give account of the actual molecular conformations adopted by the polymer and how often they are encountered. We will show that the changes observed in the shape of the distributions as a control parameter varies is highly informative about the transition and its nature.

The coil-globule transition is controlled by the relative strength of the solvent-monomer interactions, that is, by the excluded volume of a monomer, in which the center of no other monomer could enter: the excluded volume is large in a good solvent and it decreases at decreasing solvent quality. Such a decrease of the solvent quality is observed in general when the temperature decreases.^[4–5] In what follows we will thus investigate the changes in the conformational statistics observed when varying the temperature.

2.2 Finite-Size Signature of the Coil-Globule Transition

In a previous study,^[8] we evidenced that the relevant microscopic order parameter t for describing the coil-globule transition of an isolated polymer chain of length N with excluded-volume interactions is a power $t = \rho^{1/(\nu d - 1)}$ of the density $\rho = N r_g^{-3}$, where r_g is the radius of gyration of the individual chain conformation and $\nu \approx 3/5$ is the Flory exponent. In dimension $d = 3$, it comes

$$t = (N/r_g^3)^{5/4}. \quad (1)$$

The larger t , the more compact the conformation is. The distribution $P_N(t)$ at infinite temperature has been deduced from scaling arguments supplemented with numerical simulation (Monte Carlo sampling on a cubic lattice).^[8] Choosing an energy $U = -NJt$ where J is a coupling constant, that accounts within a mean-field approximation for attractive interactions at contacts,^[9] the Boltzmann–Gibbs distribution $P_N^{(\beta)}(t)$ of the chain conformation is:

$$P_N^{(\beta)}(t) \sim t^c e^{-A'(Nt)^{-q}} e^{-N[(A - \beta J)t + Bt^n]}, \quad (2)$$

where A, A', B, c , and n are numerical constants fitted on simulation data, with $n \approx 2$ and $c \approx -1.13$. We introduce a reduced temperature:

$$\tau = 1 - \theta/T = 1 - \beta J/A. \quad (3)$$

According to the sign of τ , the dominant contribution to $P_N^{(\beta)}(t)$ is located in different domains of values of t . For

$\tau > 0$, the distribution is strongly peaked around a value of t of order $1/N$, corresponding to values of r_g of order N^ν , which leads to identify this high temperature regime with a *coil phase*. For $\tau < 0$, $P_N^{(\beta)}(t)$ is now peaked in the region where t is far larger, of order 1, corresponding to values of r_g of order $N^{1/d}$, which leads to identify this low temperature regime with a *globule phase*. The temperature $\theta = J/k_B A$ gives a rough estimate of the transition temperature.

Focusing on the transition, we investigate more precisely the size and temperature dependence of the distribution $P_N^{(\beta)}(t)$. It involves a factor:

$$h(\hat{\tau}, \hat{t}) = \hat{t}^c e^{-A\hat{\tau}\hat{t} - B\hat{t}^n}, \quad (4)$$

that is scale invariant insofar as it depends on the size N

through the rescaled variables:

$$\hat{t} = tN^{1/n} \quad \text{and} \quad \hat{\tau} = \tau N^{1-1/n}, \quad (5)$$

with $n \approx 2$. The distribution now writes:

$$\hat{P}_N(\hat{\tau}, \hat{t}) = \frac{h(\hat{\tau}, \hat{t}) e^{-A'N^{-q(1-1/n)}\hat{t}^{-q}}}{\mathcal{I}_c(N, \hat{\tau})}, \quad (6)$$

where the normalization factor $\mathcal{I}_c(N, \hat{\tau})$ ensures that $\int_0^\infty \hat{P}_N(\hat{\tau}, \hat{t}) d\hat{t} = 1$. At first sight, one might presume that a scaling regime would be obtained at fixed values of $\hat{\tau}$ when $N \rightarrow \infty$. However, due to the value $c < -1$, the limiting function $h(\hat{\tau}, \hat{t}) = \hat{t}^c e^{-A\hat{\tau}\hat{t} - B\hat{t}^n}$ is not integrable in $\hat{t} = 0$. The relevance of the size-dependent contribution factor $e^{-A'N^{-q(1-1/n)}\hat{t}^{-q}}$ in $\mathcal{I}_c(N, \hat{\tau})$ breaks the scale invariance, and compels to focus on the finite-size distribution $\hat{P}_N(\hat{\tau}, \hat{t})$.

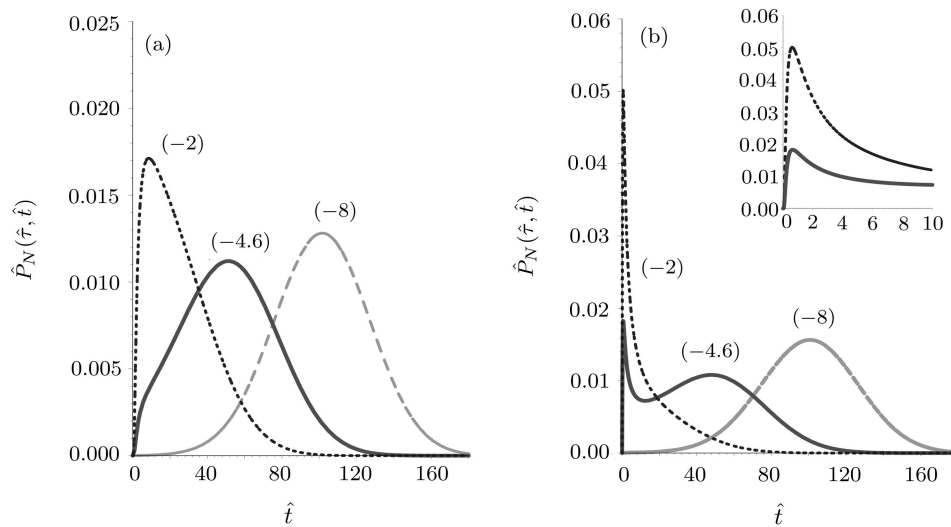


Fig. 1 (a) Continuous coil-globule transition for a polymer of small size $N < N_0 = 45$ (here $N = 20$). The evolution of the shape of $\hat{P}_N(\hat{\tau}, \hat{t})$ as the rescaled temperature $\hat{\tau}$ decreases (respectively $\hat{\tau} = -2$ (dotted line), $\hat{\tau} = -4.6$ (solid line) and $\hat{\tau} = -8$ (dashed line)) indicates a continuous transition in which the characteristics of a single population evolves smoothly with $\hat{\tau}$. (b) Evidence of a first-order-like finite-size coil-globule transition on the shape of $\hat{P}_N(\hat{\tau}, \hat{t})$, plotted with respect to the rescaled variable \hat{t} at fixed $N > N_0 = 45$ (here $N = 2000$) and for various $\hat{\tau}$. For $\hat{\tau} = -2$ (dotted line), only a coil peak is present. For $\hat{\tau} = -8$ (dashed line), only a globule peak is present. The inset shows an enlarged view of the coil region ($\hat{t} < 10$). The presence of two well-separated peaks in the distribution for $\hat{\tau} = -4.6$ (solid line) reveals the first-order-like coexistence of two distinct populations in dilute solution.

For N smaller than $N_0(c) \sim |c|^{-4}$ ($N_0 = 45$ for $c = -1.13$), the distribution $P_N^{(\beta)}(t)$ displays only one peak, which slowly shifts from the coil region towards the globule region as temperature decreases, as seen on Fig. 1(a). In this case, the coil-globule transition is continuous and no transition temperature can be clearly defined. For $N \geq N_0(c)$, a globule peak is present for rescaled temperatures $\hat{\tau} < \hat{\tau}_g(N)$, where $\hat{\tau}_g(N)$ slightly decreases from a critical value $\hat{\tau}_0 = \hat{\tau}_g(N_0) < 0$ to an asymptotic value $\hat{\tau}_g(\infty) \sim -|c|^{(n-1)/n}$ (with $n \approx 2$). The actual parameter value $c = -1.13$ gives $\hat{\tau}_0 \approx -3.2$ and $\hat{\tau}_g(\infty) \approx -4$. A coil peak exists as soon as $\tau > \tau_c(N)$ where $\tau_c(N)$ rapidly increases from τ_0 towards an asymptotic value $\tau_c(\infty) < 0$ independent of N (here $\tau_c(\infty) \approx -0.4$). The rescaled

bound $|\hat{\tau}_c|$ thus behaves as $N^{1-1/n}$, so that a coil peak always exists in the scaling region ($\hat{\tau}$ finite) for N large enough. Accordingly, $P_N^{(\beta)}(t)$ displays two peaks in some range of temperatures. The bimodal shape of this finite-size single-molecule distribution is straightforwardly related to the statistics describing a dilute solution: it indicates that a coil population and a globule population coexist in the interval of rescaled temperatures $[\hat{\tau}_c(N), \hat{\tau}_g(N)]$. As represented Fig. 1(b), the transition at fixed N proceeds through an exchange of importance between the two peaks. A crucial point is that their positions \hat{t}_c and \hat{t}_g remain well-separated when temperature varies: they are located on each side of a value $\hat{x}(N)$ increasing with N from \hat{x}_0 towards an asymptotic value $\hat{x}_\infty = [|c|/n(n-1)B]^{1/n}$

(in our case, $\hat{x}_0 \approx 13$ and $\hat{x}_\infty \approx 26$). Accordingly, our model predicts a first-order-like coil-globule transition for polymer sizes $N_0 < N < \infty$.

2.3 Assessing the Order of the Conformational Transition

A first caveat concerns the experimental reality of the coil-globule coexistence. It is actually possible to distinguish two populations only if the height of $\hat{P}_N(\hat{\tau}, \hat{t})$ at the minimum \hat{t}_m located between \hat{t}_c and \hat{t}_g differs significantly from the height of the peaks. We have checked that it is true for N not too small. It is then sensible to partition the configuration space in two disjoint macro-states: a coil state $\{\hat{t} < \hat{x}(N)\}$ and a globule state $\{\hat{t} > \hat{x}(N)\}$ as the peaks remain located on each side of $\hat{x}(N)$ as soon as they exist, even alone. Coexistence is actually observed if the fractions of molecules in each state, i.e. the areas of the two peaks, have comparable values. The ratio

$$\kappa(\hat{\tau}, N) = \frac{\text{globule peak area}}{\text{coil peak area}} \quad (7)$$

is the equilibrium constant of the transition between the two states. This view recovers the standard *two-state model* developed to describe the conformation of macromolecules. In particular, the equilibrium constant κ can be deduced from various experimental (e.g. calorimetric) data.^[10]

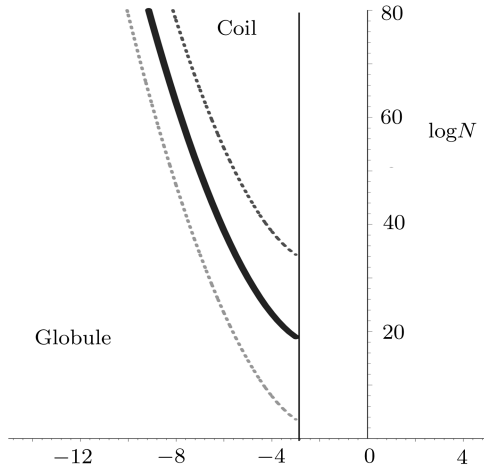


Fig. 2 Phase diagram of the polymer conformation in $(\hat{\tau}, \log N)$ -space. The vertical straight line $\hat{\tau} = \hat{\tau}_g$ bounds above the temperature domain where a well-identified globule state (a globule peak) exists. The bold curve $\hat{\tau}_{\text{coex}}(N)$ corresponding to the coexistence in equal proportions of coil and globule populations ($\kappa = 1$ in Eq. (7)) behaves as $\sqrt{\log N}$ for large N . The thin dotted curves bound the coexistence region ($\kappa = 10$ on the globule side and $\kappa = 0.1$ on the coil side), which displays a width $\Delta\hat{\tau}(N) \sim 1/\sqrt{\log N}$ for large N . A first-order coil-globule transition occurs when $\hat{\tau}$ increases at fixed N or when N increases at fixed $\hat{\tau} < \hat{\tau}_g$.

The coexistence curve $\hat{\tau}_{\text{coex}}(N)$, corresponding to the coexistence condition $\kappa(\hat{\tau}, N) = 1$, and the associated phase diagram are shown in Fig. 2. Coexistence in equal

proportions of coil and globule phases is observed only for large enough chains, and $\hat{\tau}_{\text{coex}}(N)$ behaves as $\log N$ for large N . The coexistence region has a width $\Delta\hat{\tau} \sim 1/\sqrt{\log N}$: it tends to 0 as $N \rightarrow \infty$, so that it makes sense to speak of a phase transition in the infinite-size limit.^[11] Another signature is the behavior of the densities along the coexistence curve $\lim_{N \rightarrow \infty} \rho_g/\rho_c = \infty$, showing that the physical difference between the phases increases with the size N . As expected, $\tau_{\text{coex}} = \hat{\tau}_{\text{coex}} N^{-(1-1/n)}$ tends to 0 as $N \rightarrow \infty$, supporting the estimate $\theta = J/k_B A$ of the transition temperature. The shape of the coexistence curve shows that N is not only the size but also a control parameter ruling the transition: increasing N at fixed $\hat{\tau}$ leads into the coil phase.

Coming back to the unscaled variable t , the minimal distance between the two peaks in the coexistence region satisfies $\Delta t < \hat{x} N^{-(1-1/n)}$ (with bounded \hat{x}), hence tends to 0 as N tends to infinity. The globule density right at the transition point tends to 0 as $\rho_g \sim N^{-2/5}$. This means that in the infinite-size limit, the transition occurs at $\tau = 0$ and both coil and globule densities vanish. In this respect, in infinite-size, this coil-globule transition displays some features of a second-order transition. Also the shape of the mean order parameter $\langle t \rangle$ with respect to the reduced temperature τ has the characteristic shape of a second-order transition, as shown in Fig. 3. Nevertheless, the transition occurs through the coexistence of a coil population, whose statistics is controlled by the size N , and a globule population, whose statistics is scale-invariant and controlled by the rescaled temperature $\hat{\tau}$. The first-order nature of the transition originates in the incompatible scale behaviors of the two sets of conformations. Hence, it is likely to be observed whatever the underlying model, provided the shape of the distribution $P_N(t)$ (infinite-temperature distribution, describing the purely entropic contribution), gives enough weight to the coil region.

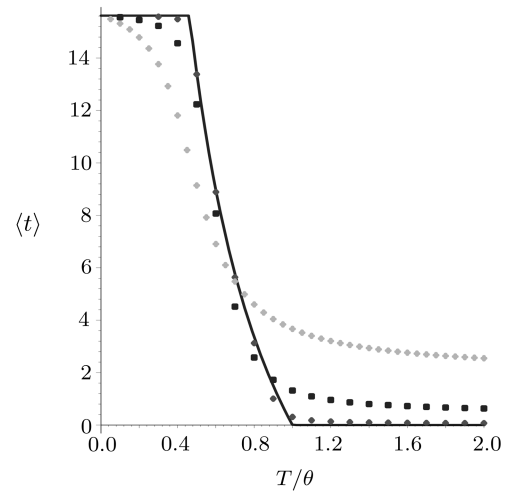


Fig. 3 Plot of the average order parameter $\langle t \rangle$ with respect to the relative temperature T/θ for $N = 20$ (+, light grey), $N = 100$ (□, black) and $N = 1000$ (◇, dark grey); in the limit as $N \rightarrow \infty$, the curve exhibits the typical shape of a second-order transition, recovering the thermodynamic prediction.^[6–7]

Our point is that the coexistence of two populations in the transition region cannot be detected on the behavior of $\langle t \rangle$. This shortcoming of the standard approach based on average thermodynamic quantities, and the ensuing need to investigate the order parameter distribution are not only due to the finite length N of the macromolecule. Rather, it originates in the special status of this length variable N . De Gennes has shown a formal analogy between the statistical scaling properties of a linear polymer and those predicted by the n -vector model for a system of spins (with a spin value $||\vec{s}||^2 = n$), in the limit $n \rightarrow 0$.^[4–5] In this analogy, the end-to-end distance plays the role of the spin system correlation length and the inverse polymer size $1/N$ the role of the reduced temperature, i.e. the control parameter. Flory exponent is then analogous to the critical exponent ν of the correlation length. The formulation of this analogy is quite technical and out of our scope. However, it underlines that in adapting concepts developed for describing phase transitions of spatially extended many-body systems to the conformational transitions of a macromolecule, the length N of the molecule does not at all play the role of a plain extensive variable of reference like the volume or the number of elements. Accordingly, the limit $N \rightarrow \infty$ has not the status and the properties of a standard thermodynamic limit.^[3]

The first-order features of the finite-size coil-globule transition allows to describe this transition within a two-state model, well-suited for analyzing experimental observations. The bimodal probability distribution reflects the fraction of macromolecules in each conformational state in a dilute solution. Experimental data available in the literature confirm the above theoretical scenario. For instance,

Yoshikawa, Baigl and their collaborators have experimentally investigated the conformation (in dilute solution) of DNA segments, marked all along their length with fluorescent probes for direct visualization.^[12–14] They evidenced that DNA displays a coil-globule transition when the concentration of an additional chemical in the solution increases, by observing a change in the DNA length distribution, passing from a distribution peaked at large lengths to a distribution peaked in the small-length region. In the transition region, the length distribution displays a bimodal shape with two well-separated coil and globule peaks, corresponding to the coexistence of two populations of macromolecules, either in the coil or in the globule state. The transition is revealed as an exchange of importance between the peaks while their positions do not change significantly.^[14] As underlined above, such experimental two-state coexistence cannot be accounted for in the standard thermodynamical picture.^[6–7]

2.4 A Finite-Size Framework: the Order Parameter Distribution

Our study demonstrates that in numerical simulation studies of macromolecule conformational transitions, the relevant quantity to be analyzed is the shape of distribution of an order parameter (denoted generally a in Fig. 4) when a control parameter T varies. This control parameter is not necessarily the temperature but for instance some feature of the solvent or the strength of a constraint applied to the polymer. The thermodynamic approach relying on the analysis of the distribution moments is not sufficient to reveal the finite-size features of the transition nor its nature.

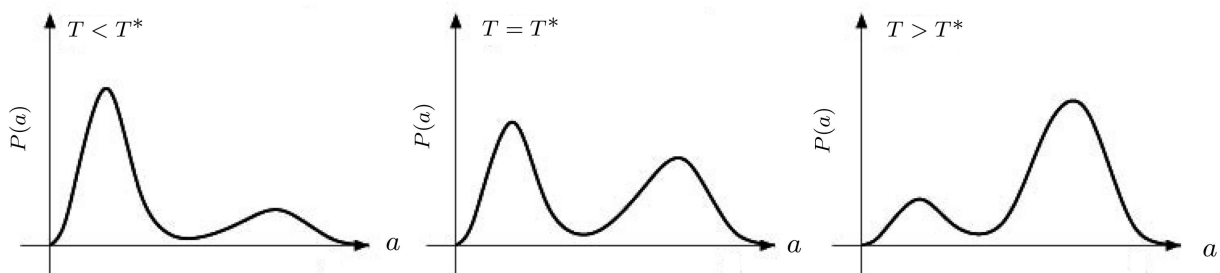


Fig. 4 Empirical notion of first-order conformational transition in finite size (sketch). What makes sense for a finite-size system, in particular a biological system, is the distribution $P(a)$ of some molecular property a . The first-order signature is the exchange of importance between two well-separated peaks at roughly fixed positions, when the control parameter T (for instance, but not necessarily, the temperature) varies. Strikingly, the observed value a_{obs} , the average value $\langle a \rangle$, and the most probable value a_{max} do not match in the transition region.

The framework presented here also provides a guideline for experimental studies in suggesting to investigate the distribution $P_T(a)$ of some finite-size, molecular quantity a as a control parameter T varies. For instance, single-

molecule techniques^[1] allow a direct determination of the distribution of the length of the molecule.

The standard two-state model used in an experimental context for the description of macromolecule confor-

mations and conformational transitions^[10] is recovered in case of a first-order-like transition, associated with a bimodal order parameter distribution $P_T(a)$ and an exchange of importance between two peaks as the control parameter T varies. It is to note that the peaks of $P_T(a)$ correspond to the wells of the free energy $F_T(a) = -\log P_T(a)$. The set of conformations corresponding to each peak of $P_T(a)$ (respectively each well of $F_T(a)$) defines each of the two states, providing a microscopic rooting of this effective model. Importantly, such a mapping onto a two-state model relates the theoretical description based on the knowledge of the configurational statistics and the experimental observations through a coarse-graining of the configuration space, and not through a thermodynamic limit.

3 DNA Conformational Transitions

We mentioned in the previous section the experimental observation of a coil-globule transition for DNA when monitoring the chemical composition of the solvent,^[14] which is compatible with the prediction of our theoretical investigation. We here extend the scope of these studies, with the aim of understanding conformational transitions of DNA and their biological role, that is, the role they play in the fulfillment and regulation of biological functions.

3.1 An Interplay of Mechanics, Electrostatics, Topology and Thermal Noise

DNA is more complicated than the model linear polymer considered in the previous section. It is a polyelectrolyte, with a negative density of charge, meaning that electrostatics matters, for instance most DNA features display a salt dependence.^[15] Observations of knotted DNA, supplemented with single-molecule experiments have revealed DNA non trivial elastic properties, partly coupled to electrostatics. As regards its response to an applied force, DNA can be considered as a semi-flexible polymer described within the worm-like-chain model and characterized by its (bending) persistence length.^[5] However,

DNA also displays torsional elasticity, characterized by a second (twist) persistence length. This additional feature implies that DNA conformation and mechanical properties will be affected by topological constraints. A closed or fixed-end stretch of DNA possesses a topological invariant: its linking number L_k defined as the number of turns one strand makes around the other. The number of extra turns with respect to the relaxed state (with linking number L_k^0) defines the supercoiling $\sigma = (L_k - L_k^0)/L_k^0$. The conservation of L_k reflects in particular in the trade-off between torsion and formation of plectonemes (braids reminiscent of those observed in old telephone wires).

Overall, DNA conformational properties thus follow from geometric, mechanical, electrostatic and topological constraints, in interplay with thermal noise and the action of specific biological factors like DNA-binding proteins, inducing local structural deformations, or topo-isomerases able to relax torsional stress. We argue that accordingly, our framework based on the analysis of the distribution of finite-size quantities will be more insightful than a thermodynamic approach based on the analysis of asymptotic (infinite-size) average quantities.

3.2 Helix-Coil Transition upon Application of a Torque

Single-molecule micro-manipulations with magnetic tweezers, now available, allow to investigate the response of DNA to an applied torque.^[16] The increase in torsional energy density monitored by the magnetic tweezers triggers the appearance of plectonemes and a helical state in the considered DNA segment (Figs. 5B, 5C, and 5D). Such experiments have been recently reproduced *in silico*, using novel simulation method. In this simulation, DNA is represented as an array of small rigid sticks linked by articulated joints accounting for the mechanical and topological constraints experienced by DNA (Fig. 5A). The dynamics is based on physics engines used in video games supplemented with thermal noise.^[17]

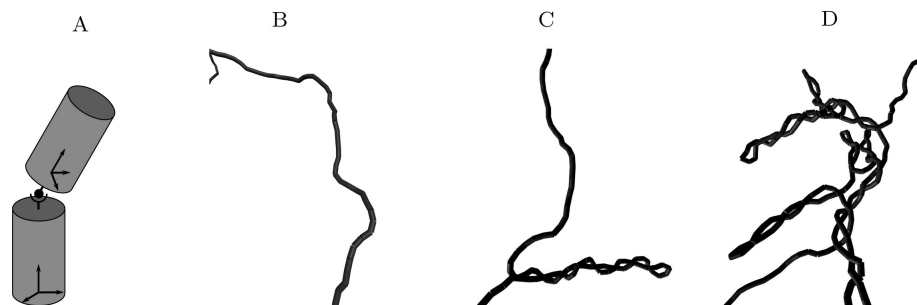


Fig. 5 DNA helix-coil transition observed in an *in silico* magnetic tweezer experiment.^[17] (A) DNA is seen as a flexible but physically constrained filament, and modeled as a chain of short rigid sticks linked by ball-in-socket joints. (B–D) At increasing applied torque Γ (from left to right), the supercoiling σ of the DNA segment increases and plectonemes appear in the DNA conformation (see also Fig. 6).

Here the control parameter is the applied torque Γ , at fixed applied stretching force, and the finite-size order parameter is the supercoiling σ or the relative extension λ of the DNA segment. The helix-coil transition is revealed by the exchange of weight between the coil region (low σ) and the helix region (high σ) on either the supercoiling distribution $P_N^{(\Gamma)}(\sigma)$ or the length distribution $P_N^{(\Gamma)}(\lambda)$ when the torque Γ increases (Fig. 6). The presence of two well-separated peaks in $P_N^{(\Gamma)}(\sigma)$, identified respectively with a coil state and helix state, demonstrates the coexistence of two distinct populations in the transition region (around $\Gamma_c = 9.8$ pN·nm for an applied

stretching force of 0.74 pN). The same two-state signature is observed on $P_N^{(\Gamma)}(\lambda)$ (Fig. 6). Although a transition is visible on the behavior of the average supercoiling $\langle \sigma \rangle$ as a function of the applied torque Γ (with a plateau at the transition value Γ_c , as presented in Ref. [17]), only an investigation of the finite-size distribution $P_N^{(\Gamma)}(\sigma)$ allows to pinpoint what occurs in terms of individual DNA conformation at the transition. These numerical results recover the experimental observations by Brutzer *et al.*^[16] on the distribution of the relative extension of a finite segment of DNA at varying torque values, in a magnetic tweezer experiment.

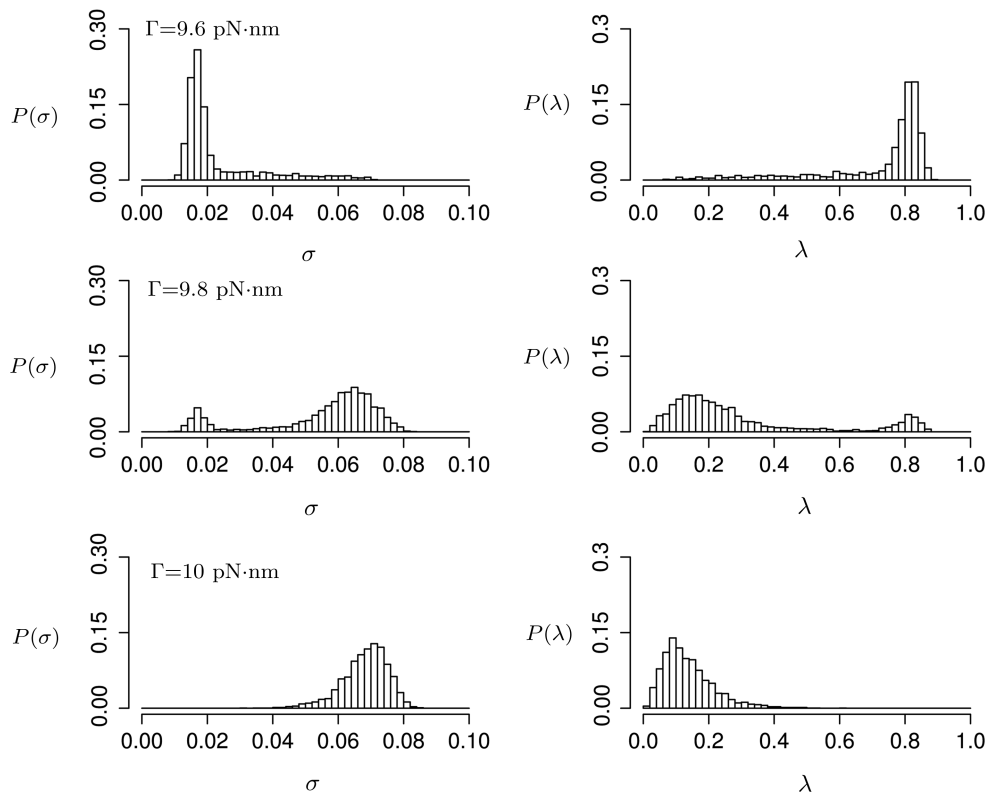


Fig. 6 *In silico* first-order helix-coil transition of a finite-size DNA segment. (Left column) Simulated distribution $P_N^{(\Gamma)}(\sigma)$ of DNA supercoiling σ at fixed applied torque Γ and DNA length N . A transition between a coil state (top) and a helix state where DNA forms plectonemes (bottom) is observed when the applied torque Γ is varied at fixed applied stretching force (of 0.74 pN) and fixed length (top: $\Gamma = 9.6$ pN·nm, middle: $\Gamma = \Gamma_c = 9.8$ pN·nm, bottom $\Gamma = 10$ pN·nm). The first-order nature of the finite-size helix-coil transition is assessed by the coexistence of two populations at the transition, reflecting in the bimodal shape of $P_N^{(\Gamma)}(\sigma)$ in the transition region. (Right column) Distribution $P_N^{(\Gamma)}(\lambda)$ of the DNA relative extension λ at fixed applied torque Γ (same simulation and Γ values as in the left column) and a fixed applied stretching force of about 1 pN. The distribution similarly displays a bimodal shape in the transition region, and the transition corresponds to an exchange of importance between the peaks.

4 DNA Higher Levels of Organization and Their Conformations

As underlined in introduction, DNA is only the first level of organization of the genome within an eukaryotic cell. We now consider the next level of organization, namely the chromatin fiber, and its tridimensional con-

formation within the cell nucleus.

4.1 Conformational Transitions of the Chromatin Fiber

A challenge in genomics is to understand the functional DNA folding into a chromatin fiber, capable at the

same time to achieve a high degree of DNA compaction and to preserve the access and binding to genes (or regulatory DNA sequences) of various biological factors and machineries. An essential step is to quantitatively determine the physical properties of the chromatin fiber,^[15,18] namely its bending and twisting persistence lengths as well as its compaction α (in $\text{bp}\cdot\text{nm}^{-1}$). Integrated scenarios can describe theoretically how the interplay between the physical features of the chromatin fiber conformation and the action of specific biological factors could achieve genomic functions.^[19] Here again, the finite size of the involved genomic regions and the finite number of interacting factors have to be taken into account. The inherent stochasticity generated by thermal noise is still perceptible in the observed processes, which promotes approaches based on the probability distribution of local features.

4.2 *In Vivo Experiments: Chromosome Conformational Capture Technologies*

Recent experimental techniques achieve the *in vivo* measurement of pairwise contacts between genomic loci, that is, the identification of pairs of DNA sites that are distant along DNA but *in vivo* very close in the tridimensional space. The initial technology, known as chromosome conformational capture (3C),^[2] has been improved into quantitative 3C-qPCR^[20] or genome-wide (but at a coarser resolution) HiC experiments.^[21] Contrary to single-molecule experiments, these technologies provide an ensemble average over a cell population of tridimensional DNA-DNA contacts at a given time. However, according to an ergodic hypothesis, the data are presumed to reflect time average fluctuations of the chromatin tridimensional conformation, and thus to provide local information about the chromatin conformational dynamics. Overall, they open a new direction to investigate the biological role of chromatin fiber conformational transitions. As such, they require a dedicated finite-size framework for their formulation and interpretation.

Raw data consist in numbers of contacts between well-identified genomic loci. In 3C technology, the considered pairs of loci are composed of a given DNA site and sites located at a varying distance s along DNA (Fig. 7, see also Ref. [20]). In contrast, HiC technology considers pairwise contacts within a set of loci spanning the whole genome, however sequencing cost prevents to reliably identify all possible pairs and only a sample of contacts is actually obtained.^[21] Raw data are then normalized by a reference number of contacts^[22] and processed into a curve $s \rightarrow q(s)$ describing for each s the *relative contact frequency* $q(s)$ for DNA loci separated by a genomic distance s (where s is measured in number of DNA base pairs). Although $q(s)$ is obtained in experiments over a cell population, i.e. for a population of macromolecules, the ergodic assumption allows to interpret $q(s)$ as a quantity accumulated over time for a single chromatin fiber. It thus essentially captures a finite-size feature of the chromatin fiber, namely its finite-range looping in a living cell.

3C observation of complex contact distributions $q(s)$ in mammals reveals the presence of constraints on chromatin conformational fluctuations, biasing its random exploration of the conformational space upon the effect of thermal noise.^[23] The conformational constraints would also reflect in the average chromatin conformation. As 3C data are not directly meaningful in terms of tridimensional structures, the statistical shape of the considered chromatin segment is derived through the fit of a polymer model.^[2] The idea is to consider a polymer model of the chromatin fiber, and to fit its prediction $q_{\text{th}}(s)$ for the contact frequency at a genomic distance s with the experimental data. The first step is to compute the statistical average $\bar{q}(L)$ over the polymer conformations (at equilibrium) of the fraction of instances where two given sites, separated by a distance L along the filament modeling the chromatin fiber, are in contact in the tridimensional space. The value $\bar{q}(L)$ varies between 0 (its value if a contact would never occur between the sites) and 1 (its value if the sites would be permanently grafted). The second step is to relate the separation L along the chromatin and the separation s along the embedded DNA, from which $q_{\text{th}}(s) = \bar{q}[L(s)]$ follows. The prediction of the simplest semi-flexible polymer model writes:^[2]

$$q_{\text{th}}(s) \sim [L(s)/L_p]^{-3/2} \exp[-2(L_p/L(s))^2],$$

where L_p is the fiber bending persistence length and $L(s) = s/\alpha$. The fit of this formula on the experimental distribution $q(s)$, up to an overall normalization factor, yields the fiber bending persistence length L_p and compaction α . However, this formula does not take into account excluded volume, nor possible pre-constraints (e.g. intrinsic bending) of the polymer, and more refined polymer models have to be developed. In particular, using a model accounting for helical constraints, we have shown that some portions of the mammalian chromatin, encompassing few hundred of kilo-bases (kb) of DNA (and deprived of locus-specific interactions), tend to adopt, statistically, a helix shape.^[23] The issue is now to identify the physical constraints and the biological factors biasing the conformational space exploration.

This analysis would strictly relate to our proposed framework, if we would consider a *probability distribution* $P(s)$, that is, a normalized curve $s \rightarrow P(s)$ such that $\sum_s P(s) = 1$. The value $P(s)$ would then give the contribution of distance s (in base pairs) to the DNA-DNA contacts within an *in vivo* chromatin fiber. In this view, the genomic distance s would appear as an individual order parameter, roughly describing the genomic size of a chromatin loop, and $P(s)$ as its probability distribution. Note that s is not exactly an order parameter since it takes several values in a single conformation, however $P(s)$ is a true probability distribution. The analysis of such a distribution would determine which genomic distances contribute most to contacts, i.e. whether the chromatin folding consists in short loops or large-scale loops. In particular, it could be interesting to evaluate the weight $P(s > s_0)$ of long-range contacts (for some

threshold s_0) and to investigate how it is modulated by the presence of specific interactions or epigenetic modifications. As in the case of a plain polymer, changes in the shape of $P(s)$ would reflect conformational transitions of the chromatin fiber. However, achieving the normalization of the experimental curves is delicate, since the abscissas s of the experimental points are prescribed by the technology (specific DNA sites responding to the action of an enzyme) and distributed in a non controlled and heterogeneous way. We suggest that a possibility would be to consider a model $q_{th}(s)$ fitting the experimental curve $q(s)$; in the model, all values of s could be considered, hence $q_{th}(s)$ could be normalized into a true probability distribution $P_{th}(s) = q_{th}(s) / \sum_{s'} q_{th}(s')$.

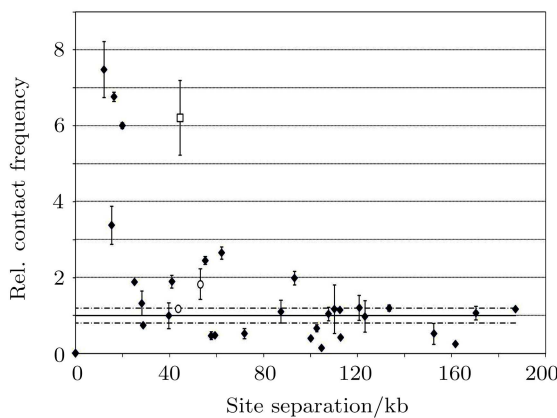


Fig. 7 Using the 3C-qPCR approach,^[20,23] the relative contact frequencies are measured along the *Pyruvate dehydrogenase beta* (Pdhb) gene locus, on mouse chromosome 14. Globally, contact frequencies are observed to decrease as site separation is increasing. Noteworthy, a locus-specific interaction is visualized as a local peak in the graph (point underlined as a white square), where contact frequency is much higher than that measured at surrounding sites (points underlined as white circles). Errors bars are obtained with three independent biological replicates.

In any cases, the shape of the curves $q(s)$ and $P(s)$ are the same, hence it is informative to investigate the s -dependence of the curve $q(s)$.

3C experiments evidence contacts following from chromatin folding and loop formation (Fig. 7). The physical contacts are often stabilized by locus-specific factors bound locally to specific DNA sequences. Such locus-specific interactions are detected in 3C-qPCR experiments as local peaks that display high contact frequencies compared to the surrounding regions. Figure 7 displays an example of a locus-specific interaction observed in a 3C-qPCR experiment performed in mouse liver cells at the Pdhb gene locus, on chromosome 14. 3C experiments moreover show that the tridimensional conformation of the chromatin in mammals is actually very dynamic. For example, at the Igf2/H19 gene locus, on the mouse chromosome 7, we have shown that chromatin can alternate between two mutually exclusive conformations mediated by locus-specific interactions.^[24]

Note that the simulation tools that have been recently developed for DNA (see Subsec. 3.2) can be extended, at a coarser scale, to the chromatin fiber.^[17] Figure 8 shows a snapshot of a plectoneme made up of chromatin fiber. Contacts between genomic loci inside plectonemes are stabilized by supercoiling constraints. The probability distribution function of such contacts at thermal equilibrium can be simulated and compared to available 3C data.

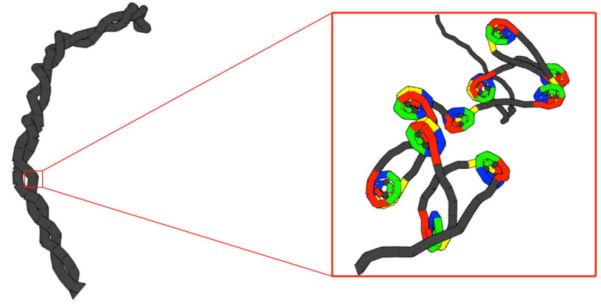


Fig. 8 (Color online) Simulation of chromatin conformational dynamics. (Left) The chromatin fiber can be seen as a constrained linear polymer, and its behavior is strongly reminiscent of DNA. (Right) An enlarged view displays what is actually simulated: rigid sticks representing either linker DNA (in black) or nucleosomal DNA (in color/grey) are linked together by articulated joints reproducing the mechanical and topological constraints experienced by DNA within the chromatin organization; the colors/levels of grey of the sticks describing nucleosomal DNA vary with the type of bound histones.

When varying the cell type (in a given organism), the contact frequencies are observed to change. This has been observed for instance in the mouse, in comparing embryonic stem cells and liver cells. The differences are best summarized in the statistical shape: the statistical helix observed in liver cells is considerably stretched in embryonic stem cells (Ea *et al.*, manuscript in preparation). The change in the probability distribution of contacts reflects a change in the physical constraints experienced by the chromosome and biasing its exploration of the conformational space. Understanding the determinants, for instance the epigenetic determinants, influencing this conformational bias, is yet a challenge. It opens a direction to investigate epigenetic regulation of genomic functions, as the occurrence of contacts is a prerequisite for specific biological interactions, involving the binding of dedicated proteins and the assembly of protein complexes.

4.3 Physical Mechanisms Mediating the Epigenetic Regulation of Gene Expression

Confrontation of chromosome conformation capture experiments with polymer model predictions allows to identify the physical component of genome dynamics and to pinpoint specific regulatory events. We suggest that physical folding is in fact a transient trend to fold, stabilized by specific interactions. In other words, physics is

what makes specific interactions possible or not. Physical constraints would control the possible events, which would then be selected and completed by specific biological factors. Such a mixed physical/biological approach opens a new research direction to understand what is termed the epigenetic regulation of gene expression. The word *epigenetic* refers to various local sequence-preserving chemical modifications of DNA or histones presumed to regulate DNA-templated processes, e.g. to underlie cell differentiation and gene regulation. 3C experiments in different cell types, or considering chromatin domains endowed with different epigenetic status, have demonstrated an epigenetic tuning of contact frequency. The challenge behind epigenetics is considerable, leading to issues such as cell differentiation (how cells with identical genomes can display such differences in shape and physiology) or pathological changes (cancer cells phenotype is not only explained by their modified genome). Important ingredients are provided by the conformational transitions experienced by the genome at each of its levels of organization, and the ensuing changes in their topological and mechanical properties.

5 Conclusion

With the advances of nanotechnologies and micro-manipulations, the notions of conformation, conformational fluctuations and conformational transitions of (bio)polymers are becoming central, with a need of developing a proper framework to investigate their finite-size behavior. Considering the infinite-size limit may yield a misleading simplified view. We have here demonstrated that an operational framework for understanding conformational transitions of a finite-size macromolecule is to study the order parameter distribution.

In a biological context, physical modeling allows to identify the constraints underlying macromolecule conformations. In particular, the functional conformational dynamics of the chromosome appears to involve an interplay between global generic physical properties and local specific biological factors, at several levels of organization. These results emphasize the importance of understanding the part of the physical properties in the chromosome biological functions.

References

- [1] F. Ritort, J. Phys.: Condens. Matter **18** (2006) R531.
- [2] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, Science **295** (2002) 1306.
- [3] D. Ruelle, *Thermodynamic Formalism, The Mathematical Structure of Equilibrium Statistical Mechanics*, Addison-Wesley Reading, MA (1978).
- [4] P.G. De Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca (1979).
- [5] M. Laguës and A. Lesne, *Scale Invariance*, Springer, Berlin (2011).
- [6] P.G. De Gennes, J. Physique Lett. **36** (1975) L55.
- [7] P.G. De Gennes, J. Physique Lett. **39** (1978) L299.
- [8] J.B. Imbert, A. Lesne, and J.M. Victor, Phys. Rev. E **56** (1997) 5630.
- [9] J.M. Victor, J.B. Imbert, and D. Lhuillier, J. Chem. Phys. **100** (1994) 5372.
- [10] C.R. Cantor and P.R. Schimmel, *Biophysical Chemistry, Part III: the Behaviour of Biological Macromolecules*, Freeman, New York (1980).
- [11] I.M. Lifshitz, A.Y. Grosberg, and A.R. Khokhlov, Rev. Mod. Phys. **50** (1978) 683.
- [12] S.M. Melnikov, V.G. Sergeyev, and K. Yoshikawa, J. Am. Chem. Soc. **117** (1995) 2401.
- [13] K. Yoshikawa and Y. Matsuzawa, Physica D **84** (1995) 220.
- [14] M. Sollogoub, S. Guieu, M. Geoffroy A. Yamada, A. Estévez-Torres, K. Yoshikawa, and D. Baigl, Chem-BioChem **9** (2008) 1201.
- [15] P. Carrivain, A. Cournac, C. Lavelle, A. Lesne, J. Mozziconacci, F. Paillusson, L. Signon, J.M. Victor, and M. Barbi, Soft Matter **8** (2012) 9285.
- [16] H. Brutzer, N. Luzzietti, D. Kalue, and R. Seidel, Biophys. J. **98** (2010) 1267.
- [17] P. Carrivain, M. Barbi, and J.M. Victor, PLoS Comput. Biol. **10** (2014) e1003456.
- [18] H. Schiessel, J. Phys.: Condens. Matter **15** (2003) R699.
- [19] A. Lesne and J.M. Victor, Eur. Phys. J. E **19** (2006) 279.
- [20] H. Hagège, P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat, and T. Forné, Nat. Protoc. **2** (2007) 1722.
- [21] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, *et al.*, Science **326** (2009) 289.
- [22] C. Braem, B. Recolin, R.C. Rancourt, C. Angiolini, *et al.*, J. Biol. Chem. **283** (2008) 18612.
- [23] F. Court, J. Miro, C. Braem, *et al.*, Genome Biol. **12** (2011) R42.
- [24] F. Court, M. Baniol, H. Hagège, *et al.*, Nucl. Acids Res. **39** (2011) 5893.