Spontaneous emergence of sequence-dependent rosettelike folding of chromatin fiber

Ph. St-Jean, C. Vaillant, B. Audit, and A. Arneodo
Université de Lyon, F–69000 Lyon, France and Laboratoire Joliot-Curie and Laboratoire de Physique, ENS-Lyon, CNRS, 46 Allée d'Italie, 69364 Lyon Cedex 07, France
(Received 20 April 2007; revised manuscript received 22 April 2008; published 27 June 2008)

In the crowded environment of the eukaryotic nucleus, the presence of intrinsic structural defects is shown to predispose chromatin fiber to spontaneously form rosettelike structures. These multilooped patterns self-organize through entropy-driven clustering of sequence-induced fiber defects by depletive forces prior to any external factors coming into play. They provide an attractive description of replication foci that are observed in interphase mammalian nuclei as stable chromatin domains of autonomous DNA replication and gene expression. Experimental perspectives for *in vivo* visualization of rosettelike organization of the chromatin fiber via the clustering of recently identified putative replication initiation zones are discussed.

DOI: 10.1103/PhysRevE.77.061923 PACS number(s): 87.14.G−, 36.20.−r, 87.15.A−, 87.15.Cc

Replication and transcription of DNA require great reproducibility and coordination, all this in the crowded environment of the cell nucleus. Regulation of these complex processes relies in part on the conformation and dynamics of the 30 nm chromatin fiber that ultimately condition DNA sequence accessibility. Chromatin fiber is a nucleoprotein filament with nonhomogeneous structural and mechanical properties [1]. This heterogeneity evidently affects how the fiber folds and organizes into higher-order structures like loops, coils, or chromonema. However, despite increasing experimental and modeling efforts [2], the so-called tertiary chromatin structure is still very controversial and the possible role, if any, of the DNA sequence at such large scales remains an enigma. Recently, a genome-wide wavelet-based multiscale analysis of human DNA sequences [3] has revealed the existence of well defined "replication domains" delimited by 1000 putative replication origins. In these domains, gene position, orientation, and breadth of expression present a high level of organization: around the replication origins, genes are abundant and broadly expressed, and their transcription is co-oriented with the replication fork progression; at replication domain centers, genes are rare and expressed in few tissues. Consistently with the experimental demonstration [4] that domains of open chromatin fiber are enriched in broadly expressed (housekeeping) genes, these replication initiation zones conserved across mammalian genomes are likely to correspond to local open chromatin fiber defects creating an environment that facilitates the initiation and coordination of transcription and replication [3]. Here, our goal is to show that these intrinsic open chromatin defects are likely to predispose the fiber to spontaneously compact itself into a rosettelike structure optimizing the recruitment of protein complexes involved in the activation of transcription and replication.

Macromolecular crowding in the cell nucleus is thought to affect the condensation and decondensation of chromatin and in turn the regulation of transcription and replication [5]. Recently, Snir and Kamien [6] have shown that short molecular chains modeled as stiff (but not rigid) impenetrable tubes, immersed in a solution of small spheres, are driven to a helix configuration by depletion forces. However, their argument is limited to uniform and rather short tubes of length of the order of a few persistence lengths l_p of the rod. For longer chains, the picture rapidly grows in complexity with a plethora of optimal configurations (e.g., hairpin, β-sheet, superhelix, torus) leading to an overwhelmingly rich phase diagram. In this study, we aim to show that the presence of chromatin fiber rosettes can be explained using depletion arguments for long tubes with "frozen" heterogeneously distributed geometric and/or elastic properties. By frozen, we mean that these defects are imprinted on the chromatin fiber by the sequence itself. Indeed, the fiber is known to be dependent upon the properties of the nucleosomal string of beads [7], which in turn is influenced by the double-helix sequence-dependent mechanical properties. As mentioned just above, the recently discovered replication origins [3] provide privileged locations for some intrinsic decondensed fiber defects.

For a semiflexible tube in a dilute environment, local repulsive potentials among parts of the fiber induce a self-avoiding random walk configuration (swollen coil [8]). In a crowded environment, the depletion action may dominate and the fiber will tend to collapse on itself, forming a globular phase. An important feature of the depletion potential lies in its simplistic geometrical nature. We thus consider a system constituted of a dense fluid of hard spheres bathing a semiflexible tube. The tube is assumed to be nonuniform with localized geometrical defects, e.g., local thickening. To make the computation easily tractable, this semiflexible tube can be modeled by a compact helix [9], whereas the geometrical defect would be local decompaction to form "open spots" as illustrated in Fig. 1(a). Note that the helix structure as well as the mechanism that led to it are of no importance for our purpose, the only critical aspect for our demonstration being that the local degree of compaction will vary along the fiber to mimic decondensed structural defects. The elastic nature of the tube (or helix) prevents the appearance of points with too high curvature; consequently the first step in the condensation of the tube is the formation of loops (Fig. 2). Loop formation involves competition between the bending energy of the tube and the entropic gain of the hard sphere fluid. The free energy cost is dominated by elastic energy for small loops and by entropy for large ones. This results in a preferential length of $(3–4)l_p$ in the wormlike chain model with a characteristic loop-formation dynamics

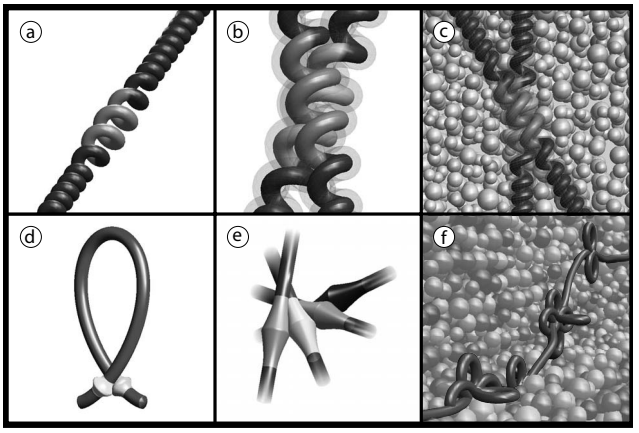

FIG. 1. Different scales in the spontaneous emergence of rosettelike folding of the chromatin fiber in the crowded environment of the cell nucleus: (a) individual "open defect" along a helicoidal fiber; (b) and (c) entanglement of a pair of defects; (d) formation of a loop resulting from the entanglement process; (e) plausible arrangement of a number of defects to form the core of a rosette; (f) chain of rosettes.

($\sim 10^{-2}$ s) which is much faster than the replication time scale (20 mn to a few hours) [10].

Once a loop is formed [Figs. 2(a) and 2(b)], contact will be maintained by depletion forces; hence the loop will preferentially relax through local gliding of the two contact points [Fig. 2(c)], which leaves the depletion potential mainly unchanged. This is where local defects come into play: when they meet from this gliding process, they act as local geometrical wells and "stick" together [Fig. 2(d)]. This defect-induced stabilization is important since it prevents further depletion mechanisms from occurring. Indeed, by modifying locally the angle of tangent vectors at the contact points, the depletion force could drive them to align in opposite directions, forming the first turn of a helix or toroidal condensate; alternatively it could align them in the same direction, favoring the formation of hairpins. The presence of defects, by favoring a specific contact geometry, *breaks* the symmetries (translational, axial) essential to the formation of these compact structures. For example, let us consider the

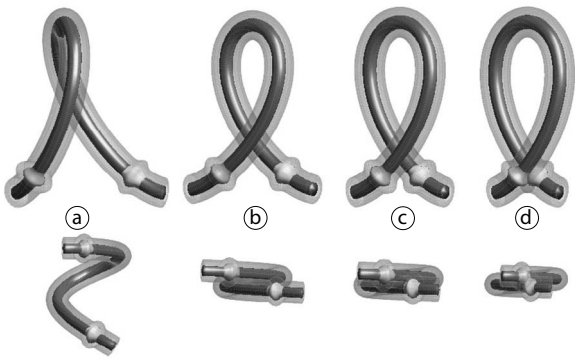

FIG. 2. Steps involved in loop formation in depleting environment: (a) free evolution of the tube; (b) formation of an unstable loop at around $(3–4)l_p$; (c) gliding of the loop; (d) trapping of the loop by local defects. The translucent surface represents the excluded volume for the fluid of hard spheres that is reduced from the overlap resulting from the loop formation.

overlap volume of excluded regions induced by the entanglement of a pair of "opened" regions along the inhomogeneously compact fiber of our pedagogical example [Figs. 1(b) and 1(c)]. For a given geometry of the defects (number of turns, pitch, radius) and putting aside the elastic and entropic contribution involved in loop formation, we can limit the description of the depletion potential in terms of two configuration parameters, namely, the angle between the two strands and the relative axial translation (or shift). For a given configuration (i.e., for given values of these two parameters), there is a minimal distance between helix axes, constrained by contact points. We have computed numerically the overlap of the excluded volume v_{ovl} for two defects involving four complete turns of both helices [Fig. 1(b)]. Figure 3 actually shows v_{ovl}; the depletion potential itself depends linearly on v_{ovl} [Eq. (4)]. What should be noticed is the rather sharp potential well centered around 32°; indeed, we can observe a narrow valley that cover angles ranging from about 28° to 36°. Note also that the bottom of the valley varies in shift with the angle, a trivial consequence of the fact that when the two defects are deeply entangled, they can still rotate with respect to one another if they are allowed to glide

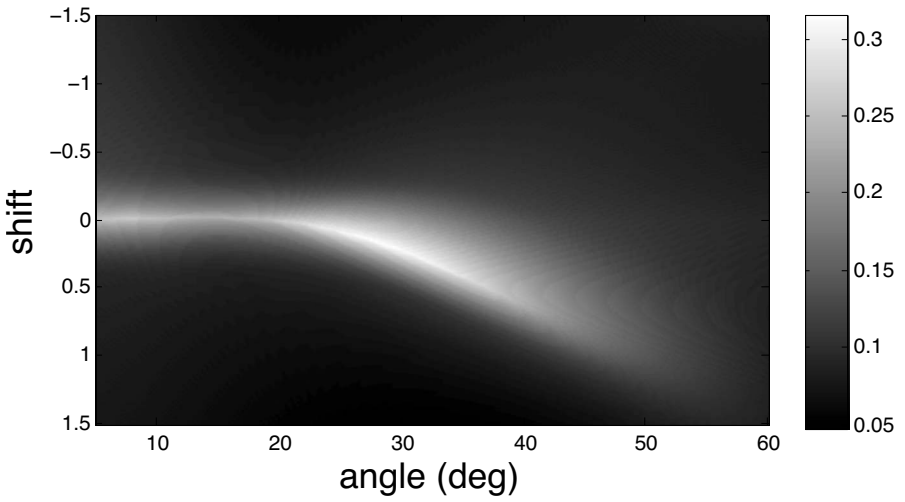

FIG. 3. Depletion potential between two "opened" regions along a helicoidal fiber as a function of the relative angle of the two strands (abscissa) and the relative shift along the axis of one of the two strands (ordinate). The volume overlap v_{ovl} is shown for the following parameter values relative to the radius of the fiber itself ($r_f=1$): radius of both helices $r_h=3$, helix pitches $p_h=9$, and sphere radius $r_s=0.2$.

axially. In any event, the message here is that a simple geometrical defect conformation can lead to the presence of these local depletion locks that may more or less select some characteristic inner loop angle in the rosette pattern self-organization.

As illustrated in Figs. 1(e) and 1(f), the emergence of rosettelike folding of the chromatin fiber in the crowded cell nucleus results from the aggregation of defects. Here, we characterize the distribution of the number of loops per rosette from a minimal parametrization of the system. We consider a dense fluid composed of a large number N_s of identical spheres bathing a tube which, for simplicity, contains N equidistant defects separated by a distance l along the tube. Consistently with the experimental observation that replicons involved in replication foci are likely to be adjacent to each other on the same chromosomal DNA molecule [[11(a)],[11(b)]], we assume that rosettes are formed while respecting the sequential order of defects along the tube. If we cannot rule out the possibility of nonlocal loops, we somehow implicitly hypothesize that the crowded environment, far from fluid, imposes additional mechanical constraints that cause the contribution of far-reaching defect collisions to the overall entropy to increase faster than logarithmically. Let n denote the number of rosettes along the tube; solitary defects are also considered as trivial "rosettes" with zero loops. The case $n=N$ represents the absence of clustering since all defects are then solitary. The case $n=1$ corresponds to a single large rosette assembling all defects. We separate the system into two parts in a natural fashion, namely, the hard sphere fluid and the tube itself. Let F, F_t, and F_s denote the free energies of the system, the tube, and the spheres, respectively (all of which depend on n), such that we can write $F=F_t+F_s$. The most probable value for n is obtained by equating to zero the *chemical potential* μ_r of a rosette:

$$\mu_r \equiv \frac{\partial F}{\partial n} = \frac{\partial F_t}{\partial n} + \frac{\partial F_s}{\partial n} = 0. \tag{1}$$

The equation of state of a fluid of hard spheres has been extensively studied in the past [12]. We follow the method used by Dinsmore *et al.* [13] and make use of the Carnahan-Starling approximation [14]:

$$\frac{P_s(n)V_s(n)}{N_s k_B T} = \frac{1+\varphi+\varphi^2-\varphi^3}{(1-\varphi)^3}, \tag{2}$$

where $\varphi=N_s v_s/V_s(n)$ is the sphere density, $P_s(n)$ the fluid osmotic pressure, v_s the volume of each sphere, and $V_s(n)$ the volume available to the spheres. The F_t term can be expressed as

$$F_t = E_0 + (N-n)\Delta F_l - k_B T \ln\binom{n}{N}, \tag{3}$$

where $\Delta F_l > 0$ is the free energy cost for the formation of a single loop, and E_0 is an energy term assumed to be independent of n. We appreciate that this assumption may seem an oversimplification; however, our goal here is to propose an approximation with minimal complexity that still leads to enough structure such that we can derive a useful result for the chemical potential μ_r. This pragmatic approach is motivated by the fact that the parameters that we just introduced cannot be measured with great precision for the time being as we will discuss below. The last term on the right-hand side of Eq. (3) corresponds to the number of arrangements of n rosettes from N defects and contributes to the entropy of the tube. From Eqs. (2) and (3) and from the thermodynamical identity $\frac{\partial F_s}{\partial V_s} = -P_s$, we get:

$$\mu_r = -k_B T \ln\left(\frac{N-n}{n}\right) - \Delta F_l + k_B T \left(\frac{v_{\mathrm{ovl}}}{v_s}\right)\left(\frac{\varphi+\varphi^2+\varphi^3-\varphi^4}{(1-\varphi)^3}\right), \tag{4}$$

where $v_{\mathrm{ovl}} = -\frac{\partial V_s}{\partial n}$ represents the overlapping excluded volume gained for the spheres due to the interaction between two defects. Thus, this simple model depends on three parameters: the free energy cost of a loop, ΔF_l, the normalized overlap volume per loop, v_{ovl}/v_s, and φ. ΔF_l can be approximated by the sum of an elastic energy term and an entropic term:

$$\Delta F_l = 2k_B T \chi l_p/l + c k_B T \ln(l/l_p) \lesssim 9 k_B T, \tag{5}$$

where $l \simeq (3-4) l_p$ is the length of a typical loop, $\chi=7$, and $c=3-4$ [10]. Physiological values for the hard sphere fluid density $\Phi=\varphi(n=N)$ vary between $\simeq 0.2$ and 0.3 [5]. For a 30 nm fiber we can expect $v_{\mathrm{ovl}} \sim 10^3$ nm^3 (see Fig. 3); the typical size of proteins is $v_s \sim 10^2$ nm^3, leading to values of v_{ovl}/v_s around 10.

As illustrated in Fig. 4, increase in the sphere density Φ or the normalized overlap volume results in an increase of the average number of loops per rosette $(N/n-1)$, and thus in a more compact structure. On the other hand, increase in the

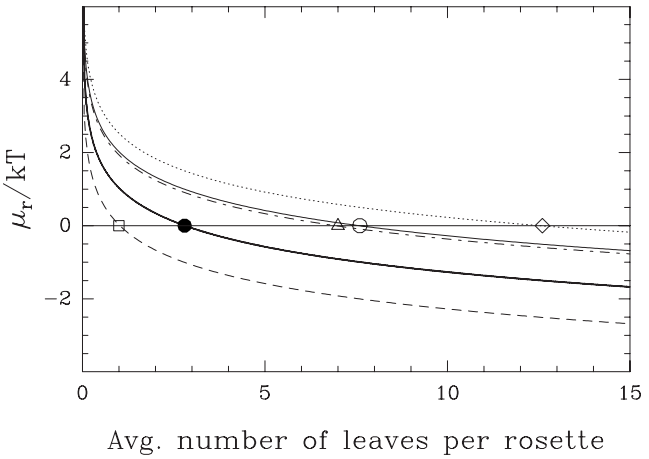

FIG. 4. Chemical potential vs the number of loops per rosette $(N/n-1)$. The thick solid curve serves as a reference and corresponds to the parameter values $\Phi=0.28$, $\Delta F_l=9k_BT$, and $v_{\mathrm{ovl}}/v_s=10$. For these values, a mean value of 2.8 loops (●) per rosette is expected. Increase of Φ to 0.29 (dash-dotted curve) increases the mean to 7 (△); similarly, increase of v_{ovl}/v_s to 11.5 (dotted curve) results in a mean value of 12.6 (◇). Increase of ΔF_l to $10k_BT$ (dashed curve) reduces the mean to 1 (□). Decrease of ΔF_l to $8k_BT$ (thin solid curve) increases the mean to 7.6 (○).

free energy cost of loop formation (stiffening of the fiber) decreases the average number of loops. Interestingly, we find that the average number of loops per rosette can be regulated by fine tuning the values of these parameters within the physiological range. For instance, for $v_{ovl}/v_s=10$ and $\Delta F_l=9k_BT$, varying Φ between 0.28 and 0.29 results in the average number of loops per rosette running from 2.8, i.e., low clustering, to 7. This provides attractive scenarios for the spontaneous emergence of chromatin rosettes in the nucleus prior to their possible stabilization by external factors, e.g., DNA binding proteins [[2(c)],[11,15]].

Various models of interphase chromatin based on a multilooped structure of the 30 nm fiber have been proposed in the literature [[2(b)]], but they all involve interaction with some nucleoproteic complexes to organize the structure, e.g., the scaffolding proteins that interact specifically at certain DNA regions (scaffold associated regions) to fold the fiber [16] or the transcription complexes strung along the genome that form clusters and consequently fold the chromatin fiber [15]. In Refs. [[15(b)],[15(c)]] nonspecific entropic forces between polymerases already engaged on the chromatin fiber have been proposed as possibly driving the compartmentalization of genomes into loops containing several thousands (and sometimes millions) of base pairs. The present work advocates the very attractive alternative that the chromatin fiber can self-organize into rosettelike patterns thanks to its heterogeneous structure. Recent modeling [[7(c)],[7(d)]] has revealed an extreme sensitivity of the internal fiber conformation to the local structural and mechanical properties of the nucleosomal string, e.g., the linker length, the entry-exit angle between the linkers, or the twist angle along a linker. The fiber local structure is known to be controlled by epigenetic modifications of these architectural nucleosomal parameters [[7(b)],[7(d)]] (DNA methylation, histone modifications, etc.). Yet, as suggested by recent modeling of the thermodynamics of DNA loops [17], the local properties of the nucleosomal string are also conditioned by the primary DNA sequence. Therefore the structural defects of the fiber can be encoded in the sequence. The entropy-driven fiber folding mechanism described above leads to the aggregation of neighboring defects into clusters that ensures high local concentration of distant DNA target sites. This clustering is likely to favor the recruiting of protein complexes involved in the activation of replication and transcription. In this context, the set of 1000 putative replication initiation zones recently identified in the human genome [3] are good candidates for some intrinsic decondensed fiber defects. This is further justified by the fact that most of these well-positioned putative replication origins are found to lie close to the promoter regions of divergent genes that are expressed in a large number of tissues and that are likely to exhibit characteristic nucleosome free transcription start sites [18]. Thus, the spontaneous emergence of rosette patterns, likely maintained by the origin recognition complexes, provides a very attractive description of the so-called replication foci [11] that have been observed in interphase mammalian nuclei as stable structural domains of autonomous replication that persist during all cell cycle stages. Furthermore, the remarkable gene organization discovered around the putative replication origins [3] (with clusters of genes mostly co-oriented with the progression of the replication fork) strongly suggests that these rosettes contribute to the compartmentalization of the genome into autonomous domains of gene transcription. Via the self-organizing structural role of the replication origins, the DNA sequence might therefore code, to some extent, for the tertiary chromatin structure. Even though one expects to observe, from one cell cycle to the next, fluctuations in the number of loops contained in each rosette [Fig. 1(f)], the genetic imprinting of defects is likely to ensure the inheritance of the interphase chromatin rosette organization.

To conclude, this study raises opportunities for future experiments. For example, by using different techniques for labeling specific chromosome regions (fluorescent *in situ* hybridization, chromosome painting), we can hope to visualize *in vivo* the clustering of the replication initiation zones into replication foci and the dynamics of the rosette pattern formation, hopefully distinguishing self-organization from DNA-protein binding driven formation, during the different stages of the cell cycle. More precisely, these fluorescent techniques will allow us to analyze the position of replication origins in chromosome territories relative to the nucleus center and their possible relation to nuclear matrix attachment sites.

We thank P. R. Cook, M. Castelnovo, and J. Moukhtar for useful discussions. This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC), the ACI IMPBio, and the Agence Nationale de la Recherche (Grant No. NT05-3_41825).

[1] P. J. Horn and C. L. Peterson, Science **297**, 1824 (2002).

[2] (a) A. S. Belmont *et al.*, Curr. Opin. Cell Biol. **11**, 307 (1999); (b) C. Münkel *et al.*, J. Mol. Biol. **285**, 1053 (1999); (c) T. Cremer and C. Cremer, Nat. Rev. Genet. **2**, 292 (2001); (d) S. M. Gasser, Science **296**, 1412 (2002); (e) W. G. Müller *et al.*, Mol. Cell. Biol. **24**, 9359 (2004); (f) N. Gilbert, S. Gilchrist, and W. A. Bickmore, Int. Rev. Cytol. **242**, 283 (2005).

[3] M. Touchon *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 9836 (2005); E.-B. Brodie of Brodie, S. Nicolay, M. Touchon, B. Audit, Y. d'Aubenton-Carafa, C. Thermes, and A. Arneodo, Phys. Rev. Lett. **94**, 248103 (2005); M. Huvet *et al.*, Genome Res. **17**, 1279 (2007).

[4] N. Gilbert *et al.*, Cell **118**, 555 (2004).

[5] A. P. Minton, J. Biol. Chem. **276**, 10577 (2001).

[6] Y. Snir and R. D. Kamien, Science **307**, 1067 (2005); e-print arXiv:cond-mat/0612243.

[7] (a) C. L. Woodcock *et al.*, Proc. Natl. Acad. Sci. U.S.A. **90**, 9021 (1993); (b) K. van Holde and J. Zlatanova, *ibid.* **93**, 10548 (1996); (c) B. Mergell, R. Everaers, and H. Schiessel, Phys. Rev. E **70**, 011915 (2004); (d) A. Lesne and J.-M. Vic-

tor, Eur. Phys. J. E **19**, 279 (2006).
[8] A. Y. Grossberg and A. R. Khokhlov, in *Statistical Physics of Macromolecules,* edited by R. Larson and P. A. Pincus, AIP Series in Polymers and Complex Materials (AIP, Woodbury, NY, 1994).
[9] C. W. Jones *et al.*, Biophys. J. **88**, 2433 (2005).
[10] H. Yamakawa and W. H. Stockmayer, J. Chem. Phys. **57**, 2843 (1972); S. Jun, J. Bechhoefer, and B.-Y. Ha, Europhys. Lett. **64**, 420 (2003); S. Jun *et al.*, Cell Cycle **3**, 223 (2004).
[11] (a) D. A. Jackson and A. Pombo, J. Cell Biol. **140**, 1285 (1998); (b) R. Berezney, D. D. Dubey, and J. A. Huberman, Chromosoma **108**, 471 (2000); (c) C. Demeret, Y. Vassetzky, and M. Méchali, Oncogene **20**, 3086 (2001).
[12] H. Reiss, H. L. Frisch, and J. L. Lebowitz, J. Chem. Phys. **31**, 369 (1959).
[13] A. D. Dinsmore, A. G. Yodh, and D. J. Pine, Phys. Rev. E **52**, 4045 (1995).
[14] N. F. Carnahan and K. E. Starling, J. Chem. Phys. **51**, 635 (1969).
[15] (a) P. R. Cook, J. Cell. Sci. **108**, 2927 (1995); (b) Nat. Genet. **32**, 347 (2002); (c) D. Marenduzzo, C. Micheletti, and P. R. Cook, Biophys. J. **90**, 3712 (2006); (d) N. M. Toan, D. Marenduzzo, P. R. Cook, and C. Micheletti, Phys. Rev. Lett. **97**, 178302 (2006).
[16] Y. Saitoh and U. K. Laemmli, Cold Spring Harbor Symp. Quant. Biol. **58**, 755 (1993).
[17] C. Vaillant, B. Audit, and A. Arneodo, Phys. Rev. Lett. **95**, 068101 (2005); C. Vaillant *et al.*, Eur. Phys. J. E **19**, 263 (2006).
[18] F. Ozsolak *et al.*, Nat. Biotechnol. **25**, 244 (2007).