

# Mesoscale simulations of two nucleosome-repeat length oligonucleosomes†

Tamar Schlick\* and Ognjen Perišić

Received 8th September 2009, Accepted 2nd October 2009

First published as an Advance Article on the web 20th October 2009

DOI: 10.1039/b918629h

The compaction of chromatin, accessed through coarse-grained modeling and simulation, reveals different folding patterns as a function of the nucleosome repeat length (NRL), the presence of the linker histone, and the ionic strength. Our results indicate that the linker histone has negligible influence on short NRL fibers, whereas for longer NRL fibers it works like, and in tandem with, concentrated positive counterions to condense the chromatin fiber. Longer NRL fibers also exhibit structural heterogeneity, with solenoid-like conformations viable in addition to irregular zigzags. These features of chromatin and associated internucleosomal patterns presented here help interpret structural dependencies of the chromatin fiber on internal and external factors. In particular, we suggest that longer-NRL are more advantageous for packing and achieving various levels of fiber compaction throughout the cell cycle.

## Introduction

### Chromatin structure

The  $30 \times 10^9$  nucleotide base pairs in each human cell are densely packed within the nucleoprotein complex that forms the chromatin fiber. The detailed architecture of the chromatin fiber, not to speak of its dynamics and energetics, have occupied experimental and theoretical scientists for decades. With recent discoveries of a nucleosome positioning code,<sup>1</sup> the need to understand chromatin organization has intensified, because the structural puzzle that explains nucleosome, and hence chromatin organization, underlies the most basic cellular functions, including transcription activation, gene silencing and epigenetic control.<sup>2,3</sup>

### Nucleic acid simulations

While all-atom simulations of nucleic acids have steadily increased in accuracy, scope and length (*e.g.*, microsecond simulations of solvated B-DNA dodecamer<sup>4</sup>), coarse grained models are required to simulate macromolecular systems that are very large and highly dynamic, requiring sampling of millions of configurations to represent behavior over millisecond time frames. Creation of such general multiscale or coarse grained models requires as much art as science, to resolve key functional components of the molecular system while approximating others. Examples of such tailored models include various models of macromolecular assemblies of membrane systems<sup>5–7</sup> and simplified models of the chromatin fiber.<sup>8–11</sup>

The chromatin fiber is made up of the nucleosome core particle building block which organizes 147 bp of DNA in a left-handed supercoil around an octamer of four core histone proteins (H2A, H2B, H3 and H4).<sup>12,13</sup> Nucleosomes are connected by linker DNA, and various linker histones can bind as well as dissociate with the polymer (Fig. 1).

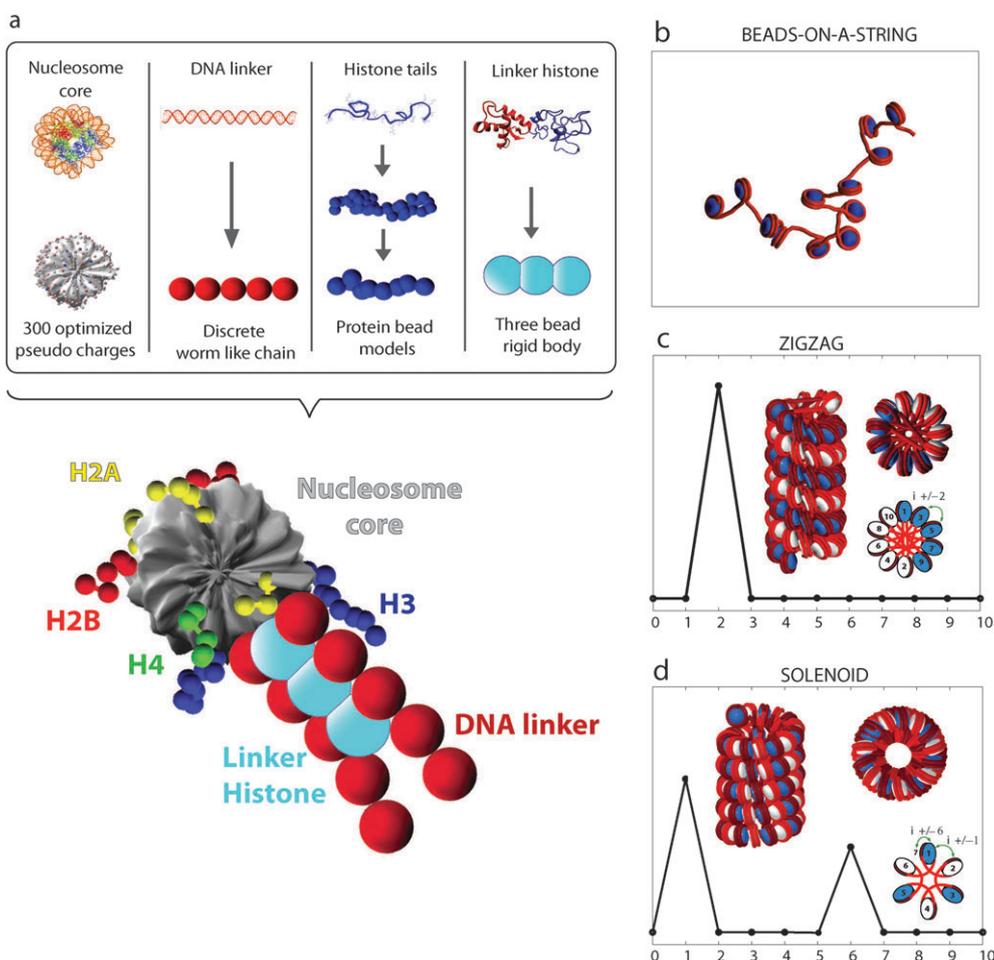
### The chromatin structure puzzle

One of the current problems that chromatin researchers address is elucidation of the 30 nm chromatin fiber architecture. As described in recent reviews (*e.g.*,<sup>14,15</sup>), the detailed structure of the chromatin fiber has remained controversial for over three decades, with evidence supporting both zigzag (“two-start”) and solenoid (“one-start”) models (see Fig. 1) coming from various sources, including X-ray crystallography and electron microscopy. Only recently, have researchers begun to dissect the influence of key internal and external factors such as length of the connecting DNA segments between nucleosomes (which can vary from 10 to 70 bp), the binding of linker histones, and the presence of various concentrations of monovalent and divalent ions on chromatin structure.

For example, a crystallographic structure of the tetranucleosome<sup>16</sup> with short linker DNAs between nucleosomes and without linker histones produced irrefutable evidence for the zigzag model, in which DNA linkers are straight, and alternate nucleosomes ( $i \pm 2$ ) are in closest spatial proximity (Fig. 1). However, electron microscopy of reconstituted fibers with long linker DNAs and with linker histone and divalent ions produced strong evidence for the opposite, solenoid, model in which linker DNAs are strongly bent and nucleosomes crisscross one another, bringing neighboring nucleosomes and neighbors 5 or 6 nucleosomes away in close proximity ( $i \pm 5$ ,  $i \pm 6$ )<sup>17</sup> (see Fig. 1). More recently, Rhodes and colleagues continued to show that, depending on the linker DNA length and presence of linker histone, different fiber dimensions are produced; in particular, short linker DNAs

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012. E-mail: schlick@nyu.edu; Fax: 212-995-4152; Tel: 212-998-3116

† Electronic supplementary information (ESI) available: Fig. S1–5. See DOI: 10.1039/b918629h



**Fig. 1** The mesoscale model and basic chromatin fiber structures. (a) The mesoscale model of the basic chromatin building block. (b) Beads-on-a-string fiber structure at low monovalent salt (0.01 M). (c) An ideal zigzag configuration of the chromatin fiber with the nucleosome contact pattern. (d) An ideal solenoid configuration of the chromatin fiber with the nucleosome contact pattern.

cannot produce compact fibers.<sup>17</sup> Modeling by Wong *et al.*<sup>18</sup> also showed the dependence of fiber width on the nucleosome repeat length (NRL) and the orientation of linker histones. Modeling of simplified coarse-grained nucleosome models by the Rippe group reinforced the large effect of the NRL and nucleosome twist angles on the extent of fiber compaction.<sup>10</sup> More recently, single-molecule force microscope studies<sup>19</sup> subjecting 25-nucleosome arrays with two NRL (167 and 197 bp) to forces up to 4 pN suggested a fundamental solenoid organization, stiffer fibers with short NRL, and only a stabilizing but not structure-determining effect of the linker histone.

### Our mesoscale modeling of chromatin

Over the past several years, our group has developed a mesoscale model of chromatin (Fig. 1) in which the nucleosome, excluding the histone tails, with wrapped DNA is treated as an electrostatically charged object with Debye-Hückel charges approximating the atomistic electric field computed by the Poisson-Boltzmann equation using the DiSCO (discrete surface charge optimization) algorithm.<sup>20,21</sup> Linker DNA is treated as beads in the wormlike chain model used for supercoiled DNA, and linker histones and histone tails are coarse grained by beads from united-atom protein

model. All the components are flexibly attached using parameters derived from experimental measurements. Full model and Monte Carlo simulation details, including model validation and various applications, were recently summarized in Arya and Schlick,<sup>11</sup> where the role of histone tails in compacting fiber structure was analyzed. The Methods section here summarizes key features for easy reference. A recent collaborative study with experimentalists probed nucleosome interaction patterns with and without linker histone, and with and without divalent ions, to dissect how each factor affects chromatin structure.<sup>22</sup> These studies showed that linker histones tighten and strengthen an ordered zigzag fiber structure, and that divalent ions lead to bending of some linker DNAs to produce a zigzag fiber accented with solenoid-like features; the latter organization allows a more compact overall fiber. These studies thus give support for the both solenoid and zigzag models; moreover, they underscore the heterogeneous and polymorphic nature of the chromatin fiber.

### A study of two NRL

In this communication, we report on structural features of the chromatin fiber at two different nucleosome repeat lengths: 173 and 209 bp; the former has NRL close to the value used in

the tetranucleosome (167 bp),<sup>16</sup> while the latter is a typical value of chromatin fibers occurring in Nature. Besides varying the NRL, we also consider the presence and absence of the linker histones and the presence and absence of divalent ions. Divalent ions are modeled as depicted in ref. 22 by a first-order approximation by reducing the DNA persistence length from 50 to 30 nm and by setting the Debye length of the linker DNA beads to the DNA diameter (2.5 nm), to allow beads to touch one another (see also Methods). We sample 12-oligonucleosome fibers at the two NRL values at two salt conditions: typical monovalent salt (0.15 M), and monovalent salt (0.15 M) as well as low concentrated divalent ions. The two NRL values at both salt conditions are modeled with and without linker histone. Configurations are sampled with local and global Monte Carlo moves as described recently.<sup>11</sup> The only change in the model required to account for the different NRL is the different twist between nucleosomes. For our standard model of 6 DNA linker beads (7 segments) or NRL of 209 bp, 6 full turns of DNA occur, but for the short, 173 bp NRL of 2 beads (3 segments, 9 nm, or 26.47 bp), 2.57 DNA turns occur. The difference from an integral number of turns (0.43) is accounted for as twisting energy penalty of  $0.43 \times 360^\circ/26.47 = 6^\circ$  per base pair or a total twist of  $155^\circ$  over the entire 26-bp segment of linker DNA.

For each system (NRL, salt condition, with and without linker histone), we conducted 24 simulations on 12 core arrays with 10 million Monte Carlo steps per trajectory; 12 simulations were started from the zigzag starting structure and 12 from the solenoid. Of these 12 simulations, three different intrinsic DNA twist values were sampled (0,  $\pm 12^\circ$  about the mean), to account for small variations of the DNA twist from one nucleosome to the next, with four different random seeds each. The last 2 million steps per trajectory were included in the statistical analysis. For visualization of compact fibers, analogous simulations of 24-core arrays started from the final configurations of the 12-core systems were also performed. Analysis techniques follow those previously developed;<sup>11,22</sup> see also Methods section here for additional details on computing the nucleosome/nucleosome interaction patterns of core/core interactions, sedimentation coefficients, fiber widths and packing ratios.

The results presented below reflect averages over the ensembles above. (See discussion below and Figs. S1 and S2†). Because of fiber heterogeneity, there are fluctuations in internal variables, especially for the solenoid starting configurations. The overall energies converge quickly (1–3 million steps) and the sedimentation coefficients follow the same trend (Fig. S3†).

## Results

### Fiber packing

Table 1 shows the various sedimentation coefficients, packing ratios (# nucleosomes per 11 nm) and fiber widths for the short and regular NRL values for 12-core arrays. The packing ratio is calculated as the number of nucleosomes divided by the fiber length and multiplied by 11. A linear regression in all three coordinate axes was used to calculate the fiber length for each analyzed simulation frame (see Fig. S5†).

We note a large difference in sedimentation coefficients for the longer NRL. The linker histone has a limited effect on short NRL fibers (173 bp), as seen in the nominal increase of the sedimentation coefficient, packing ratio and fiber width.

The effect of  $Mg^{2+}$  on the packing ratio in short NRL fibers is not large. In contrast, the ion concentration and the presence of the linker histone have a strong compaction effect on the longer NRL (209 bp), as evident by the increase of the average sedimentation coefficient and of the packing ratio. The fiber width for the longer NRL is smaller than the experimentally measured values,<sup>17,23</sup> which reflected highly concentrated divalent ions.

### Fiber architecture

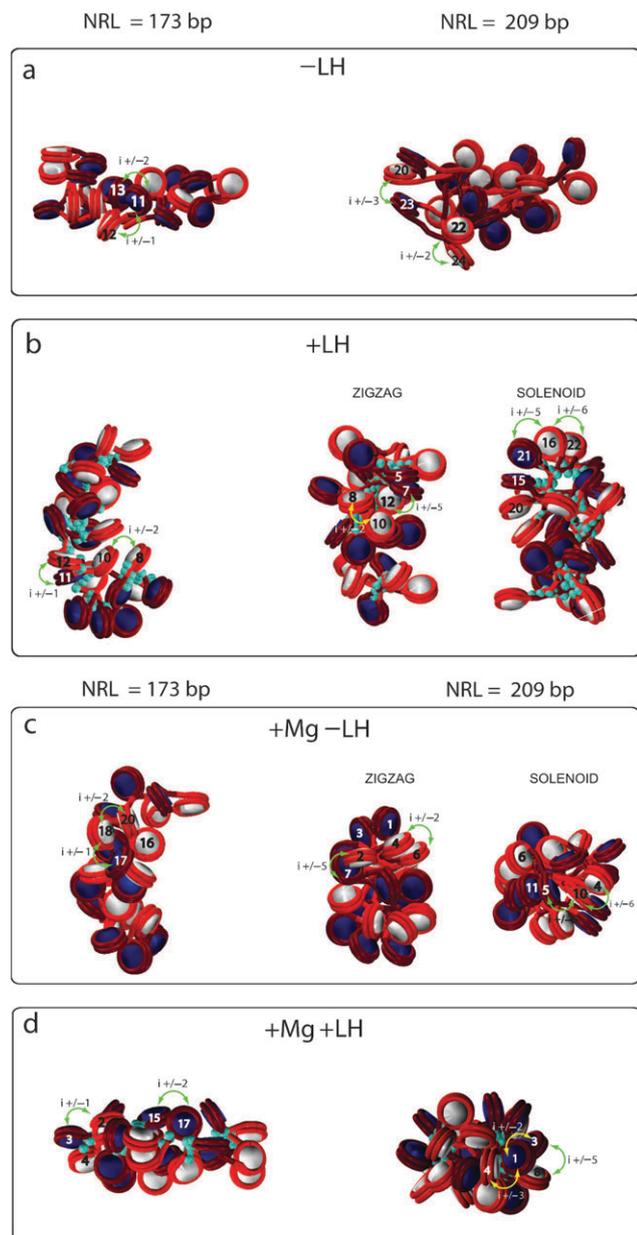
Fig. 2 shows the fiber configurations at various conditions for 24 core arrays. In 6 of the 8 cases, only zigzag configurations are viable, whereas for 2 cases with NRL = 209 bp (–LH + Mg, +LH – Mg), distinct zigzag and solenoid configurations are possible. Those two structures are shown in Fig. 2. Without divalent ions, the short NRL arrays maintain the ladder-like zigzag organization at all conditions, while longer NRL compact into irregular structures. Both arrays exhibit a dependence on the linker histone: sedimentation coefficients and packing ratios increase with the addition of linker histone. For the longer NRL, a tighter zigzag structure is also evident, with long-distance  $i \pm 5$  interactions (Fig. 3), indicating increased compaction and order (analyzed below). The addition of divalent ions produces very compact structures for both NRL systems.

The dominant internucleosomal patterns in Fig. 3a help interpret these trends further. Fig. 3 shows that the short NRL fibers at typical monovalent salt (0.15 M) have strong  $i \pm 1$  interactions simply by construction (and not an indicator of solenoid-like forms), regardless of the presence of the linker histone. However, the addition of linker histone produces more compact fibers with zigzag-type  $i \pm 2$  dominant contacts. The short NRL fibers have negligible  $i \pm 5$  contacts, an indication of their ladder-like structure.

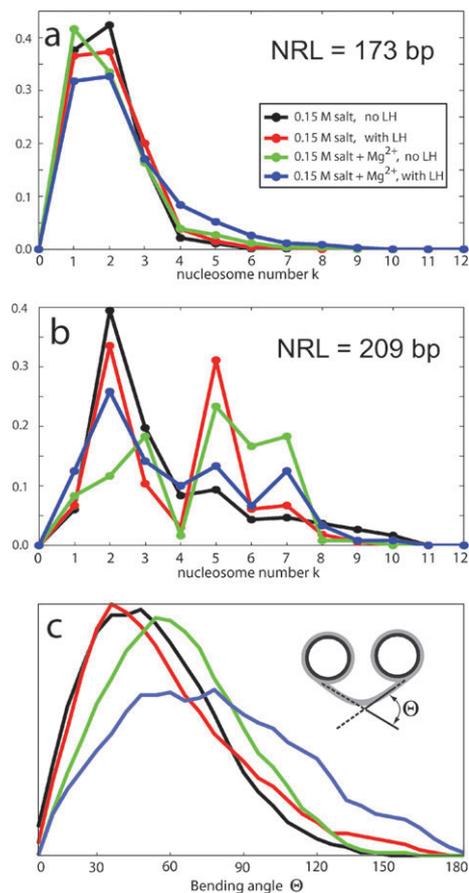
As mentioned above, our results reflect averages from both zigzag and solenoid starting configurations (see supplemental Fig. S1†). As discussed above, the separate analyses show that, in 6 out of 8 cases considered, both starting conformations (zigzag and solenoid) produce only zigzag forms, while two cases preserve both zigzag and solenoid organizations of nucleosomes. Note from Fig. S3† that both ensembles converge fairly quickly (1–3 million steps): zigzag curves fluctuate about converged values thereafter, whereas the solenoid curves continue to fluctuate significantly about the value reached. This behavior reflects the heterogeneity of the chromatin fiber ensemble and the existence of multiple minima rather than lack of convergence. Fig. S3† shows the total potential energy, sedimentation coefficients, the triplet angles measuring the orientation of each three successive nucleosomes, and nucleosome dimer distances. These quantities are representative: the energy behavior mirrors global configurational properties (like sedimentation coefficients and packing ratios) while the triplet angle is more sensitive to the specific state (zigzag or solenoid). Fig. S4† shows behavior of three single trajectories

**Table 1** Fiber properties for 12-core arrays at the two NRL (nucleosome repeat length) values (in base pairs)

	NRL	Sed. coeff./S	Nucl. per 11 nm	Fiber width/nm
-LH, monovalent salt	173	51.13 ± 1.46	3.48 ± 0.37	20.48 ± 5.29
	209	42.15 ± 2.04	3.25 ± 0.76	25.26 ± 10.25
+LH, monovalent salt	173	54.95 ± 1.35	3.99 ± 0.35	19.57 ± 5.02
	209	49.14 ± 2.97	5.74 ± 1.18	26.10 ± 9.16
-LH, monovalent salt & Mg	173	50.10 ± 2.27	3.31 ± 0.42	20.15 ± 5.55
	209	48.56 ± 3.34	6.82 ± 1.54	25.26 ± 10.26
+LH, monovalent salt & Mg	173	50.61 ± 2.10	3.56 ± 0.65	20.97 ± 5.96
	209	48.57 ± 4.21	5.65 ± 2.00	24.88 ± 11.09



**Fig. 2** Fiber structures for two NRLs under different conditions for 24-core arrays. For each case, the dominant zigzag conformation is shown and, for two cases where a solenoid conformation also exist, both the zigzag and solenoid models are shown. The turquoise beads indicate linker histones. Various nucleosome contacts are illustrated to help interpret the profiles in Fig. 3.



**Fig. 3** Contact patterns and bending angles under different experimental conditions. (a, b) The contact patterns for the 173 (Fig. 3a) and 209 bp (Fig. 3b) NRL fibers at 4 conditions as shown in the inset. (c) The bending angle distribution for the 209 bp NRL fibers at 4 conditions as shown in the inset. See Fig. 2 for nucleosome contact illustrations.

for each case. Clearly, each trajectory converges rapidly, but the global ensemble reflects variations in chromatin structure, as expected for a floppy polymer in solution.<sup>18,22</sup> The results clearly indicate that zigzag conformations are always viable and that solenoid conformations are less stable but still viable under some conditions.

Fig. S2† shows that the zigzag architecture is dominant (from the prominent  $i \pm 2$  contacts) at all conditions for NRL = 173 bp, whereas the solenoid configuration is viable when either LH or divalent ions are present for NRL = 209 bp. Note that  $i \pm 1$  contact for short linkers originate from

proximal distances and not from solenoid like architecture. The longer NRL fibers also have prominent higher order interactions, namely  $i \pm 3$ ,  $i \pm 5$  and  $i \pm 7$ .

The DNA bending angle (Fig. 3c), which measures the orientation between the DNA linker exiting one nucleosome and entering another (see Fig. 3c inset), shows that the addition of linker histone to long NRL fibers at typical salt concentrations straightens the DNA linkers (the peak of the bending angle distribution moves from  $40^\circ$  to  $30^\circ$ ) and favors the zigzag structure (Fig. S2†). Divalent ions widen the bending of DNA between nucleosomes because of some bending of linker DNAs, to effectively minimize linker clashes along the fiber axis.<sup>22</sup>

The DNA bending angle distributions in Fig. S2† confirm the above finding and show that linker histones and divalent ions strongly affect DNA linker orientation. For NRL = 209 bp with both divalent ions and linker histones, the bending angle distributions from both starting configurations converge to the same orientation (Fig. S2d†), as found in,<sup>22</sup> indicating a mostly zigzag conformation accented by some linker DNA bending. With LH without divalent ions, or without LH with divalent ions, the configurational ensemble of chromatin for NRL = 209 bp is more heterogeneous, with both zigzag and solenoid architecture feasible. These are the two cases out of 8 considered here where the zigzag is not the only viable form (Fig. 2).

### Tail interactions

Fig. 4 depicts the positional distributions of the histone tails at 0.15 M salt for the two NRLs, with linker histone, along the nucleosome and dyad planes of the nucleosome and within the whole chromatin fiber, projected on the planes perpendicular to the fiber axis. The tail distributions for divalent ions are very similar (not shown). The H3 tails (blue) in longer NRL fibers are positioned in the interior of the fiber along the fiber axis close to the entry/exit positions of the parental DNA linkers with whom they mostly interact. These interactions help maintain linker DNA near the linker histone due to effective screening. Therefore, the DNA linkers in longer NRL fibers, together with the H3 tails and linker histones, reinforce the basic zigzag structural element, the DNA stem formed by the negatively charged linker DNA is screened effectively by the H3 tails and the linker histone.

In the short NRL fibers, the interaction of the H3 tails with the parental DNA linkers is reduced by the presence of adjacent positively charged parental and non parental linker histones and the presence of adjacent nucleosomes. These factors move the H3 tails away from the dyad axis and thus reduce their effectiveness.

The distribution of H4 tails (green) does not depend on the DNA linker length. Their short length and location on the flat surface of the nucleosome move them from the fiber axis. They interact with non-parent nucleosomes<sup>24</sup> to increase fiber compaction.

The H2A and H2B tails are mostly positioned on the periphery of the nucleosome and on the periphery of the fiber. They are involved in histone core aggregation,<sup>25</sup> transcription control, and probably mediate inter-fiber interactions due to their position on the periphery of the nucleosome.<sup>24</sup>

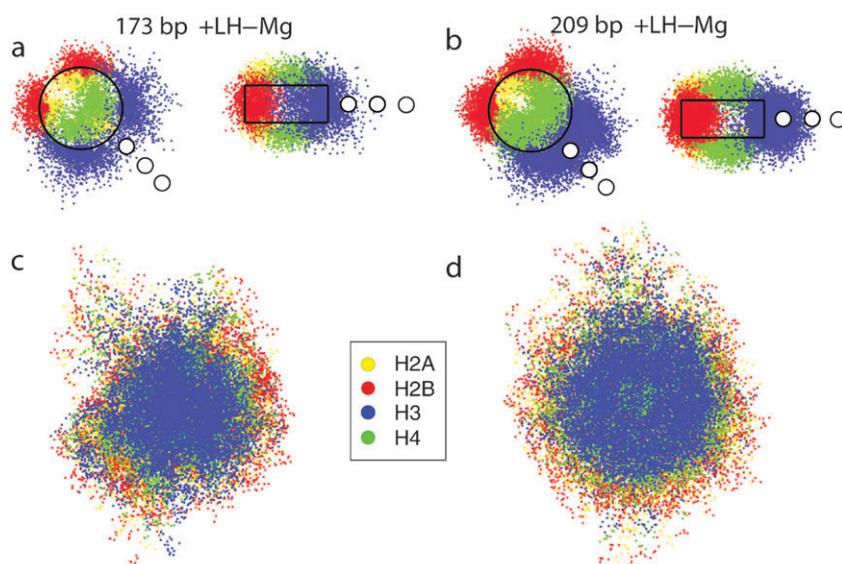
As shown in Fig. 5, where linker DNA distributions are projected on the nucleosome and dyad planes, the linker histone restricts the trajectory of the linker DNA by forming a rigid stem with the longer NRL but affects the shorter NRL much less significantly. Divalent ions have a similar tightening effect but they also introduce heterogeneity in fiber orientation due to some bending, and this is reflected by wider distributions. The wider distributions are also caused by the increased DNA linker length.

### Conclusion

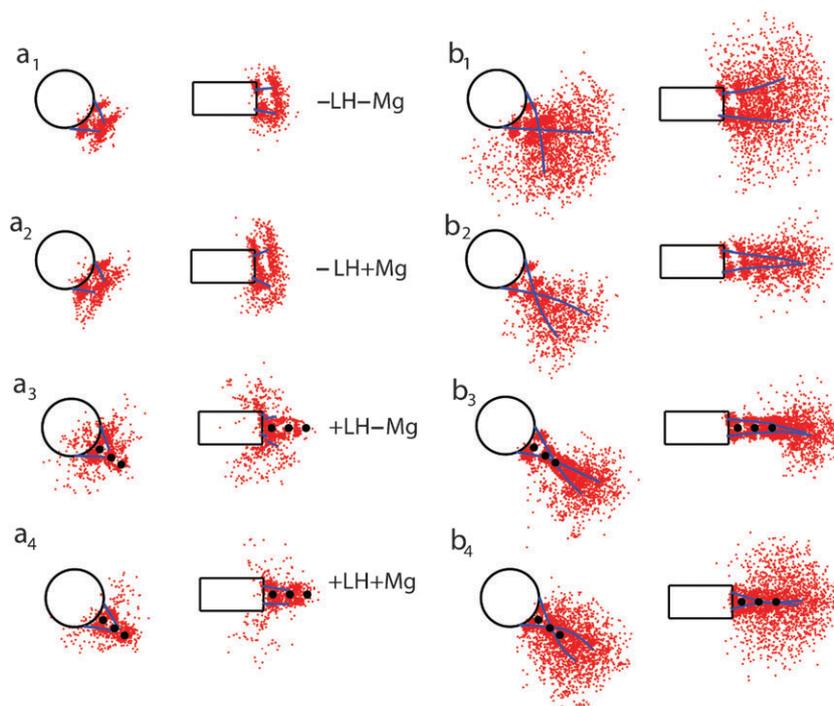
Our studies which examined the configurations of chromatin fiber at NRL = 173 bp and NRL = 209 bp using a mesoscopic model help dissect the effects of NRL, LH and the ionic strength on chromatin fiber packing and geometry. Our results show the negligible effect of LH on fibers with short NRL, which fold into zigzag structures. For longer NRL, both solenoid and zigzag configurations are viable and represent distinct feasible arrangements when either LH or divalent ions are present. However, when both LH and divalent ions exist, the longer-NRL fibers favor one type of hybrid conformation that resembles a zigzag accented by linker DNA bending. The important effect of linker histones and various salt conditions on chromatin organization, especially on long NRL systems, underscore the essential regulatory roles of these factors. The greater configurational heterogeneity in longer NRL systems may favor them over short DNA linker systems for performing fundamental biological processes that require various level of compaction to promote or hamper access to the genetic material.

Recent computational studies using a coarse-grained model described by several tunable parameters such as the linker DNA opening angle and twisting angle between successive nucleosomes<sup>10</sup> found periodic patterns in fiber dimensions for NRL from 202 to 222 bp, a strong effect of NRL on the viable chromatin conformations (two-start and three-start were found), and increased structural irregularity for NRL > 214 bp. While such patterns agree with X-ray scattering studies,<sup>26</sup> Robinson *et al.*<sup>17</sup> recently reported two discrete classes of reconstituted fiber dimensions (radii of 33 nm and 44 nm, respectively) and densities (11 and 15–17 nucleosomes per 11 nm) for NRL < 207 bp and NRL > 217 bp, with a transition between those values. Modeling based on EM measurements of reconstituted fibers, however, show a range of possible conformations as NRL changes;<sup>18</sup> the authors emphasize “the multiplicity of fiber structures!” tuned by the NRL.

Our results here show precisely how NRL, along with LH and positive couterions, shift the ensemble conformations of chromatin fibers at various conditions, making zigzag conformations always viable and also dominant for short NRL. The increased heterogeneity and increased polymorphic nature of the fiber as NRL increases agrees with the above studies<sup>10,18</sup> and adds further details to the associated inter-nucleosome patterns and internal geometry. As argued previously,<sup>10,22</sup> the increased heterogeneity and irregularity may be energetically favorable and biologically important, for example, for transitioning to higher-order forms in



**Fig. 4** Tail distributions for a typical fiber configuration within a nucleosome and within the fiber with linker histone at 0.15 M salt. The dots represent tail bead positions according to the color code shown. The positions of the three linker histone beads are shown as white circles.



**Fig. 5** Linker DNA positional distribution for a typical fiber configuration for various conditions projected on the nucleosome and dyad plane. The red dots represent linker DNA bead positions. The mean trajectories of the DNA linkers are shown as blue lines. The positions of the three linker histone beads are shown as black circles.

chromatin that involve fiber/fiber interactions and higher orders of packing during the cell cycle.

While our studies consider two type of models, other topologies, such as ribbon-like conformations,<sup>27</sup> 3-start helix,<sup>28</sup> and 5-start helix,<sup>29</sup> are also possible. Such a multiplicity of conformations and a large number of tunable internal and external factors (LH, NRL, ionic conditions, variations in NRL from one core to the next) make the chromatin fiber not only infinitely interesting and suitable for performing a rich

array of functions in the cell but also an attractive and continuously challenging subject for experimentalists and theoreticians alike.

## Model and methods

### Chromatin model and analysis techniques

The chromatin model as developed over several years and updated recently to model histone tails, divalent ions and

linker histones was recently described thoroughly in Arya & Schlick (2009).<sup>11</sup> Included in that description were the different treatments of the multiscale chromatin model components (core, tails, linker DNA and linker histone), the mechanical and energetic interactions among them, monovalent and divalent salt modeling, interactions neglected or approximated in the mesoscale model, model validation studies and Monte Carlo sampling methodology. A recent review paper also summarized the history of the developed model, from a macroscopic to mesoscale approach.<sup>30</sup> Essentially, our first-generation model<sup>20,21</sup> captured the fundamental monovalent-salt dependent mechanics of chromatin and the thermal fluctuations of the nucleosome and linker; the next-generation mesoscale model<sup>24,31–33</sup> accounts for the irregular surface of the nucleosome, the flexibility of the histone tails, the presence of linker histone and divalent ion effects (to a first-order approximation). Because all these details are described comprehensively in ref. 11, only bare-bones aspects are described here for quick reference.

**Multiscale chromatin model.** The nucleosome core—the histone octamer and wound DNA minus the N-termini of all four histones and short C-termini of H2A—is treated as a charged, rigid body (Fig. 1) as an irregular charged surface parameterized using the discrete surface charge optimization (DiSCO) algorithm.<sup>21,33</sup> This approach uses optimization<sup>34,35</sup> of a Debye–Hückel approximation to the salt dependent electric field of the atomistic nucleosome core to parameterize 300 pseudo-charges uniformly distributed across the nucleosome surface. An error of less than 10% over a large range of salt concentrations is considered acceptable. In addition to a point charge, each core point is assigned an effective excluded volume modeled using a Lennard-Jones (LJ) potential.

The linker DNA connecting adjacent nucleosome cores is treated using the discrete elastic wormlike chain model of Allison *et al.*<sup>36</sup> with electrostatic potential developed by Stigter.<sup>37</sup> Each monovalent-ion dependent charged bead segment represents 3 nm (~10 bp) of relaxed DNA (Fig. 1) and is also assigned an excluded volume using the LJ potential to prevent overlap with other components of chromatin. Moreover, the DNA beads have stretching, bending and twisting potential energy components following macroscopic supercoiled DNA models.

The ten histone tails—eight N-termini of H2A, H2B, H3 and H4 and two C-termini of H2A—are treated using a coarse-grained protein bead model. Each tail bead represents five protein residues and has a corresponding charge modified by a scale factor close to unity, an excluded volume (LJ) potential, and intramolecular bond stretching and bond-angle bending energy components.<sup>32</sup>

The linker histone (LH) is modeled after the rat H1d linker histone modeled by Bharath *et al.*<sup>38,39</sup> using three charged beads to represent the globular domain (one charged bead) and the C-terminal domain (two charged beads). These beads are placed in a straight line separated by a distance of 2.6 nm (see Fig. 1), and assigned Debye–Hückel charges to reproduce the electric field of the atomistic model as well as excluded volume parameters to account for the interaction of linker

histones with parental cores, all linker DNA and all other linker histones.

The components above are integrated to yield the repeating oligonucleosome motif shown in Fig. 1. The histone tails are attached to the nucleosome core *via* stiff harmonic springs; the rigidly-attached linker histone beads are placed on the dyad axis of each nucleosome at distances  $r = 6.2, 8.8$  and  $11.4$  nm from the nucleosome center as suggested experimentally,<sup>38–40</sup> and each core is linked to an entering and exiting linker DNA segments where the linker DNA entering/exiting orientation has a corresponding value of  $\theta_0 = 108^\circ$  and wrapping of 1.7 turns of wound DNA according to the recent high-resolution nucleosome crystal with tails resolved.<sup>41</sup>

Each linker DNA bead and nucleosome is allowed to twist about the DNA axis with a twisting energy penalty implemented through local Euler coordinates for each bead. In addition, to measure DNA bending and twisting at core attachment points, two other coordinate systems describe the trajectory of the wrapped DNA on the nucleosome cores at the point where it diverges from the core to connect the two linker DNA. Additional coordinate frames are defined to transform each linker DNA bead to the next and each nucleosome core to/from preceding/succeeding DNA beads.

**Governing energetics.** The total potential energy is expressed as the sum of stretching, bending and torsional components of linker DNA, stretching of histone tails, intramolecular bending to the histone tails, total electrostatic energy (accounting for all interactions among nucleosomes, linkers, tails and cores with all other units) and excluded volume terms.

**Ion modeling.** Physiological salt conditions of monovalent and divalent cations are important for compacting chromatin by electrostatic repulsion between the linker DNAs. The solvent surrounding oligonucleosomes is treated as a continuum with a dielectric constant of 80. The effect of monovalent ions is treated using effective Debye–Hückel charges that are monovalent-salt dependent. Divalent ions are treated to a first-order approximation following studies on DNA bending,<sup>42,43</sup> which suggest a reduction of the DNA persistence length to promote linker bending. Specifically, we reduce the repulsion among linker DNA in linker/linker interactions by setting an inverse Debye length of  $\kappa = 2.5 \text{ nm}^{-1}$  to allow DNA to almost touch one another, and reduce the persistence length of the linker DNAs from 50 to 30 nm according to experimental findings.<sup>42,43</sup>

**Model validation and approximations.** When properly developed and parameterized, a mesoscale model allows sampling of oligonucleosome systems for investigating key features systematically that depend on basic electrostatics and conformational flexibility. By construction and out of necessity, all-atom detail (*e.g.*, sequence dependence) is not considered, and ion correlation effects are omitted. Behavior thus represents an average over sequence, protein binding, atomistic fluctuations, higher-order ionic correlations, explicit solvation interactions, *etc.*

Our mesoscale chromatin model has been validated for many experimentally-measured properties (see ref. 11, 24, 31 and 32). These properties include salt-induced compaction of

oligonucleosomes to reproduce experimental sedimentation coefficients<sup>44</sup> and nucleosome packing ratios;<sup>45–47</sup> diffusion and salt-dependent behavior of mononucleosomes, dinucleosomes and trinucleosomes;<sup>48–50</sup> salt-dependent extension of histone tails measured *via* the tail-to-tail diameter of the core and radius of gyration for mononucleosomes over a broad range of monovalent salt concentrations;<sup>51</sup> the irregular zigzag topology of chromatin fibers consistent with experimental models<sup>16,27,47,50</sup> and its enhanced regularity upon LH binding;<sup>16</sup> linker crossing orientations in agreement with various experiments;<sup>47,52–54</sup> and internucleosome interaction patterns consistent with cross-linking and EM experiments.<sup>22</sup>

**Monte Carlo simulations.** We use four tailored MC moves—pivot, translation, rotation and tail re-growth—with relative frequencies of 0.2 : 0.1 : 0.1 : 0.6, respectively.<sup>24,32</sup> The tail re-growth method which we use for the histone tail beads involves a configurational bias MC<sup>55–57</sup> to re-grow a randomly selected histone tail bead by bead by using the Rosenbluth sampling criterion<sup>58</sup> starting with the core-attached bead.

We employ two starting oligonucleosome configurations (zigzag and solenoid models) and employ both 12 and 24-core nucleosome simulations, the latter only for suggesting compact configurations of fibers as obtained from the extensive 12-array studies.

**Calculation of internucleosome patterns.** Nucleosome/nucleosome interaction patterns are calculated by an interaction intensity matrix that measures the fraction of time (or fraction of MC configurations) each nucleosome's histone tail interacts with the other nucleosome, that is approach each other within 80% of their van der Waals radii. Such matrices<sup>22,24</sup> can be further decomposed to one-dimensional plots that depict the magnitude of each  $i, i + k$  interaction, as shown in Fig. 3.

**Calculation of sedimentation coefficients.** We applied the method developed by Bloomfield *et al.*<sup>59,60</sup> to calculate the sedimentation coefficient of a given oligonucleosome array conformation, from the intercore distances.<sup>44,61</sup> The sedimentation coefficient  $S_{20,w}$  is approximated from  $S_N$ , where

$$\frac{S_N}{S_1} = 1 + \frac{R_1}{N} \sum_i \sum_j \frac{1}{R_{ij}} \quad (1)$$

Here,  $S_N$  represents  $S_{20,w}$  for a rigid structure consisting of  $N$  nucleosomes of radius  $R_1$ ,  $R_{ij}$  is the distance between the centers of two nucleosomes, and  $S_1$  is  $S_{20,w}$  for a mononucleosome. This approach assumes spherical nucleosomes, a reasonable approximation. We use  $R_1 = 5.5$  nm and  $S_1 = 11.1$  Svedberg (1 S =  $10^{-13}$  s) as done previously.<sup>61</sup> Similar results can be obtained by a more complex procedure implemented in the program HYDRO<sup>62</sup> which calculates  $S_{20,w}$  using the radii of both the nucleosome core particle (5.0 nm) and the DNA bead (1.5 nm).

**Calculation of fiber packing ratio.** To calculate the fiber packing ratio (number of nucleosomes per 11 nm of fiber length) for each simulation frame, we first compute the length of the fiber axis passing through a chromatin fiber core

(see Fig. S5†). The fiber axis is a parametric curve defined by three functions  $f(i)$ ,  $g(i)$  and  $h(i)$  (where  $i$  is a nucleosome core index) which are calculated by a method of least squares, using the nucleosome core center vectors as input parameters for each simulation frame. The least squares method produces  $n + 1$  polynomial coefficients for a polynomial of an order  $n$ . We use  $n = 2$ , as it gives the best estimate of the fiber axis. The described procedure produces 12 points per spatial dimension for a 12-oligonucleosome fiber and 24 points per spatial dimension for a 24-oligonucleosome fiber and a 3D curve for the fiber axis. We sum the Cartesian distances between every consecutive odd point ( $f(i), g(i), h(i)$  and  $f(i + 2), g(i + 2), h(i + 2), i = 1, 3, 5, \dots$ ) to obtain the fiber axis length  $L$ . We use odd points only to avoid overestimating the fiber length. The packing ratio (# of cores per 11 nm) is calculated as the number of cores multiplied by  $11/L$ .

**Calculation of fiber width.** From the fiber axis computed above we define the fiber radius, for a given nucleosome core, to be the perpendicular distance between a nucleosome core center and its closest linear fiber axis segment plus the nucleosome radius (5.5 nm). We repeat this procedure for each configuration and average the value to obtain a mean fiber radius. The diameter is twice that value.

## Acknowledgements

This work was supported by NSF grant MCB-0316771 and NIH grant R01 GM55164 to T. Schlick. is also made to the donors of the American Chemical Society (Award PRF39225-AC4) Petroleum Research Fund and Philip Morris USA and to Philip Morris International.

## References

- 1 E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J. P. Wang and J. Widom, *Nature*, 2006, **442**, 772–778.
- 2 T. J. Richmond, *Nature*, 2006, **442**, 750–752.
- 3 G. Felsenfeld and M. Groudine, *Nature*, 2003, **421**, 448–453.
- 4 A. Pérez, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2007, **129**, 14739–14745.
- 5 A. Arkhipov, Y. Yin and K. Schulten, *Biophys. J.*, 2008, **95**, 2806–2821.
- 6 P. J. Bond and M. S. Sansom, *J. Am. Chem. Soc.*, 2006, **128**, 2697–2704.
- 7 K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid and M. S. Sansom, *Structure*, 2008, **16**, 621–630.
- 8 K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. Langowski and J. C. Smith, *J. Comput. Chem.*, 2008, **29**(9), 1429–1439.
- 9 F. Mühlbacher, H. Schiessel and C. Holm, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 031919.
- 10 R. Stehr, N. Kepper, K. Rippe and G. Wedemann, *Biophys. J.*, 2008, **95**, 3677–3691.
- 11 G. Arya and T. Schlick, *J. Phys. Chem. A*, 2009, **113**(16), 4045–4059.
- 12 J. T. Finch, L. C. Lutter, D. Rhodes, R. S. Brown, B. Rushton, M. Levitt and A. Klug, *Nature*, 1977, **269**, 29–36.
- 13 K. Luger, A. W. Mäder, R. K. Richmond, D. R. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251–260.
- 14 K. van Holde and J. Zlatanova, *Semin. Cell Dev. Biol.*, 2007, **18**, 651–658.
- 15 D. J. Tremethick, *Cell*, 2007, **128**, 651–654.
- 16 T. Schalch, S. Duda, D. F. Sargent and T. J. Richmond, *Nature*, 2005, **436**, 138–141.

- 17 P. J. J. Robinson, L. Fairall, V. A. T. Huynh and D. Rhodes, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 6506–6511.
- 18 H. Wong, J.-M. Victor and J. Mozziconacci, *PLoS One*, 2007, **2**, e877.
- 19 M. Kruithof, F.-T. Chien, A. Routh, C. Logie, D. Rhodes and J. van Noort, *Nat. Struct. Mol. Biol.*, 2009, **16**, 534–540.
- 20 D. A. Beard and T. Schlick, *Structure*, 2001, **9**, 105–114.
- 21 D. A. Beard and T. Schlick, *Biopolymers*, 2001, **58**, 106–115.
- 22 S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock and T. Schlick, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13317–13322.
- 23 A. Routh, S. Sandin and D. Rhodes, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8872–8877.
- 24 G. Arya and T. Schlick, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16236–16241.
- 25 A. Bertin, D. Durand, M. Renouard, F. Livolant and S. Mangenot, *Eur. Biophys. J.*, 2007, **36**, 1083–1094.
- 26 S. P. Williams, B. D. Athey, L. J. Muglia, R. S. Schappe, A. H. Gough and J. P. Langmore, *Biophys. J.*, 1986, **49**, 233–248.
- 27 S. H. Leuba, G. Yang, C. Robert, B. Samori, K. van Holde, J. Zlatanova and C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 11621–11625.
- 28 V. L. Makarov, S. I. Dimitrov, V. Smirnov and I. G. Pashev, *FEBS Lett.*, 1985, **181**, 357–361.
- 29 J. R. Daban and A. Bermúdez, *Biochemistry*, 1998, **37**, 4299–4304.
- 30 T. Schlick, in *Bridging The Scales in Science in Engineering*, ed. J. Fish, Oxford University Press, New York NY, 2009, pp. 514–535.
- 31 J. Sun, Q. Zhang and T. Schlick, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 8180–8185.
- 32 G. Arya, Q. Zhang and T. Schlick, *Biophys. J.*, 2006, **91**, 133–150.
- 33 Q. Zhang, D. A. Beard and T. Schlick, *J. Comput. Chem.*, 2003, **24**, 2063–2074.
- 34 T. Schlick and A. Fogelson, *Assoc. Comput. Mach., Trans. Math. Software*, 1992, **18**, 46–70.
- 35 T. Schlick and A. Fogelson, *Assoc. Comput. Mach., Trans. Math. Software*, 1992, **18**, 71–111.
- 36 S. A. Allison, R. Austin and M. Hogan, *J. Chem. Phys.*, 1989, **90**, 3843–3854.
- 37 D. Stigter, *Biopolymers*, 1977, **16**, 1435–1448.
- 38 M. M. S. Bharath, N. R. Chandra and M. R. S. Rao, *Nucleic Acids Res.*, 2003, **31**, 4264–4274.
- 39 M. M. S. Bharath, N. R. Chandra and M. R. S. Rao, *Proteins: Struct., Funct., Genet.*, 2002, **49**, 71–81.
- 40 S. Sheng, D. M. Czjkwosky and Z. Shao, *Biophys. J.*, 2006, **91**, L35–L37.
- 41 C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond, *J. Mol. Biol.*, 2002, **319**, 1097–1113.
- 42 C. G. Baumann, S. B. Smith, V. A. Bloomfield and C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 6185–6190.
- 43 I. Rouzina and V. A. Bloomfield, *Biophys. J.*, 1998, **74**, 3152–3164.
- 44 J. C. Hansen, J. Ausio, V. H. Stanik and K. E. van Holde, *Biochemistry*, 1989, **28**, 9129–9136.
- 45 S. P. Williams and J. P. Langmore, *Biophys. J.*, 1991, **59**, 606–618.
- 46 S. E. Gerchman and V. Ramakrishnan, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 7802–7806.
- 47 J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Kostner and C. L. Woodcock, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 14173–14178.
- 48 J. Yao, P. T. Lowary and J. Widom, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 7603–7607.
- 49 J. Yao, P. T. Lowary and J. Widom, *Biochemistry*, 1991, **30**, 8408–8414.
- 50 J. Bednar, R. A. Horowitz, J. Dubochet and C. L. Woodcock, *J. Cell Biol.*, 1995, **131**, 1365–1376.
- 51 A. Bertin, A. Leforestier, D. Durand and F. Livolant, *Biochemistry*, 2004, **43**, 4773–4780.
- 52 K. Toth, N. Brun and J. Langowski, *Biochemistry*, 2006, **45**, 1591–1598.
- 53 K. van Holde and J. Zlatanova, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 10548–10555.
- 54 J. F. Kepert, K. F. Tóth, M. Caudron, N. Mucke, J. Langowski and K. Rippe, *Biophys. J.*, 2003, **85**, 4012–4022.
- 55 J. I. Siepmann and D. Frenkel, *Mol. Phys.*, 1992, **75**, 59–70.
- 56 J. J. de Pablo, M. Laso and U. W. Suter, *J. Chem. Phys.*, 1992, **96**, 2395–2403.
- 57 D. Frenkel, G. C. A. M. Mooij and B. Smit, *J. Phys.: Condens. Matter*, 1992, **4**, 3053–3076.
- 58 M. N. Rosenbluth and A. W. Rosenbluth, *J. Chem. Phys.*, 1955, **23**, 356–359.
- 59 V. Bloomfield, W. O. Dalton and K. E. van Holde, *Biopolymers*, 1967, **5**, 135–148.
- 60 J. G. Kirkwood, *J. Polym. Sci.*, 1954, **12**, 1–14.
- 61 M. Garcia-Ramirez, F. Dong and J. Ausio, *J. Biol. Chem.*, 1992, **267**, 19587–19595.
- 62 J. García de la Torre, S. Navarro, M. C. Lopez Martinez, F. G. Diazand and J. J. Lopez Cascales, *Biophys. J.*, 1994, **67**, 530–531.