CHAPTER 9

Monte Carlo Simulations of Nucleosome Chains to Identify Factors that Control DNA Compaction and Access

KARSTEN RIPPE,*[a] RENE STEHR[b] AND GERO WEDEMANN*[b]

[a] Deutsches Krebsforschungszentrum (DKFZ) & BioQuant, Research Group Genome Organization & Function, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; [b] University of Applied Sciences Stralsund, CC Bioinformatics, Zur Schwedenschanze 15, 18435 Stralsund, Germany
*Email: Karsten.Rippe@dkfz.de; Gero.Wedemann@fh-stralsund.de

9.1 Introduction

The nucleosome is the basic organizing unit of the genome in higher eukaryotes.[1,2] It comprises two copies each of the core histone proteins H2A, H2B, H3, and H4 in an octameric protein core around which 145–147 base pairs (bp) of DNA are wrapped in almost two left-handed turns. The core histone proteins have a globular part that comprises three well-structured α-helices and long protruding N-terminal and H2A C-terminal tails. The tails extend from the globular histone fold and lack a specific secondary structure. They can carry multiple post-translational modifications like acetylation, methylation and phosphorylation that represent important signals for controlling genome

RSC Biomolecular Sciences No. 24
Innovations in Biomolecular Modeling and Simulations: Volume 2
Edited by Tamar Schlick

Published by the Royal Society of Chemistry, www.rsc.org

function.[3–5] The four core histones (and their variant forms) are stably bound to the DNA and have average residence times on the hour time scale.[6] This feature renders nucleosomal DNA partly inaccessible to other protein factors and makes nucleosome positioning a regulatory element for DNA binding.[7,8] In addition, access to the linker DNA between nucleosomes can be modulated by folding the nucleosome chain into higher order structures. This effect has been demonstrated experimentally by comparing linker DNA binding of proteins to dinucleosomes with binding to a folded chain of 17 nucleosomes, which showed a 50-fold difference in its accessibility.[9]

A central parameter of the nucleosome chain conformation is its degree of compaction. The fully decondensed state is represented by an extended nucleosome chain of $\sim$10 nm diameter. At physiological ionic strength chromatin fragments of 10 to 100 nucleosomes can reversibly fold into fiber structures that are referred to as 30 nm fibers according to their approximate diameter.[1,10–17] For this state the reported experimental determinations of its very basic parameters like diameter or linear mass density show striking differences, indicating that the 30 nm chromatin fiber can adopt different conformational states. Measurements of the fiber diameter yielded values that varied between 20 and 45 nm for chicken erythrocyte chromatin and for fibers reconstituted *in vitro*.[18,19] Likewise, the linear fiber mass density given as the number of nucleosomes per 11 nm fiber contour length varies from about 1–2 nucleosomes per 11 nm chain length in yeast[20] up to 10 or even 17 nucleosomes per 11 nm fiber for fully compacted chains.[19] Thus, despite the vast amount of available data it remains a challenging task to decipher folding properties of the nucleosome chain in dependence of the factors that govern its conformation for a given set of conditions. As discussed in further detail below, crucial parameters in this context are the spacing of nucleosomes referred to as the nucleosome repeat length (NRL),[21–23] the presence and type of linker histones,[1,10,22] the ionic conditions,[24–28] and post-translational histone modifications.[29,30] How these parameters cause conformational transitions and regulate the accessibility of the linker DNA is an important question of research in the field. It can be addressed by developing quantitative descriptions for modeling the nucleosome chain and predicting its conformation. This task is further complicated by findings from a number of recent studies arguing that the 30 nm fiber does not persist within the cell nucleus in an environment that is highly enriched with nucleosomes,[31–35] which is in contrast to other observations of fiber structures in the nucleus.[1,14–17,36] Thus, we face the additional challenge to reconcile a potential preference for the decondensed 10 nm chain conformation state within the cell with the vast amount of evidence in favor of the formation of 30 nm type fiber structures with native and reconstituted chromatin fragments *in vitro*.[1,10–13,37–40]

In summary, a number of pertinent questions on the folding of the nucleosome chain remain to be elucidated. Since the nucleosome as the fundamental building block of chromatin is known at atomic resolution and experimentally well-characterized one approach is to develop coarse-grained descriptions for the nucleosome and connecting linker DNA. Coarse graining can be conducted on different length scales as depicted in Figure 9.1.

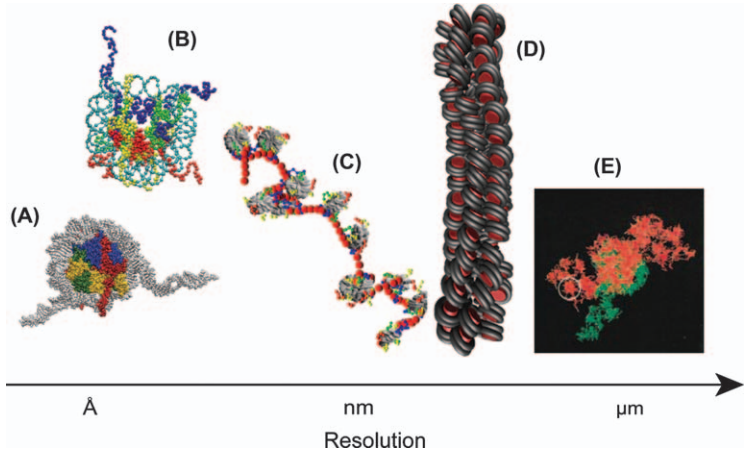

Figure 9.1 Chromatin models at different degrees of coarse graining. (A) Molecular dynamics simulations of a single nucleosome at atomic resolution, as conducted for example for studying unwrapping of nucleosomal DNA.[41] (B) Molecular dynamics simulation of a nucleosome with beads that represent protein residues and DNA nucleotides.[42] (C) Monte Carlo simulation of a chain with 12 nucleosomes.[43] (D) Monte Carlo simulation of a chromatin fiber with 100 nucleosomes.[44] (E) Monte Carlo simulation of human chromosome 3.[45]

Here, we will focus on the mesoscale regime of models that describe conformation and other features adopted by chains containing up to 100 nucleosomes. A number of analytical approaches have been developed for this regime as described for example in refs. 46–51. However, these are limited to providing a static picture of the geometrically possible conformations of nucleosome chains without evaluating their stability or dynamics. This shortcoming can be addressed by applying numerical methods to conduct computer simulations for a given nucleosome chain. With these techniques the stability of chain conformations and the effects from energy and entropy and solvent-macromolecule interactions that determine their dynamics can be evaluated. The numerical methods comprise three main approaches: In molecular dynamics (MD) and Brownian dynamics (BD) simulations the time-dependent structural fluctuations are investigated. In Monte Carlo (MC) simulations a representative ensemble of configurations at equilibrium is obtained.

The MD method is a deterministic approach, in which Newton's equations of motion are solved numerically in small time steps to compute conformational dynamics.[52] Its main application to chromatin lies currently in simulations of single nucleosome dynamics[41,42,53–55] as well as histone tail-induced chain folding as studied with coarse-grained nucleosome arrays.[56–58] When the time scale or size of the system of interest exceeds the available computational resources an alternative approach is to conduct BD simulations.[52,59] In these, the solvent is modeled as a continuum that exerts frictional and random stochastic forces on the particles. Thus, specific non-homogenous features of the

solvent as, for example, local ion-macromolecule interactions are not considered. From the MD and BD trajectories, the pathway of the conformational rearrangements and their kinetics can be obtained, but they are expensive in terms of computation time. This limits the size and complexity of the simulated systems that can be studied.

If only information on the ensemble conformation and thermodynamic properties of a system at equilibrium are needed, Metropolis Monte Carlo (MC) approaches are sufficient and in general much faster.[60–62] In these, a representative ensemble of configurations at thermal equilibrium is computed as described in further detail below. The MD and BD dynamics simulation methods are complementary to MC in nature and lead to the same averages of statistic quantities, given that the system under consideration is ergodic and the same statistical ensemble is sampled.

In this review, we will focus on MC simulations since they are ideally suited to investigate equilibrium features of the nucleosome chain in coarse-grained models with comparatively moderate requirements of computing time. MC simulations have been used successfully to investigate chain conformation, mass density and stiffness,[44,63–68] the contributions from nucleosome-nucleosome and histone-DNA interactions,[21,69–71] the salt-dependent compaction of the chain,[72,73] the role of histone tails,[43,74,75] and the effect of linker histone binding.[23,76] Here, we will discuss the implementation and application of MC simulations for studying the properties of the nucleosome chain.

9.2 Experimentally Determined Features of the Nucleosome Chain

Numerous experimental studies have characterized features of the nucleosome chain. Since any theoretical description of nucleosome chain folding needs to be critically evaluated against these experimental data it is important to understand how they are derived and why some of them appear to be contradictory. Accordingly, we will briefly discuss in the following the experimental variables that need to be considered.

9.2.1 Experimental Systems

Native chromatin fiber fragments are typically isolated from cells by a partial digestion with micrococcal nuclease (MNase). The length of nucleosome chains isolated in this manner can be adjusted to enrich a certain fragment size but is in general below 100 nucleosomes. The resulting chromatin fragments are heterogeneous with respect to the DNA sequence (although approaches exist to isolate fragments with defined DNA sequence),[25,77,78] post-translational histone modifications and the presence of non-histone proteins. Frequently, chromatin was isolated from chicken erythrocytes since purification of relatively large amounts is straightforward.[24,25,79–85] Other studies characterized native chromatin fragments from rat liver,[38,85–87] bovine thymus[88] or brain

tissue,[87] sea urchin sperm,[89,90] yeast[91] as well as immortalized mammalian cell lines like HeLa.[92,93]

To study nucleosome chains of a defined composition nucleosome chains are reconstituted.[10,11,94–97] Typically, a gradient of decreasing salt concentration is applied to deposit histone octamers and linker histones onto the DNA.[98] The *in vivo* chromatin assembly process involves histone chaperones and chromatin remodelers, and results in a more regular spacing of nucleosomes around a certain NRL.[99] In contrast, the salt gradient reconstitution method leads to a large variation of the distance between nucleosomes with natural DNA sequences. Nucleosomes can be reconstituted at defined positions by using arrays of high affinity binding sites for the histone octamer like the 5S DNA repeat from sea urchin[100] or the "601" sequence determined from an *in vitro* selection of random DNA sequences[101,102] to obtain equal distances between nucleosomes. This technique results in a highly regular spacing for the *in vitro* assembly that exceeds that of native chromatin.[103,104]

Thus, large variations between the experimental systems exist that need to be considered when modeling the folding of the nucleosome chain into fibers or other higher order structures. For example, chicken erythrocyte chromatin with an NRL of 212 bp appears to represent a more repressive overall conformational state and is enriched with 30 nm chromatin fibers[16] with a mass density of about 6–7 nucleosomes per 11 nm fiber.[24,79] It contains the avian-specific linker histone type H5 instead of H1. Nucleosomes from chicken erythrocytes display a characteristic stem-like structure in which H5 mediates the association of the two DNA segments leaving the nucleosome core particle over a distance of 3–5 nm before the linker DNA diverges.[79,81] It is unclear if this type or DNA organization by linker histones is also present in chromatin from other sources. In contrast, yeast chromatin has unusually short nucleosome repeat length between 154 and 165 bp and no canonical linker histones but the functional homologue Hho1p.[105] It adopts a more decondensed conformation of the nucleosome chain with a low mass density of 1.2–2.4 nucleosomes per 11 nm chain.[20,91]

9.2.2 Nucleosome Repeat Length

The NRL varies between species and cell type from 154 and 237 bp corresponding to a DNA linker of 10–100 bp between two nucleosomes. For example, NRL values of around 154 bp (~7 bp linker) in *S. pombe*, 165 bp (~18 bp linker) in *S. cerevisiae*, 175 bp (~28 bp linker) in *D. melanogaster* and *C. elegans*, and 185 bp (~38 bp linker) in *H. sapiens*,[1,8,106] 212 bp for chicken erythrocytes[1] and 237 bp in sea urchin sperm[90] have been determined. The distribution of spacer lengths is not random but follows a ~10-bp periodicity,[107] which closely resembles a helical turn of DNA (10.4 bp). Thus, certain sterical requirements of nucleosome spacing appear to exist that need to be accommodated in the higher-order folding of the chain. Furthermore, linker histone knockout studies demonstrate a linear relationship between the ratio of H1 per nucleosome and the NRL, with a lengthening of 37 bp being induced by

the linker histone.[12] In mammals, the typical NRL is around 200 bp but can show large variations between tissues.[1] The NRL can be determined from a partial MNase digestion and subsequent analysis of the DNA length distribution by gel electrophoresis with an accuracy of 1–2 bp.[1] These experiments suggest that the region of regular nucleosome spacing that can be identified as a set of distinct bands comprises less than ~10 nucleosomes, which is consistent with the results from genome-wide mapping of nucleosome positions.[8,106,108]

9.2.3 Variations in Protein Composition of Nucleosome Chains

9.2.3.1 Histone Modifications

Histones are subject to numerous post-translational modifications like acetylation, methylation and phosphorylation, particularly at the unstructured N-terminal tails. These modifications are set or removed in a dynamic manner by specific enzymes that have been associated with various functions.[3,4] Histone modifications can serve as binding sites for protein domains that specifically interact with the post-translationally modified histone state. For example, chromo- and bromodomains recognize methylated or acetylated histones.[5] A number of experimental studies indicate that histone tails are important for mediating internucleosomal interactions and the folding of the nucleosome chain. Removal of the histone tails leads to some increase of nucleosome flexibility and affects nucleosome-nucleosome interactions as well as the binding of other proteins to the nucleosome and/or its associated DNA.[30,109–119] In particular, acetylation of histone lysines can have a direct effect on the stability of the nucleosome core particle, and on its higher-order interactions, since the positively charged lysine is neutralized in the acetylated state.[29,30,120–124] Likewise, also for methylation of histones[125] or DNA[126] a direct effect on the nucleosome stability and interactions has been inferred from *in vitro* studies.

The N-terminal tails of H2B and H3 mediate internucleosomal interaction, possibly by binding in the continuing groove of the DNA superhelix formed by two stacked nucleosomes.[116,117,127–130] The interaction of the H4 tail with the acidic patch on the surface of H2A is particularly important for interactions between nucleosomes. Accordingly, clipping off the H4 tail or its acetylation at lysine residue 16 can strongly reduce the compaction of nucleosome arrays.[29,30,115,116,123]

9.2.3.2 Histone Variants

The canonical core histones H2A, H2B and H3 can be replaced by variant histones with different amino acid sequence. These substitutions can modulate the folding of the nucleosome chain.[131–133] In particular, variants of histone H2A appear to significantly change nucleosome-nucleosome interactions via the H4 tail.[116,127,132,134] The canonical H2A core histone provides an acidic patch that interacts with the positively charged H4 tail in the nucleosome crystal structure.[127,134] In contrast, the H2A variant H2A.Bbd lacks three

acidic amino acids in this region, and its incorporation into the nucleosome chain inhibits folding.[132] On the other hand, the H2A.Z variant has an extended acidic patch, which appears to favor nucleosome-nucleosome interactions and chain compaction.[133]

9.2.3.3 Linker Histones and Other Chromosomal Proteins

The linker histone H1 is present in 5 isoforms (H1.1 to H1.5). In avian erythrocytes its H5 variant is present while in yeast the Hho1p homologue is found. Linker histones interact with the nucleosome and an additional ~20 base pairs of flanking DNA to form a complex that is referred to as a chromatosome.[105,135–137] H1/H5 have a tripartite protein domain structure consisting of a compact globular domain flanked by two highly positively charged N- and C-terminal domains.[138–142] The latter two are mostly unstructured in the free protein and neutralize negative charges of the DNA phosphate backbone.[143,144] The effects of the linker histones on nucleosome chain folding are complex and have been reviewed previously.[12,38,140,141,145] Since a high-resolution structure of the chromatosome is missing, our current view of the chromatosome is based on model structures for the interaction of the linker histone and the nucleosome and flanking DNA.[139,146–151] Based on these structures and a large body of experimental studies the effects of linker histones on chromatin conformation appear to originate from three major contributions: (i) changes to the entry/exit angle of the DNA geometry at the nucleosome[146–150,152–154] (ii) neutralization of negative DNA phosphates by positively charged linker histone residues,[140,146,155] and (iii) an increase of the NRL with linker histone stoichiometry.[12]

9.2.3.4 Other Chromosomal Proteins

In addition to the core, variant and linker histones, a number of non-histone proteins are found in native chromatin and affect its structure. These architectural chromosomal proteins can both compact or open up chromatin:[95,156] (i) Heterochromatin protein 1 (HP1) is involved in establishing and maintaining the repressive state of pericentric heterochromatin.[157–160] Its chromatin binding properties depend on the methylation status of histone H3 at lysine 9 since the N-terminal chromodomain of HP1 interacts preferably with H3 histone tails that carry a K9me2/3 modification.[161–163] (ii) Proteins with the high mobility group (HMG) motif bind chromatin and are classified into the HMGA,[164] HMGB[165] and HMGN[166] groups. They can counteract linker histone-mediated chromatin compaction (HMGN5), affect post-translational histone modifications (HMGN and HMGB) and regulate nucleosome positioning (HMGB1, HMGN1 and HMGN2).[167,168] (iii) MeCP2 binds nucleosomes and methylated CpG sites and compacts the nucleosome chain.[169–171] (iv) The CTCF transcription factor has chromatin organizing activities, presumably by promoting the formation of DNA loops.[172,173] (v) By electron

microscopy, it was shown that core components of the polycomb repressive complex 1 induce compaction of nucleosomal arrays.[174] (vi) A number of recent reports demonstrate that cohesin and condensin protein complexes not only compact DNA in the mitotic chromosome but also act as organizers of the higher order interphase chromatin structure.[175,176]

9.3 Protein-protein and Protein-DNA Interactions in the Nucleosome Chain

9.3.1 Nucleosome-nucleosome Interactions

Nucleosome chain folding is driven by nucleosome-nucleosome interactions that compensate for the unfavorable energetic terms for linker DNA bending/twisting and its electrostatic repulsion as well as the decrease in conformational entropy.[21,67,72,177] At a concentration above 50 mM salt, *i.e.* at the physiologically relevant ionic strength, this interaction becomes attractive as observed by a variety of methods.[1,178–180] Experimentally, the salt-depended folding of the chain has been studied with both reconstituted nucleosome arrays,[11,22,26] and native chromatin fragments.[27,83,86] The strength of internucleosomal interactions can be quantitated by force spectroscopy experiments.[181–183] In these a nucleosome chain is bound at one end to a solid support and is then extended by pulling at the other end with forces in the range of 0.1 pN to 40 pN. From the resulting extension a force-distance curve is obtained. These experiments were conducted at physiological ionic strength for native chromatin fibers from chicken erythrocytes[182] as well as reconstituted nucleosome arrays.[181,183–186] The resulting values for nucleosome-nucleosome interaction energies cover a rather broad range from 3.4 to 14 k_BT with k_B being the Boltzmann constant and T the temperature for breaking this interaction or 2–8 kcal mol^{-1}. These values are similar to those of -4 to -7 kcal mol^{-1} found for typical unspecific protein-DNA interactions.[187]

9.3.2 DNA Interactions with the Histone Octamer Protein Core

Histone-DNA interaction maps with a $\sim$10-bp periodicity were derived from the crystal structure of the nucleosome[127,128] and from stretching experiments.[188–190] A total of 14 main interaction sites of protein and DNA at regions where the minor groove faces inwards were identified. Each of these sites can be considered as comprising two contacts separated by $\sim$5 bp between each of the two individual DNA strands and the histone octamer as inferred from recent molecular dynamics studies and force spectroscopy experiments.[41,55,188] The breaking of these contact sites at the DNA entry-exit site occurs spontaneously and leads to partial DNA unwrapping. This process was studied both experimentally and theoretically.[55,191–196] Lifetimes of several seconds for the fully wrapped state were reported that are interrupted by periods of a few tenths of seconds during which up to 80 bp of nucleosomal DNA are unwrapped from the histone octamer protein core.

Force spectroscopy experiments of nucleosomal DNA unwrapping were conducted to evaluate the interaction strength of DNA and histone protein core.[55,184,188,190,197–200] For a single nucleosome, the initiation of DNA unwrapping was observed already at ~3 pN.[190,200] In contrast, no unwrapping of nucleosomal DNA was observed below extension forces of ~5 pN for chains with 25 nucleosomes.[71,181,183] This points to a stabilization of the nucleosome structure in the chromatin fiber. However, also for nucleosomes within a fiber DNA unwrapping becomes significant at forces above ~5 pN.[201,202] In these experiments, the outer turn (67 bp, 23 nm of DNA) dissociates first and more easily than the inner DNA turn (80 bp, 27 nm).[55,184,185,188,190,199,203] From the force spectroscopy experiments and competitive protein binding experiments values of 10 k_BT, [190] 15 k_BT [191] and 20 k_BT [184] were derived for unwrapping the DNA of the outer turn. For subsequent unwrapping of the inner DNA turn an energy barrier appears to exist[71,184,188,190,204] that could be due to higher affinity DNA histone interactions flanking the dyad axis.[188,190]

9.4 Chromatin Fiber Conformations

Numerous models for the fiber geometry have been proposed. These are based on *in vitro* studies with native chromatin fragments and reconstituted nucleosome chains as well as on studies of chromatin in cells.[1,14–17,19,37–40,134,205–208] The different fiber types can be classified according to their nucleosome stacking as proposed previously.[47] In this nomenclature the conformation is described by two parameters [N_{stack}, N_{step}]: These refer to the number of nucleosome stacks N_{stack} and the step size between connected nucleosome stacks N_{step}.

An example is shown for a seven-start helix in Figure 9.2. Additional parameters that characterize the fiber conformations are the orientation nucleosomes to the helix axis (tilt angle), the position of the linker histone and the degree of linker DNA bending.

9.4.1 Solenoid Fibers

The classical solenoid fiber model has a one-start [1,1] helical organization of the chain, in which consecutive nucleosomes stack on top of each other[37,38,40,210] (Figure 9.3A). The interactions between nucleosomes adjacent on the DNA require bending of the intervening linker DNA. This is energetically unfavorable and could be facilitated by association with linker histones.[18,27] Other solenoid models were proposed to allow a higher nucleosome density than 6–7 nucleosomes per 11 nm fiber as reported for the [1,1] conformation.[10,19,206] These are characterized by an interdigitation of nucleosomes between adjacent turns of the helix, but differ in the nucleosome tilt angle with respect to the chromatin fiber axis.[67] In the fiber conformations proposed by Daban nucleosomes have high tilt angles of 40–60° forming [n,1] fibers with $n = 3$–6.[206]

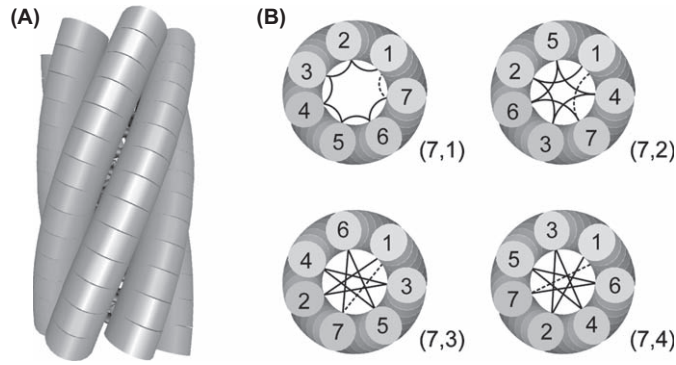

Figure 9.2 Classification of chromatin fiber structures by number of nucleosome stacks and DNA linker path. (A) Side view of a chromatin fiber in a seven-start helix conformation, *i.e.* the chain folds into seven nucleosome stacks. (B) Top view of the seven-start helix from panel A for four different paths of the linker DNA that is indicated by black lines. The nomenclature to describe the four different types of fibers is that proposed by Depken and Schiessel. The first number gives the nucleosome stacks and the second the number of steps along the nucleosome stacks to reach the nucleosome that is adjacent on the chain.[47] The [7,1] conformation would correspond to a solenoid fiber type, while [7,2], [7,3] and [7,4] have a crossed-linker DNA path. The dashed line shows the linker DNA connection to the nucleosome of the next turn. The image has been adapted from reference.[44]

Examples for a [6,1] conformation of this type are shown in Figure 9.3D,E for two different NRLs. In the model from the Rhodes group the tilt angle is $\sim 20^{\circ}$ and the nucleosome stacking follows a zig-zag path, which cannot be described in terms of nucleosome stacks[10,19] (Figure 9.3B). The high compaction ratios of interdigitated fibers were experimentally observed in the electron microscopy study that identified two distinct structural classes of fibers.[19] For NRLs of 187–207 bp a diameter of 33–34 nm and a nucleosome packing ratio of $\sim$11 nucleosomes per 11 nm fiber was measured. Longer repeat lengths of 217–237 bp associated into thicker fibers with a diameter of $\sim$44 nm and a linear mass density of $\sim$15 nucleosomes per 11 nm fiber.

9.4.2 Crossed-linker Fibers

In crossed-linker DNA chromatin fibers nucleosomes interact with each other that are not adjacent on the nucleosome chain. This allows for straight linker DNA with crossings in the interior of the fiber along a zig-zag path as in the [7,3] and [7,4] conformations shown in Figure 9.2B and the two-start fibers with straight linker DNA in Figure 9.3C and F. In these structures the fiber diameter would be expected to depend linearly on the length of the linker DNA.[211] However, results on this issue are contradictory. An increase of diameter with NRL was observed in two studies,[211,212] while others reported no

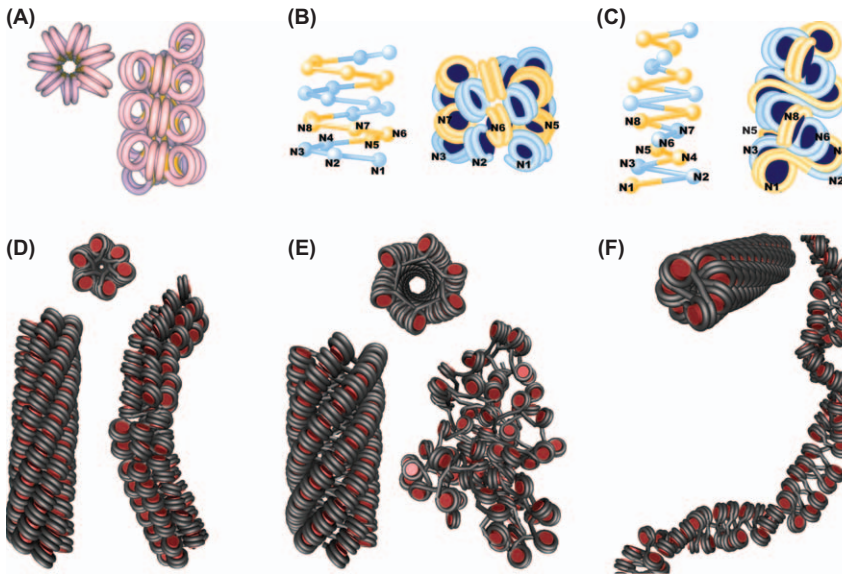

Figure 9.3 Geometric models and MC simulations of different chromatin fiber model conformations. (A) Classical solenoid model[37] with a one-start [1,1] helical organization. The nucleosomal DNA is colored in light violet cyan and the linker DNA in yellow (image from reference[134]). (B) Interdigitated solenoid model with low nucleosome tilt angle according to the model from Robinson *et al.*[19] (image from reference[35]). Alternating nucleosome pairs are colored blue and yellow, and the nucleosomes at positions 1–7 of the chain are indicated. (C) Two-start helix crossed-linker DNA fiber in [2,1] geometry derived by extending the tetranucleosome crystal structure[209] (image from reference[35]). The color coding is the same as in panel B. (D) MC simulations of a [6,1] fiber conformation with relatively high nucleosome tilt angles and different NRLs that are based on the conformation proposed by Daban.[206] The left structure shows the initial configurations and the right fiber is a representative conformation in thermal equilibrium obtained after MC simulations.[67] NRL = 189 bp, linear mass density 7.6 nucleosomes/11 nm fiber, diameter 33 nm. (E) Same as in panel D but for NRL = 207 bp. The initial fiber structure transformed into a random aggregate at thermal equilibrium. This aggregation is driven by the increased electrostatic repulsion of the longer linker DNA. (F) MC simulations of two-start helix crossed-linker chromatin fibers with NRL = 169[67] that mimics the conformation shown in panel C is shown. The start structure (left) and a representative configuration at thermal equilibrium are shown. The simulations yielded linear mass densities of 3.1 nucleosomes/11 nm fiber.

change[16,40] or an increase of the fiber diameter by ~10 nm only between an NRL of 207 bp and 217 bp.[19]

An experimentally well-established folding state is the crossed-linker DNA two-start chromatin fiber conformation with stacking of nucleosomes i and $i+2$ and adjacent nucleosomes connected by more or less straight linker DNA, *i.e.* a [2,1] geometry.[134,209,212,213] For this type of fiber the nucleosome

orientations and path of the linker DNA can be derived from the crystal structure of a tetranucleosome at 167 bp NRL and in the absence of linker histones.[209] The structure has a resolution of 9 Å, and was solved by molecular replacement with the high-resolution nucleosome core structure determined previously.[128] The tetranucleosome structure can be extended into a continuous fiber,[209] the stability of which was investigated by Monte Carlo simulations.[67] The resulting conformation at thermal equilibrium is in good agreement with structures observed by electron microscopy for NRLs of 167 bp and 197 bp.[22]

9.5 Coarse-grained Representations of the Nucleosome Chain

Molecular dynamics simulations with atomic resolution have been used to explore the properties of individual nucleosomes.[41,54,214,215] However, even the simulation of only two nucleosomes in all-atom models is at the limit of what is technically feasible at present. Accordingly, different levels of coarse graining have been introduced to simulate nucleosome chains. In one approach, 5–20 atoms were integrated into "super atoms" to investigate dynamic features of a single nucleosome.[42] In so-called mesoscale models for polynucleosomes, bigger units like spheres or cylinders are used to describe several base pairs of DNA or the nucleosomes. The actual degree of coarse graining depends on the properties to be studied (Figure 9.1). Here, we focus on MC simulation studies of nucleosome chains with 12 to 100 nucleosomes. These provided valuable insight into a number of chromatin features.

To our knowledge, the first model that went beyond the purely static geometric fiber models and accounted for the dynamic properties of the nucleosome chain was that from Ehrlich *et al.* in 1997.[216] The authors investigated hydrodynamic parameters like the diffusion coefficient of the nucleosome chain by computing Brownian dynamics trajectories for chains with 2, 4 and 25 nucleosomes according to the two-angle conformation model proposed by Woodcock *et al.*[205] A few years later, Monte Carlo simulations with similar coarse-grained models were first applied to model force spectroscopy experiments[69] to evaluate the contribution of nucleosome electrostatics in a combined BD and MC study[63] and to investigate the fiber conformation in comparison with experimental data sets.[64] Further developments and applications of MC simulations of nucleosome chains in subsequent studies can be loosely classified according to their major goals of (i) investigating the conformation of the chromatin fiber and its characteristic parameters like diameter, linear mass density and persistence length,[44,65–68] (ii) dissecting the energetics of nucleosome-nucleosome and histone-DNA interactions[21,70,71,217] and the salt-dependent chain compaction,[57,72,73] and (iii) determining the contribution of histone tails[43,74,75] or linker histone binding[23,67,76,217] to the conformation.

9.5.1 Coarse Graining

In the various mesoscale MC studies of nucleosome chains different descriptions of coarse-grained protein and DNA components were used. By accounting for the energy potentials that describe the interaction between these units representative ensembles of configurations in thermal equilibrium can be sampled with MC or BD protocols. For modeling a whole chromosome or even complete genomes the degree of coarse graining is further increased, and segments that represent parts of a chromatin fiber are used as building blocks in the model[45,218] (Figure 9.1).

9.5.1.1 Basic Units

The nucleosome chain models are composed of elements that represent its protein and DNA parts. The discretization of DNA into segments comprising several base pairs is well established from studies with coarse-grained models of DNA alone.[219–221] A nucleosome is either described by a sphere,[69,216] an ellipsoid,[64,65] a spherocylinder[21,67,70,71] or in finer granularity by an ensemble of smaller spheres.[42,57,73–75,222] Modeling a nucleosome by a single sphere or ellipsoid appears to result in unrealistic chain structures.[21] To describe the nucleosome properties in more detail it can be represented by a group of smaller beads,[42] for example, to consider the linker histone and core histone tails explicitly in the model.[23,43,74] Applying this approach increases the number of interactions that have to be computed between individual units, and up to now the largest chains described in this manner consist of 48 nucleosomes.[23] For studying systems in the range of 100 nucleosomes and above more coarse-grained descriptions of a nucleosome as a single unit have been applied.[21,44,64–68,70,71] For these models, the practical limit in terms of computation time is currently at about 1000 nucleosomes or 200 kb of DNA as discussed below.

9.5.1.2 Fiber Geometry

Units representing DNA, nucleosomes or nucleosome elements are referred to as segments. The position of a segment i in the chain can be described by a position vector $\vec{p}_i$, and a local coordinate system $(\vec{f}_i, \vec{u}_i, \vec{v}_i)$ with $\vec{v} = \vec{u}_i \times \vec{f}_i$. The segment vector $\vec{s}_i$ is defined by $\vec{s}_i = \vec{p}_{i+1} - \vec{p}_i$ with the segment length $b_i = |\vec{s}_i|$ and $\vec{s}_i = \vec{u}_i \cdot b_i$.[219] To include the position and orientation of nucleosomes additional geometric parameters are required. In the simplest model the overall structure of the fiber is represented by two angles, α and β (Figure 9.4A).[205] The parameter, α is the angle between incoming and outgoing DNA at the nucleosome seen at the flat side of the nucleosome, while β is the torsion angle between two adjacent nucleosomes. However, additional degrees of freedom exist. In our own work the geometrical center of the nucleosome is determined by a distance from the line connecting the attached DNA-segments and additional angles that describe the orientation of the relative position vector and the orientation

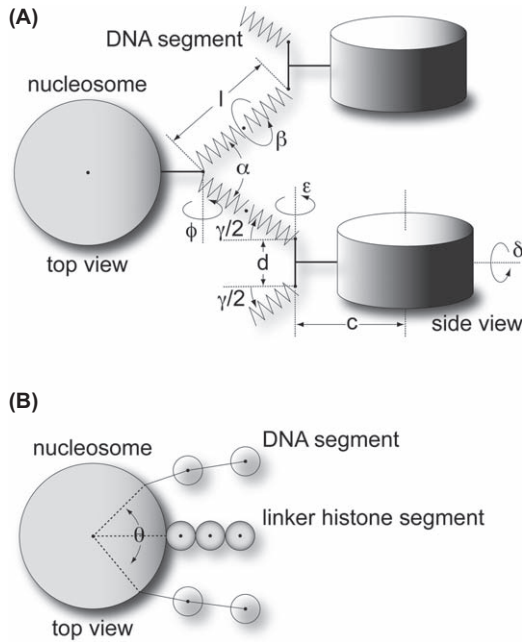

Figure 9.4 Two coarse-grained models used in MC simulations for the discretization of the nucleosome chain. (A) Nucleosomes are modeled as single units connected by DNA segments, with the indicated six angles describing the orientation of the nucleosome relative to the DNA.[44,67] (B) Model in which the nucleosome is represented by a group of spheres.[43] The position of the DNA and the tails is described relative to the center of the nucleosome. The linker histone and the tails (only one tail is shown exemplarily) are modeled explicitly.

of the nucleosome relative to the segment (Figure 9.4A).[44,67] A similar approach was used by Schlick and coworkers (Figure 9.4B).[43] The angles to represent a certain chain geometry can be derived from comparison with experimental data that contain information about the linker DNA topology, as for example the tetranucleosome crystal structure.[67,209]

9.5.2 Interactions

The local elastic properties of the chain such as stretching, bending and torsion are modeled by energy potentials between neighboring units.[43,44] In addition, non-adjacent segments may interact with each other: DNA experiences an electrostatic repulsion due to its negatively charged phosphate backbone, and for nucleosomes excluded volume effects and their interaction potential with other nucleosomes and DNA need to be considered. The computation time for the elastic energy contributions scales linearly with N as the number of segments. For the non-local energies the computation time scales with $O(N^2)$, which can be reduced to $O(N)$ by using cell structures.[223]

9.5.2.1 Elastic Energies

The elastic interactions described are generally assumed to be harmonic, which has been justified for DNA.[221] Accordingly, the DNA stretching energy is computed from the difference to the DNA equilibrium length and the DNA stretching module. For calculating bending and torsion energies the three Euler angles α_i, β_i, γ_i for the transformation of the local coordinate system of the bead i to the next bead $i+1$ are determined. In this representation α_i and γ_i are the rotation angles around the segment vector, and β_i is the rotation around the orientation vector f_i. Bending is computed from β_i, and torsion is calculated from the sum of α_i and γ_i.[220] For some segments, *e.g.* at the entry of DNA in the nucleosome, the equilibrium position of one segment to the next is intrinsically bent in the relaxed state. The equilibrium direction $\vec{B}_i$ can be described by two angles ω_i and ξ_i.

$$\vec{B}_i = \vec{f}_i \sin\omega_i \cos\xi_i + \vec{v}_i \sin\omega_i \sin\xi_i + \vec{u}_i \cos\omega_i \tag{9.1}$$

The degree of bending is the difference between $\vec{B}_i$ and $\vec{B}_{i+1}$. If the path of the DNA around the histone core is not modeled explicitly, the twist introduced by the geometric setup and the histone-DNA interactions between the two neighbor segments must be considered. This is accounted for by an intrinsic twist contribution, which is subtracted from the computed twist. The parameterization of the different elastic energy terms for the DNA can be derived from single molecule experiments.[221,224]

9.5.2.2 Electrostatic Energy of the DNA

The electrostatic repulsion between DNA segments is due to the negative charges of the phosphate backbone that are partly shielded by mobile ions of the solvent. This potential can be described by the Poisson-Boltzmann-equation or the Debye-Hückel-approximation. If segments are short, DNA-charges can be represented by charged spheres with sufficient accuracy.[44] For longer segments, DNA can be described by line charges, and the electrostatic interaction is obtained by integrating the solution of the Debye-Hückel equation for a point charge over two charged line segments.[64,220,221]

$$E_{ij}^{(e)} = \frac{v^2}{D} \int d\lambda_i \int d\lambda_j \frac{\exp(-\kappa r_{ij})}{r_{ij}} \tag{9.2}$$

In eq. (9.2), D is the dielectric constant of water, v the linear charge density, κ the inverse of the Debye length, and r_{ij} is the distance between the current positions at the segments i and j with the corresponding integration parameters λ_i an λ_j. To reduce computation time, a table of the double integral can be used in the implementation.[220] For calculating the DNA charge, the linear charge density v is chosen such that the potential at the radius of the DNA coincides

with the solution of the Poisson-Boltzmann equation for a cylinder with charge per length v_0^*. For DNA in the presence of the Gouy layer of confined counterions, this parameter can be computed as $v_0^* = qv_0$. In this relation, $v_0 = -2e/\Delta$ is the charge per length of the naked DNA, e is the proton charge, and $\Delta = 0.34\,\text{nm}$ is the distance between base pairs. As derived from experiments of single DNA molecules under tension and torsion, the value of q is 0.42[221] and not 0.73, which is the value for the effective charge of DNA in an external field.[225,226] To account for the presence of multivalent ions in the solvent, *e.g.* Mg^{2+} a modified Debye-Hückel-approximation was developed.[43,227] Furthermore, distributions of mono- and divalent ions around the nucleosome chain were calculated based on Poisson-Boltzmann theory.[73]

9.5.2.3 Nucleosome-nucleosome Interactions Potential

In the following, two methods are discussed in further detail for computing the interactions between nucleosomes. In the DiSCO (Discrete Surface Charge Optimization) model the nucleosome and the histone tails are represented by smaller spheres,[23,43,74] while the other approach uses a series expansion in S-functions.[21,44,67,70,228,229]

The DiSCO-model describes the electrostatic field predicted by Poisson-Boltzmann theory from a discrete set of Debye-Hückel charges. These are distributed on the surface of the cylindrically shaped nucleosome.[230] The level of detail of the model was increased in later studies by including also the histone tails.[23,43,74] The nucleosomal core particle surface is modeled by ~300 small spheres with fixed relative location, and the tails are represented by spheres located on a flexible line. For these spheres the electrical charge is computed based on the atomic structure.[127] The interaction between two nucleosomes described in this manner is computed by the Debye-Hückel-potential similar to that used for DNA, as discussed above. Additional excluded volume and attractive forces between nucleosomes are accounted for via 12-6 Lennard-Jones potentials.[43] This approach provides a detailed representation of the electrostatic interactions including the dynamics of the tails. However, for each pair of nucleosomes in a given chain conformation the Debye-Hückel potential between the ~300 charges of each nucleosome plus the associated Lennard-Jones potentials have to be calculated.

An alternative computationally less expensive method approximates nucleosome-nucleosome interactions *via* a series expansion in S-functions.[21,70,228,229] This yields a 12-6 Lennard-Jones potential that is shifted depending on the relative orientation of two nucleosomes.

$$U(\hat{o}_1, \hat{o}_2, \vec{r}) = 4\varepsilon(\hat{o}_1, \hat{o}_2, \hat{r})\left[\left(\frac{\sigma_0}{|\vec{r}| - \sigma(\hat{o}_1, \hat{o}_2, \hat{r}) + \sigma_0}\right)^{12} - \left(\frac{\sigma_0}{|\vec{r}| - \sigma(\hat{o}_1, \hat{o}_2, \hat{r}) + \sigma_0}\right)^{6}\right], \tag{9.3}$$

In eq. (9.3) $\hat{o}_1$ and $\hat{o}_2$ are unit vectors defining the orientation of the particles, $\vec{r}$ is the vector of the distance of the particles and σ_0 scales the potential width.

The potential strength ε and range σ parameters depend on the orientation and center-to-center difference vectors of two nucleosomes, and define the anisotropy of the potential. The dependency of ε and σ is described by a series expansion in S-functions, which are based on Wigner 3j-symbols.[228,229] Current models use a third order series expansion for particles with rotational symmetry.[21,70,229] The expansion coefficients for the strength and range are responsible for the dimension of the nucleosome shape and for the ratio of the energy strength between differently oriented nucleosomes (*e.g.* top-on-top and side-by-side), respectively. While the range coefficients reflect the dimensions of the nucleosome (width 11 nm, height 5.5 nm), the strength expansion coefficients are typically chosen in order to yield a ratio of 1/12 between side-by-side and top-on-top oriented nucleosomes. The parameterization and the shape of the potential can easily be redefined to evaluate other geometric dependencies, or to include additional nucleosome features. The series expansion approach allows for an efficient computation of nucleosome-nucleosome interactions that becomes relevant for simulations of larger systems.

9.6 MC simulations of Nucleosome Chains

9.6.1 General Considerations

A simulation using a Metropolis MC protocol samples a statistically representative ensemble of configurations in thermal equilibrium that follows a Boltzmann distribution.[60–62] Starting with an arbitrary configuration a Markov chain of new configurations is derived iteratively. The generation of one configuration based on the preceding configuration is called an MC step. The energy of the new configuration is computed as described in Section 9.5.2. If the energy difference ΔE between the new and the preceding configuration is less or equal zero the new configuration is accepted. In case that the new configuration is energetically less favorable, it is only accepted if $e^{-\Delta E/k_B T} > \zeta$ with k_B being the Boltzmann constant, T the temperature and a random value of $\zeta \in [0, 1)$. Otherwise, it is rejected. The configurations generated before equilibrium is reached are usually omitted from further analysis. Since in one MC step only parts of the chain are modified, successive conformations are correlated. This correlation needs to be considered to evaluate the significance of the results, and to compute correct error bars. Accordingly, for a given parameter like mass density or end-to-end-distance its correlation length l_c in terms of MC steps is computed. Two configurations separated by $2 \cdot l_c$ can be considered as being uncorrelated. This number typically has a linear relation to the number of beads.

9.6.2 Monte Carlo Moves

For fully exploring the conformational space of the coarse-grained nucleosome chain, suitable MC moves are needed. These originate from simulations of polymers, and comprise so-called pivot, rotation, crank-shaft, segment length

variation, local translation and rotation moves (Figure 9.5): (i) The pivot move rotates a terminal subchain part[231,232] (Figure 9.5A). A segment of the linear chain is selected randomly. In the next step, an arbitrary rotation axis is selected and the following segments are rotated around this axis by an angle chosen randomly from a given interval. (ii) For the rotation move a segment is chosen randomly, and a rotation axis is defined by the position of the preceding segment and the position of the subsequent segment[233,234] (Figure 9.5B). The chosen segment is rotated around this axis by a random angle within a given interval. (iii) The crank-shaft move is an extension of the rotation move.[235–237] Here, an entire subchain section is rotated (Figure 9.5C). From a randomly chosen start segment a defined number of segments is rotated around the axis spanning from the first bead to the last bead by an angle selected randomly within a given interval. This move is mostly used with different section lengths within a single simulation in order to produce movements at different scales. (iv) The local translation moves a single randomly chosen segment in a random direction by a randomly chosen distance from a given interval[43] (Figure 9.5D). (v) In local rotation moves a selected segment is rotated around a randomly chosen axis by a random angle out of a given interval[43] (Figure 9.5E). (vi) The segment length variation alters the length of a single randomly selected chain segment by a distance randomly chosen from a given interval[71] (Figure 9.5F). For nucleosomes, this move alters the distance d between the DNA entry and exit site (Figure 9.4A), while DNA segments are varied in their length. Intervals of the different moves are chosen to minimize the correlation length of

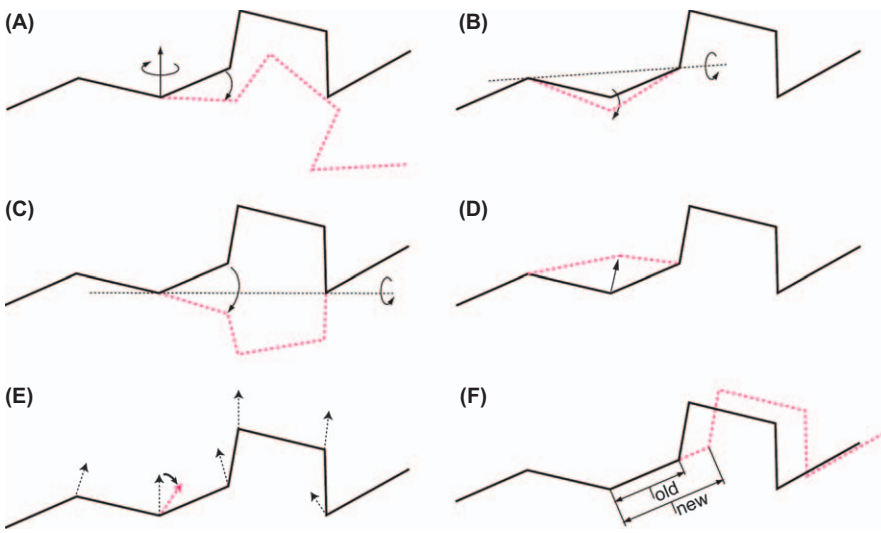

Figure 9.5 Schematic representation of the MC moves applied in different studies. (A) Pivot move. (B) Rotation move. (C) Crank-shaft move. (D) Local translation. (E) Local rotation. The vectors describing the orientation of a segment are shown. (F) Segment length variation.

subsequent configurations. This is typically achieved, if the acceptance rate is in the order of 50% for moves acting on single beads and 10–30% for moves acting on multiple beads. For the simulations that include explicit descriptions of the histone tails, the tails are sampled using an additional special MC move called tail regrowth.[43,74]

9.6.3 Reaching Thermodynamic Equilibrium in the MC Simulations

Upon increasing the strength of nucleosome-nucleosome interactions, the energy barriers for significant conformational changes become higher as nucleosomes tend to stick together. Once the configuration has reached a local minimum, the probability to accept a new configuration is very small and the system becomes trapped in this region of the state-space. Thus, thermal equilibrium is not reached, and the results represent only a non-representative section of the state-space.[21] Different *ad hoc* approaches were developed in order to avoid this problem.[65,70] A systematic approach to address this issue is based on feedback-optimized replica exchange.[21,238] In the first step a single configuration is simulated, starting with a high temperature. In a so-called simulated annealing step the temperature is decreased in small increments until the desired temperature is reached. The result of the simulated annealing process is used as input for the third step, where N systems are computed in parallel using standard Metropolis MC in which each replica has a different temperature. After a certain number of MC steps temperatures between neighboring replicas are swapped with a probability weighted by the energy difference between the systems – similar to the Metropolis criteria. Thus, the replicas are heated up and cooled down randomly and allowing them to escape from local minima. This approach generates an ensemble that follows a Boltzmann distribution.[239,240] In practice, the choice of an appropriate set of temperatures is difficult. With a so-called feedback optimized approach a viable set can be generated in a reproducible manner.[238] For configurations with an internucleosomal interaction strength of 9 k_BT or more a plain Metropolis MC protocol is no longer suitable, whereas the replica exchange method generates good results.[21]

9.7 Effect of Protein-protein and Protein-DNA Interactions on the Folding of the Nucleosome Chain

9.7.1 Orientation Dependence and Shape of Nucleosome-nucleosome Interactions

Several experimental findings indicate that the stacked alignment of two nucleosomes provides the most favorable conformation at physiological salt concentrations (Figure 9.6). This is inferred from nucleosome crystal structures,[209,241] studies of nucleosome liquid crystals,[242] electron microscopy observations,[22,134,243] and the nucleosome nearest neighbor distance

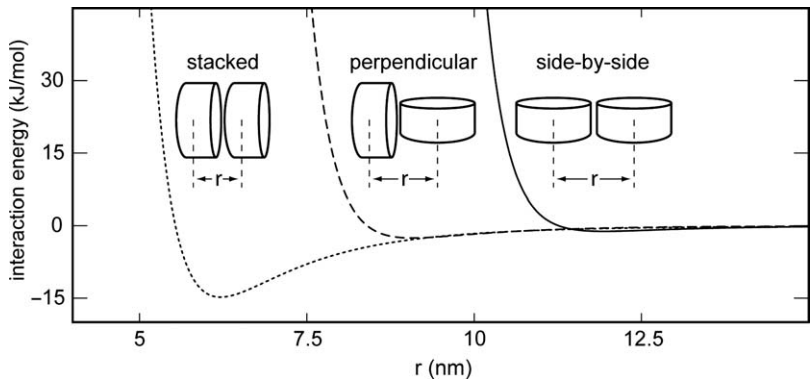

Figure 9.6 Estimated orientation dependence of nucleosome-nucleosome interaction potential. The interaction energy is shown as a function of the center-to-center distance *r* for differently oriented nucleosome pairs according to a potential used in coarse-grained computer simulations (scheme adapted from reference[177]). Interaction energies vary with distance r and have been parameterized in this example to reach about 14.6 kJ mol^{-1} (stacked nucleosomes), 0.5 kJ mol^{-1} (perpendicular) and 1.2 kJ mol^{-1} (side-by-side) at the optimal distance.

distributions determined by atomic force microscopy.[120,244] In addition to the stacked orientation, some favorable interaction energies may also be provided by histone octamers and nucleosomes that interact in a side-by-side orientation[242,245,246] (Figure 9.6). This type of associations is also observed for salt-dependent fiber-fiber interactions.[40,116–118,123,129,247] Both the stacked as well as the perpendicular/side-by-side interactions are critically dependent on histone tails as mentioned above.

9.7.2 The Strength of Nucleosome-nucleosome Interactions

Force spectroscopy studies provide a wealth of information on chromatin fiber features like their mechanical/elastic properties, stability and conformation/shape. Unfortunately, the available analytical descriptions do not provide the level of detail required to fit the experimental data sets appropriately and/or to faithfully extract parameters like the nucleosome-nucleosome interaction potential from them. Furthermore, the experimentally determined nucleosome interaction energies cover a rather broad range from 2 kcal mol^{-1} in native chromatin fibers[182] to 8 kcal mol^{-1} for reconstituted nucleosomal arrays[183] as mentioned above. Some insight on these observations can be obtained from MC simulations.[66,69,71,217] As discussed in reference[71] the large range of values is likely to reflect to which degree specific features of a given nucleosome chain allow for the establishment of optimal interactions between nucleosomes, as well as the solution environment studied. Thus, parameters like the average NRL, the regularity of nucleosome spacing as well the presence of linker

histones and divalent cations will be translated into differences in the effective nucleosome-nucleosome interactions. Under conditions where interactions between nucleosomes are optimal the attractive energies of up to 8 kcal mol^{-1} are significant and similar to that of unspecific binding of a protein to DNA. However, constraints imposed by the DNA linker and/or the local nucleosome geometry may counteract a favorable alignment. This can render the effective nucleosome-nucleosome interaction energy insufficient to establish a compact chromatin fiber-like structure. For example, it was shown that the unfavorable electrostatic repulsions and DNA bending/twisting energies of a ~60 bp long linker DNA (NRL = 207 bp) reduced the favorable contribution of nucleosome-nucleosome interaction to an effective value of $\sim 2\, k_B T$ in the absence of linker histones and divalent cations.[67,71] Open structures with low fiber mass density form in these simulations that are similar to those observed by electron microscopy images for nucleosome arrays without linker histones.[22]

9.7.3 Contributions of Histone Tails to Nucleosome-nucleosome Interactions

Modeling studies indicate that histone tails are important for mediating internucleosomal interactions and the folding of the nucleosome chain.[74,248] The results indicate that the positively charged tails neutralize negative phosphate charges of the DNA backbone and promote interactions between neighboring nucleosomes.[43,116,117,249,250] The contribution of the histone tails to nucleosome-nucleosome interactions is significant. In a theoretical study it was concluded that acetylation of a single H4K16 can reduce its value by almost $2\, k_B T$. [248] From comparison of complete nucleosomes and those with trypsinized tails interaction energies of $2\, k_B T$ [251] and 5–10 $k_B T$[113] were derived. When considering the results from computer simulations a total tail contribution of $\sim 5\, k_B T$ to the nucleosome-nucleosome interaction energy seems to be a reasonable estimate.[248,252]

9.7.4 Salt Dependence

The salt-dependent compaction of the nucleosome chain was experimentally studied in dependence of the chain length for samples with 2 to ~60 nucleosomes by measuring hydrodynamic parameters like the sedimentation coefficient and the diffusion coefficient for samples from rat liver nuclei,[86,87] chicken erythrocytes,[82,83] nuclei from bovine thymus,[88] HeLa cells[93] and reconstituted nucleosome arrays.[11,22,26,253] The expected dependence of the sedimentation coefficient s on the molecular weight M of the fiber is $s \propto \ln M$ for short rodlike shapes and $s \propto M^{1/2}$ for longer chains that are in a wormlike coil conformation with deviations of the exponent from 1/2 reflecting excluded-volume effects.[67] From these studies several findings are noteworthy: (i) Polynucleosomes as well as short trinucleosome samples display a significant compaction at concentrations above 30 to 40 mM monovalent ions.[83,86] This reflects a transition

from the decondensed 10 nm chain to the 30 nm fiber conformation that is driven by an increased strength of the nucleosome-nucleosome interactions in the physiological concentrations. (ii) Divalent ions like Mg^{2+} at typical concentrations of 1–3 mM seem to be particularly effective in inducing chain compaction.[22,253] (iii) An additional favorable electrostatic contribution is provided by binding of linker histones.[10,22,140,146,155]

These complex electrostatic effects are not fully accounted for in many coarse-grained models that are parameterized to provided appropriate interaction energies only for a solution environment of 0.1 M monovalent salt, *e.g.* reference.[21,67] However, more elaborate models account explicitly for salt-dependent surface site charges and the electrostatic energy between DNA-linkers and nucleosomes. In the corresponding MC simulations the experimentally observed compaction behavior is reproduced.[72] Raising the salt concentration from 10 mM to 100 mM decreased the repulsion between DNA-linkers and increased the attraction energy between nucleosomes for a 12mer array with a total energy difference of 54 kcal mol^{-1} for a 12 nucleosome array (equivalent to 2.7 k_BT per nucleosome) in the simulations. The associated conformational changes from an extended chain with 27 S at 10 mM salt to 40 S in the compacted state at 200 mM were in good agreement with experimental results.[26] This approach was recently extended to account for the contribution of divalent ions and linker histones in combination with monovalent salt.[43,73] Arrays of 12 and 24 nucleosomes were sampled in MC simulations to compute the ion distribution and conformation of the chain. An alternative approach to MC simulations are coarse-grained MD simulations of nucleosome arrays investigating the effect of ions and histone tails to modulate electrostatic interactions and to induce chain folding.[56–58]

9.8 The Effect of Nucleosome Positioning and Local Geometry on Chain Conformation

9.8.1 Effect of DNA Linker Length on the Chromatin Fiber Conformation

The effect of the linker DNA length and entry-exit angle on fiber formations was investigated for two conformations: a crossed-linker chain with [2,1] and an interdigitated [7,3] geometry. To account for the helical twist of the DNA, the torsion angle of the DNA linker was changed by 36° per base pair. The results revealed a large dependency of the fiber compaction on the linker DNA length (Figure 9.7). Increasing the NRL reduced fiber compaction, while the opening angle had only a moderate impact over the range of values studies. A similar 10-bp periodicity was also found for the model of chicken erythrocyte chromatin type fibers that adopted right-handed [3,1] fiber conformations for NRLs of 202 and 212 bp, while for NRLs of 206 and 216 bp, more open left-handed [2,1] fiber conformations were detected.[21] It is noted that both geometries display a 10-bp periodicity for the peaks of the linear mass density,

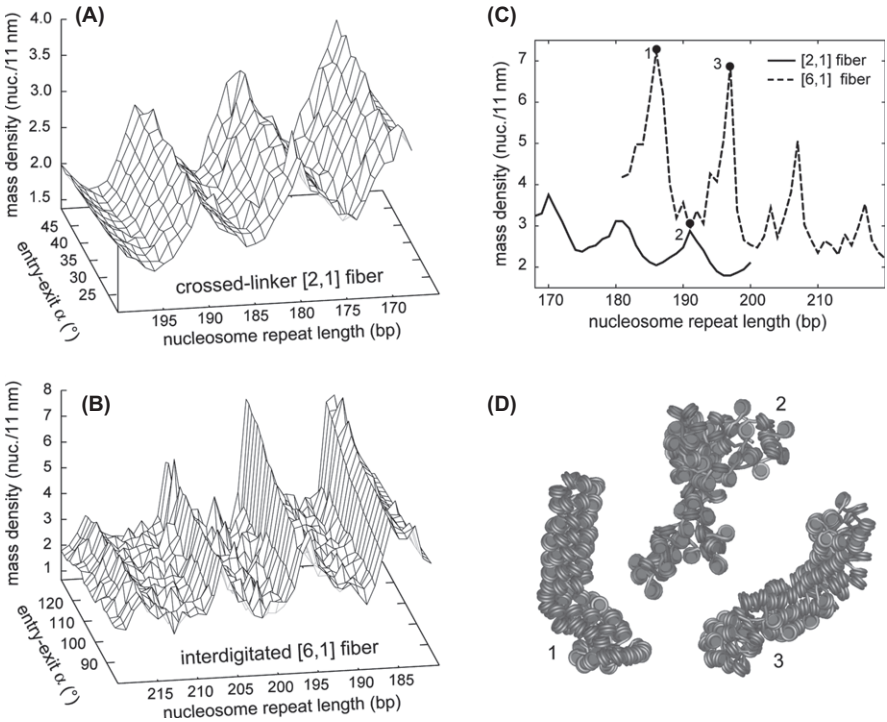

Figure 9.7 Dependence of the mass density on the NRL and local nucleosome geometry (adapted from ref. 67). A crossed-linker [2,1] and an interdigitated [6,1] fiber geometry with high nucleosome tilt angles are compared. Each data point represents the mean value of the equilibrium ensemble. (A) [2,1] fiber. (B) [6,1] fiber. (C) Comparison of [2,1] and [6,1] fibers for entry-exit angles of 35° and 117.5°, respectively. (D) Configurations from the indicated points for the [6,1] fiber data shown in panel C illustrate the observed fluctuations in mass density for NRLs of 186 bp (#1), 191 bp (#2), and 197 bp (#3).

which is consistent with the distribution of NRLs found in natural sequences.[107]

9.8.2 Transitions between Fiber Conformation

The results reviewed here clearly demonstrate that the nucleosome chain is polymorphic and can organize into a variety of conformations. These states are determined by the positioning of nucleosomes along the DNA (both with respect to the spacing regularity and separation distance), the protein composition of the chromatin fragment (histone variants, presence/type of linker histone, other architectural proteins) as well as post-translational histone modifications. Relatively small variations of these parameters can induce changes of the local nucleosome geometry that translate into large scale

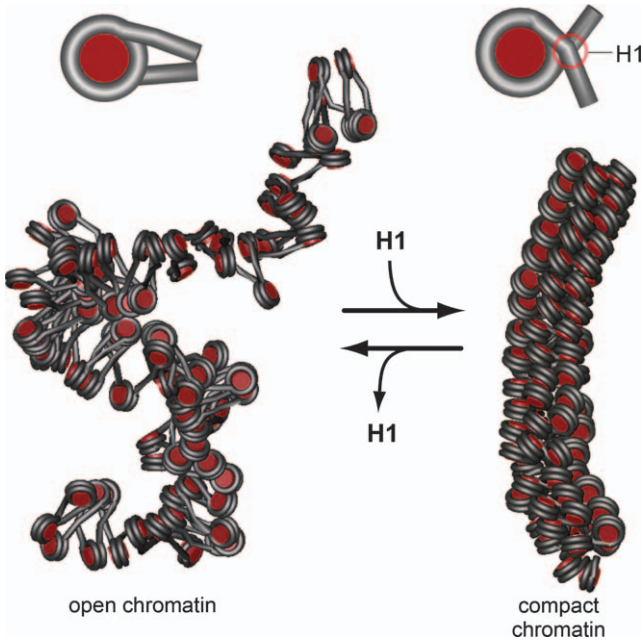

Figure 9.8 Model for chromatin fiber compaction induced by changing the local nucleosome geometry via binding of linker histone H1. A coarse-grained model of a chain with 100 nucleosomes was subjected to Monte Carlo simulations. The putative change of the DNA geometry due to binding of linker histone H1 at the DNA entry-exit site of the nucleosome leads to a compaction of the chain into a condensed fiber structure with a diameter of about 30 nm.[67]

rearrangements of the overall chain conformation. An example for such a conformational transition as studied by MC simulations is given in Figure 9.8. Two local nucleosome geometries, one without and one with bound linker histone H1 are considered. In the latter, H1 induces a change in the conformation of the DNA at the nucleosome entry-exit sites and neutralizes DNA phosphate charges from about 20 bp of linker DNA. If H1 is missing an open [2,1] fiber conformation is observed in the MC simulations for a local nucleosome geometry that mimics the one found in the tetranucleosome crystal structure (Figure 9.3C, F) but with an NRL extended to ~189 bp. In this open state, other protein factors could easily access most of the linker DNA. Upon binding of H1 the local geometry of the DNA at the nucleosome entry-exit is changed as depicted in the coarse-grained model (Figure 9.8). As a result, a different folding of the chain into a [6,1] fiber is induced. This conformation has a higher linear mass density and the linker DNA becomes located in the interior of the fiber so that binding of other factors to this part of the DNA is impeded. Thus, the observed changes of nucleosome chain compaction over a ~10fold range from 1–2 nucleosomes per 11 nm in a very open conformation[20] up to 17 nucleosomes per 11 nm fiber if fully condensed[19] could regulate linker

DNA access for other protein factors. This effect has been demonstrated in experiments that reveal large differences for the binding of protein factors to a folded and unfolded 17mer nucleosomal array.[9]

9.8.3 Systematically Exploring the Conformation Space with MC Simulated Phase Diagrams

For the crossed-linker DNA fiber family conformations with [3,1], [5,2] and [7,3] geometry were proposed,[47,51,254] which suggests a potentially high conformational variability. However, the range of geometrically possible structures needs to be evaluated in terms of their stability. One approach to systematically search for low energy fiber conformations is illustrated in Figure 9.9.[44] An energy-minimized phase diagram based on the [2,1] fiber conformation derived for chicken erythrocyte chromatin fibers by Woodcock and coworkers[79,205] is

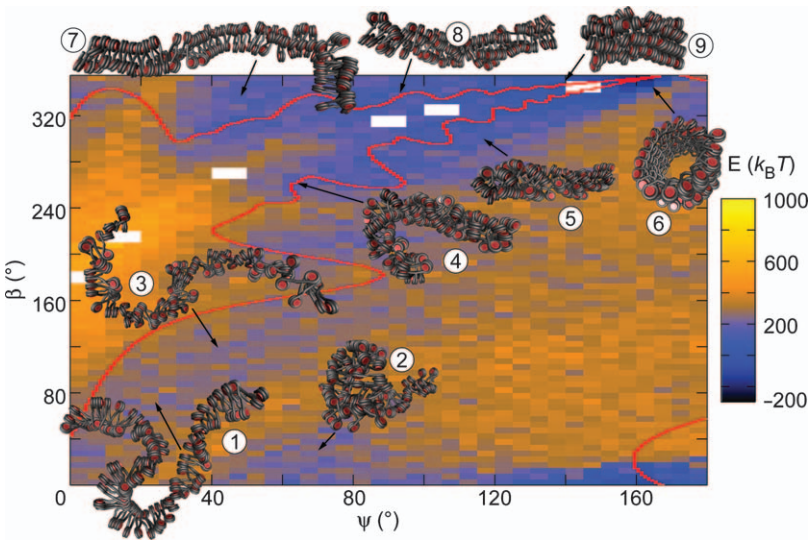

Figure 9.9 Phase diagram for energy minimized crossed-linker fibers with a nucleosome stem structure. The initial structure was parameterized to fit the data of native chromatin of chicken erythrocytes with an NRL of 212 bp.[79,205] This corresponds to fiber #1. The local geometry of the nucleosome was changed by varying the initial values of linker DNA torsion angle β and the opening angle ψ of the DNA for its entry-exit site at the nucleosome, and then minimizing the energies of the resulting structures. Within the conformational space explored in this manner, distinct subgroups of fiber conformations can be identified that vary in their stability as reflected by the color-coding. Stable conformations comprised [2,1] fibers (#1–3), a [3,1] fiber (#4) as well as [n,1] fiber conformations with $n > 3$ (fibers #5, 6, 8, and 9). White regions indicate sterically impossible conformations. The red contour line marks the border between sterically possible and impossible conformations in the initial structures, *i.e.* without allowing linker DNA bending and twisting. The image is from reference.[44]

shown. By systematically varying the local nucleosome geometry a number of additional fiber structures could be identified that were stable in the computer simulations.[44]

9.8.4 Resolution of the Chromatin Fiber Structure

Studies of fiber structures *in vitro* are typically conducted with short fragments (<100 nucleosomes) at nucleosome concentrations in the 1 μM range.[19,134] However, the nucleus represents an environment that is highly enriched with nucleosomes and DNA. Nucleosome concentrations during the interphase of the cell cycle are estimated to vary between 60–450 μM during interphase and can reach ~1.2 mM in the mitotic chromosome.[255] Under these conditions the 30 nm fiber conformation might resolve into a "sea of nucleosomes".[31–35] Alternatively, more irregular and aggregate-like structures could form where nucleosomes from distant parts or from other chromosomes would intermingle. These nucleosome-nucleosome interactions in *cis* and in *trans* become more favorable as the nucleosome concentration is raised and the chain length is increased, which will facilitate its back-folding. Technical improvements of MC simulations make it possible to compute realistic ensembles of nucleosome chains with up to ~1000 nucleosomes on the time scale of weeks on current multiprocessor computer systems. The influence of chain length in conjunction with strength and geometry of the internucleosomal interaction potential on the spatial organization of the fiber is illustrated in Figure 9.10. At a maximal interaction energy of $E_{max} = 6\,k_B T$ (corresponding to an effective average energy of $E_{eff} \sim 3.5\,k_B T$) arrays of 100 nucleosomes always formed fiber structures for the local nucleosome geometry used in these simulations, while chains with 250 nucleosomes displayed already some tendency to fold back (Figure 9.10A). This effect was enhanced when further extending the chain to 500 nucleosomes and/or increasing the value of E_{max} to 12 $k_B T$ (Figure 9.10B). Another factor that will promote chain back folding or formation of aggregation is the ratio of the lateral to the perpendicular maximum internucleosomal interaction strength (Figure 9.6). When this parameter was decreased to 1 to 5 as opposed to the standard value of 1 to 10 the chains folded into more compact structures (Figure 9.10C). Such a modulation of the orientation dependence of the nucleosome interaction potential could be induced *in vivo via* the binding of other chromosomal proteins or histone modifications as discussed above.

9.9 Perspectives

The nucleosome chain represents a highly complex, dynamic and polymorphic supramolecular system. While the structure of its basic building block, the nucleosome, is known at atomic resolution, the conformational flexibility of the nucleosome chain that exists even for homogenous samples in conjunction with variations in protein composition and post-translational histone modification found in native chromatin is incompatible with a high resolution structure

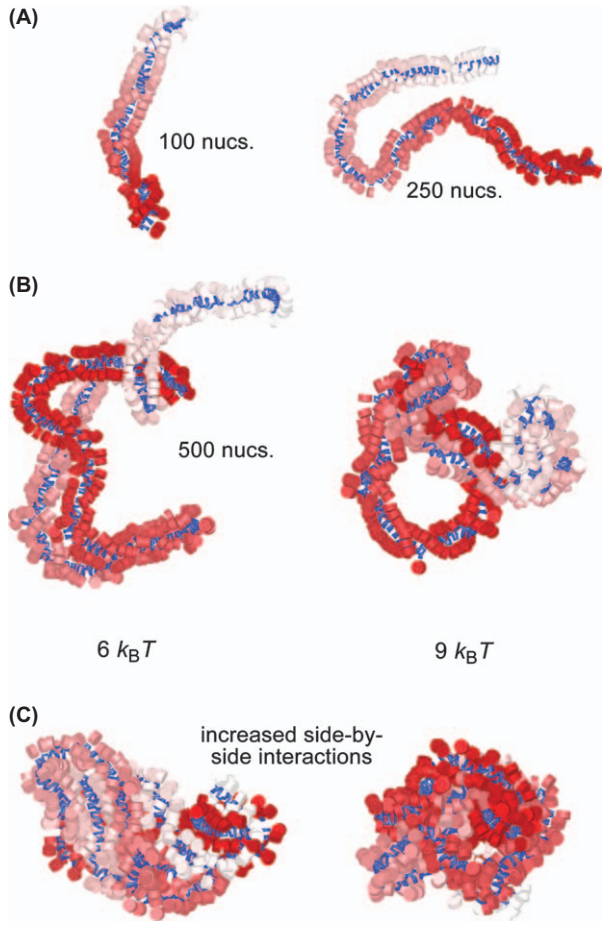

Figure 9.10 Resolution of a 30 nm chromatin fiber in dependence of the nucleosome-nucleosome interaction potential and chain length. All structures are computed with a simulated annealing step and subsequent replica exchange MC simulations of $\sim 10^7$ simulation steps. (A) At a maximal interaction energy of $E_{max} = 6\ k_BT$ (corresponding to an effective average energy of $E_{eff} \sim 3.5\, k_BT$) 100 nucleosome arrays always formed fiber structures while chains with 250 nucleosomes displayed some tendency to fold back. (B) Fibers with 500 nucleosomes always displayed some higher order folding. This feature became more pronounced when increasing the value of E_{max} from 6 to $9\, k_BT$ (C) The ratio of the lateral to the perpendicular maximum internucleosomal interaction strength was decreased to 1 to 5 as opposed to 1 to 10 in panel B (see Figure 9.6). For the higher side-by-side interactions the chains fold into more compact structures.

analysis beyond that of a few nucleosomes. MC simulations are ideally suited to address the need for characterizing the folding of the nucleosome chain since they derive a representative ensemble of conformations that reflects the conformational flexibility of the nucleosome chain according to the distribution

obtained at thermodynamic equilibrium. They allow it to take advantage of our detailed and increasing knowledge of the nucleosome in all its variant forms as the building block of chromatin to evaluate the resulting differential organization of the chromatin fiber. To further investigate the potentially large impact of the nucleosome chain conformation to regulate DNA access and associated molecular biological processes in MC simulations a number of issues need to be addressed: (i) The conformation(s) and effects of linker histone binding to the nucleosome and associated linker DNA remains to be determined. Currently, a number of largely different model structures have been proposed,[139,146–151] which translate into very different conformations when extended into fiber structures. (ii) The strength of the nucleosome-nucleosome and its dependence on the spatial orientation of interacting nucleosomes has to be determined more precisely. Important information has been obtained from the analysis of single force spectroscopy and other experiments (see section 9.3 and 9.7). However, a direct approach to measure the distance and orientation dependence of the interaction between two nucleosome core particles would be a valuable contribution to parameterize this potential in MC simulations as discussed in the context of Figures 9.6 and 9.10. (iii) It will be necessary to extend current experimental *in vitro* studies as well as their simulations to much larger nucleosome chains and/or higher nucleosome concentration. As discussed in section 9.8.4, this will provide information on nucleosome chain organization in an environment that is more similar to that encountered in the nucleus. (iv) Currently, much progress is made in experimental studies to identify different functional chromatin states and characterize these states in terms of protein composition and histone and DNA modifications.[256–258] It will be a new challenging task to integrate the chromatin states defined in this manner into chromatin models and investigate their conformational and thermodynamic features in numerical simulations. Advancements in these four areas will be crucial to improve coarse-grained nucleosome chain description and advancing the application of MC simulations of chromatin. We expect these developments to lead to a more comprehensive quantitative description of nucleosome chain folding to understand its organization and function *in vitro* as well as in the cell nucleus.

Acknowledgements

We are grateful to Nick Kepper, Ramona Ettig and Vladimir Teif for discussions and Oliver Müller and Robert Schöpflin for critical reading of the manuscript. Our work on chromatin conformation is supported within the project EpiGenSys by the BMBF as a partner of the ERASysBio+ initiative supported under the EU ERA-NET Plus scheme in FP7.

References

1. K. E. van Holde, *Chromatin*, Springer, Heidelberg, 1989.
2. M. L. Dechassa and K. Luger, in *Genome Organization and Function in the Cell Nucleus*, ed. K. Rippe, Wiley-VCH, Weinheim, 2012, pp. 55–87.

3. E. I. Campos and D. Reinberg, *Annu. Rev. Genet.*, 2009, **43**, 559–599.
4. B. M. Lee and L. C. Mahadevan, *J. Cell. Biochem.*, 2009, **108**, 22–34.
5. S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis and D. J. Patel, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1025–1040.
6. M. Wachsmuth, M. Caudron-Herger and K. Rippe, *Biochim. Biophys. Acta*, 2008, **1783**, 2061–2079.
7. M. Radman-Livaja and O. J. Rando, *Dev. Biol.*, 2010, **339**, 258–266.
8. C. Jiang and B. F. Pugh, *Nat. Rev. Genet.*, 2009, **10**, 161–172.
9. M. G. Poirier, M. Bussiek, J. Langowski and J. Widom, *J. Mol. Biol.*, 2008, **379**, 772–786.
10. P. J. Robinson and D. Rhodes, *Curr. Opin. Genet. Dev.*, 2006, **16**, 336–343.
11. J. C. Hansen, *Annu. Rev. Biophys. Biomol. Struct.*, 2002, **31**, 361–392.
12. C. L. Woodcock, A. I. Skoultchi and Y. Fan, *Chromosome Res.*, 2006, **14**, 17–25.
13. A. Bassett, S. Cooper, C. Wu and A. Travers, *Curr. Opin. Genet. Dev.*, 2009, **19**, 159–165.
14. M. P. Marsden and U. K. Laemmli, *Cell*, 1979, **17**, 849–858.
15. B. A. Hamkalo and J. B. Rattner, *Q. Rev. Biol.*, 1980, **55**, 409–417.
16. C. L. Woodcock, *J. Cell. Biol.*, 1994, **125**, 11–19.
17. K. Andersson, B. Bjorkroth and B. Daneholt, *J. Cell Biol.*, 1984, **98**, 1296–1303.
18. J. Widom, *Annu. Rev. Biophys. Biophys. Chem.*, 1989, **18**, 365–395.
19. P. J. Robinson, L. Fairall, V. A. Huynh and D. Rhodes, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 6506–6511.
20. J. Dekker, *J. Biol. Chem.*, 2008, **283**, 34532–34540.
21. R. Stehr, N. Kepper, K. Rippe and G. Wedemann, *Biophys. J.*, 2008, **95**, 3677–3691.
22. A. Routh, S. Sandin and D. Rhodes, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 8872–8877.
23. O. Perisic, R. Collepardo-Guevara and T. Schlick, *J. Mol. Biol.*, 2010, **403**, 777–802.
24. S. E. Gerchman and V. Ramakrishnan, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 7802–7806.
25. R. Ghirlando and G. Felsenfeld, *J. Mol. Biol.*, 2008, **376**, 1417–1425.
26. J. C. Hansen, J. Ausio, V. H. Stanik and K. E. van Holde, *Biochemistry*, 1989, **28**, 9129–9136.
27. K. van Holde and J. Zlatanova, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 10548–10555.
28. J. Bednar, R. A. Horowitz, J. Dubochet and C. L. Woodcock, *J. Cell Biol.*, 1995, **131**, 1365–1376.
29. M. Shogren-Knaak, H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie and C. L. Peterson, *Science*, 2006, **311**, 844–847.
30. P. J. J. Robinson, W. An, A. Routh, F. Martino, L. Chapman, R. G. Roeder and D. Rhodes, *J. Mol. Biol.*, 2008, **381**, 816–825.
31. J. Dubochet, M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowall and P. Schultz, *Q. Rev. Biophys.*, 1988, **21**, 129–228.

32. K. Maeshima, S. Hihara and M. Eltsov, *Curr. Opin. Cell Biol.*, 2010, **22**, 291–297.
33. E. Fussner, R. W. Ching and D. P. Bazett-Jones, *Trends Biochem. Sci.*, 2011, **36**, 1–6.
34. M. Eltsov, K. M. Maclellan, K. Maeshima, A. S. Frangakis and J. Dubochet, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 19732–19737.
35. K. Maeshima, S. Hihara and H. Takata, *Cold Spring Harbor Symp. Quant. Biol.*, 2010, **75**, 439–444.
36. M. P. Scheffer, M. Eltsov and A. S. Frangakis, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 16992–16997.
37. J. T. Finch and A. Klug, *Proc. Natl. Acad. Sci. USA*, 1976, **73**, 1897–1901.
38. F. Thoma, T. Koller and A. Klug, *J. Cell. Biol.*, 1979, **83**, 403–427.
39. J. B. Rattner and B. A. Hamkalo, *J. Cell Biol.*, 1979, **81**, 453–457.
40. J. Widom and A. Klug, *Cell*, 1985, **43**, 207–213.
41. R. Ettig, N. Kepper, R. Stehr, G. Wedemann and K. Rippe, *Biophys. J.*, 2011, **101**, 1999–2008.
42. K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. Langowski and J. Smith, *J. Comput. Chem.*, 2008, **29**, 1429–1439.
43. G. Arya and T. Schlick, *J. Phys. Chem. A*, 2009, **113**, 4045–4059.
44. R. Stehr, R. Schöpflin, R. Ettig, N. Kepper, K. Rippe and G. Wedemann, *Biophys. J.*, 2010, **98**, 1028–1037.
45. C. Münkel, R. Eils, S. Dietzel, D. Zink, C. Mehring, G. Wedemann, T. Cremer and J. Langowski, *J. Mol. Biol.*, 1999, **285**, 1053–1065.
46. H. Schiessel, W. M. Gelbart and R. Bruinsma, *Biophys. J.*, 2001, **80**, 1940–1956.
47. M. Depken and H. Schiessel, *Biophys. J.*, 2009, **96**, 777–784.
48. P. M. Diesinger and D. W. Heermann, *Phys. Rev. E*, 2006, **74**, 031904.
49. E. F. Koslover, C. J. Fuller, A. F. Straight and A. J. Spakowitz, *Biophys. J.*, 2010, **99**, 3941–3950.
50. A. Scipioni, G. Turchetti, S. Morosetti and P. De Santis, *Biophys. Chem.*, 2010, **148**, 56–67.
51. H. Wong, J.-M. Victor and J. Mozziconacci, *PLoS ONE*, 2007, **2**, e877.
52. S. A. Adcock and J. A. McCammon, *Chem. Rev.*, 2006, **106**, 1589–1615.
53. T. C. Bishop, *J. Biomol. Struct. Dyn.*, 2005, **22**, 673–686.
54. T. Bishop, *Biophys. J.*, 2008, **95**, 1007–1017.
55. V. B. Teif, R. Ettig and K. Rippe, *Biophys. J.*, 2010, **99**, 2597–2607.
56. A. Allahverdi, R. Yang, N. Korolev, Y. Fan, C. A. Davey, C. F. Liu and L. Nordenskiold, *Nucleic Acids Res.*, 2011, **39**, 1680–1691.
57. N. Korolev, A. Allahverdi, Y. Yang, Y. Fan, A. P. Lyubartsev and L. Nordenskiold, *Biophys. J.*, 2010, **99**, 1896–1905.
58. N. Korolev, A. P. Lyubartsev and L. Nordenskiold, *Adv. Colloid Interface Sci.*, 2010, **158**, 32–47.
59. D. L. Ermak and J. A. McCammon, *J. Chem. Phys.*, 1978, **69**, 1352–1359.
60. N. Metropolis, A. Rosenbluth, M. Roenbluth, A. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1092.

61. K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction*, 5th edn, Springer, Berlin, 2010.
62. D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, 3rd edn, Cambridge University Press, Cambridge, 2009.
63. D. A. Beard and T. Schlick, *Structure*, 2001, **9**, 105–114.
64. G. Wedemann and J. Langowski, *Biophys. J.*, 2002, **82**, 2847–2859.
65. B. Mergell, R. Everaers and H. Schiessel, *Phys. Rev. E*, 2004, **70**, 011915.
66. F. Aumann, F. Lankas, M. Caudron and J. Langowski, *Phys. Rev. E*, 2006, **73**, 041927.
67. N. Kepper, D. Foethke, R. Stehr, G. Wedemann and K. Rippe, *Biophys. J.*, 2008, **95**, 3692–3705.
68. F. Aumann, J. Sühnel, J. Langowski and S. Diekmann, *Theor. Chem. Acc.*, 2010, **125**, 217–231.
69. V. Katritch, C. Bustamante and W. K. Olson, *J. Mol. Biol.*, 2000, **295**, 29–40.
70. G. Cinacchi, G. La Penna and A. Perico, *Macromolecules*, 2007, **40**, 9603–9613.
71. N. Kepper, R. Ettig, R. Stehr, G. Wedemann and K. Rippe, *Biopolymers*, 2011, **95**, 435–447.
72. J. Sun, Q. Zhang and T. Schlick, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 8180–8185.
73. H. H. Gan and T. Schlick, *Biophys. J.*, 2010, **99**, 2587–2596.
74. G. Arya and T. Schlick, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 16236–16241.
75. Y. Yang, A. P. Lyubartsev, N. Korolev and L. Nordenskiold, *Biophys. J.*, 2009, **96**, 2082–2094.
76. P. M. Diesinger, S. Kunkel, J. Langowski and D. W. Heermann, *Biophys. J.*, 2010, **99**, 2995–3001.
77. R. Ghirlando, M. D. Litt, M. N. Prioleau, F. Recillas-Targa and G. Felsenfeld, *J. Mol. Biol.*, 2004, **336**, 597–605.
78. J. Dubochet, M. Adrian, P. Schultz and P. Oudet, *EMBO J.*, 1986, **5**, 519–528.
79. J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster and C. L. Woodcock, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 14173–14178.
80. S. H. Leuba, G. Yang, C. Robert, B. Samori, K. van Holde, J. Zlatanova and C. Bustamante, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 11621–11625.
81. A. Hamiche, P. Schultz, V. Ramakrishnan, P. Oudet and A. Prunell, *J. Mol. Biol.*, 1996, **257**, 30–42.
82. J. Ausio, N. Borochov, D. Seger and H. Eisenberg, *J. Mol. Biol.*, 1984, **177**, 373–398.
83. P. J. Butler and J. O. Thomas, *J. Mol. Biol.*, 1998, **281**, 401–407.
84. D. L. Bates, P. J. Butler, E. C. Pearson and J. O. Thomas, *Eur. J. Biochem.*, 1981, **119**, 469–476.
85. M. H. Koch, M. C. Vega, Z. Sayers and A. M. Michon, *Eur. Biophys. J.*, 1987, **14**, 307–319.

86. P. J. Butler and J. O. Thomas, *J. Mol. Biol.*, 1980, **140**, 505–529.
87. E. C. Pearson, P. J. Butler and J. O. Thomas, *EMBO J.*, 1983, **2**, 1367–1372.
88. J. M. Gale and M. J. Smerdon, *Biochemistry*, 1988, **27**, 7197–7205.
89. J. O. Thomas, C. Rees and P. J. Butler, *Eur. J. Biochem.*, 1986, **154**, 343–348.
90. C. Spadafora, M. Bellard, J. L. Compton and P. Chambon, *FEBS Lett.*, 1976, **69**, 281–285.
91. J. B. Rattner, C. Saunders, J. R. Davie and B. A. Hamkalo, *J. Cell Biol.*, 1982, **93**, 217–222.
92. A. T. Annunziato, L. L. Frado, R. L. Seale and C. L. Woodcock, *Chromosoma*, 1988, **96**, 132–138.
93. J. F. Kepert, J. Mazurkiewicz, G. Heuvelman, K. Fejes Tóth and K. Rippe, *J. Biol. Chem.*, 2005, **280**, 34063–34072.
94. R. A. Horowitz-Scherer and C. L. Woodcock, *Chromosoma*, 2006, **115**, 1–14.
95. C. L. Woodcock, *Curr. Opin. Genet. Dev.*, 2006, **16**, 213–220.
96. H. J. Szerlong and J. C. Hansen, *Biochem. Cell Biol.*, 2011, **89**, 24–34.
97. G. Li and D. Reinberg, *Curr. Opin. Genet. Dev.*, 2011, **21**, 175–186.
98. K. Rippe, J. Mazurkiewicz and N. Kepper, in *DNA interactions with polymers and surfactants*, eds. R. S. Dias and B. Lindman, Wiley, London, 2008, pp. 135–172.
99. G. Längst, V. B. Teif and K. Rippe, in *Genome organization and function in the cell nucleus*, ed. K. Rippe, Wiley-VCH, Weinheim, 2012, pp. 111–138.
100. R. T. Simpson, F. Thoma and J. M. Brubaker, *Cell*, 1985, **42**, 799–808.
101. P. T. Lowary and J. Widom, *J. Mol. Biol.*, 1998, **276**, 19–42.
102. E. Segal and J. Widom, *Trends. Genet.*, 2009, **25**, 335–343.
103. K. van Holde and J. Zlatanova, *J. Biol. Chem.*, 1995, **270**, 8373–8376.
104. J. Zlatanova, S. H. Leuba and K. van Holde, *Biophys. J.*, 1998, **74**, 2554–2566.
105. A. Levy, M. Eyal, G. Hershkovits, M. Salmon-Divon, M. Klutstein and D. J. Katcoff, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 11703–11708.
106. A. B. Lantermann, T. Straub, A. Stralfors, G. C. Yuan, K. Ekwall and P. Korber, *Nat. Struct. Mol. Biol.*, 2010, **17**, 251–257.
107. J. Widom, *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 1095–1099.
108. A. Valouev, S. M. Johnson, S. D. Boyd, C. L. Smith, A. Z. Fire and A. Sidow, *Nature*, 2011.
109. K. J. Polach, P. T. Lowary and J. Widom, *J. Mol. Biol.*, 2000, **298**, 211–223.
110. J. M. Vitolo, C. Thiriet and J. J. Hayes, *Mol. Cell. Biol.*, 2000, **20**, 2167–2175.
111. C. Vogler, C. Huber, T. Waldmann, R. Ettig, L. Braun, I. Chassignet, A. J. Lopez-Contreras, O. Fernandez-Capetillo, M. Dundr, K. Rippe, G. Längst and R. Schneider, *PloS Genet.*, 2010, **6**, e1001234.
112. A. Bertin, A. Leforestier, D. Durand and F. Livolant, *Biochemistry*, 2004, **43**, 4773–4780.
113. A. Bertin, M. Renouard, J. S. Pedersen, F. Livolant and D. Durand, *Biophys. J.*, 2007, **92**, 2633–2645.

114. T. M. Fletcher and J. C. Hansen, *J. Biol. Chem.*, 1995, **270**, 25359–25362.
115. C. Tse and J. C. Hansen, *Biochemistry*, 1997, **36**, 11381–11388.
116. B. Dorigo, T. Schalch, K. Bystricky and T. J. Richmond, *J. Mol. Biol.*, 2003, **327**, 85–96.
117. F. Gordon, K. Luger and J. C. Hansen, *J. Biol. Chem.*, 2005, **280**, 33701–33706.
118. P. M. Schwarz, A. Felthauser, T. M. Fletcher and J. C. Hansen, *Biochemistry*, 1996, **35**, 4009–4015.
119. X. Wang and J. J. Hayes, *Mol. Cell. Biol.*, 2008, **28**, 227–236.
120. F. J. Solis, R. Bash, J. Yodh, S. M. Lindsay and D. Lohr, *Biophys. J.*, 2004, **87**, 3372–3387.
121. M. Garcia-Ramirez, C. Rocchini and J. Ausio, *J. Biol. Chem.*, 1995, **270**, 17923–17928.
122. C. Tse, T. Sera, A. P. Wolffe and J. C. Hansen, *Mol. Cell Biol.*, 1998, **18**, 4629–4638.
123. P. Y. Kan, T. L. Caterino and J. J. Hayes, *Mol. Cell Biol.*, 2009, **29**, 538–546.
124. X. Wang and J. J. Hayes, *Biochem. Cell Biol.*, 2006, **84**, 578–588.
125. X. Lu, M. D. Simon, J. V. Chodaparambil, J. C. Hansen, K. M. Shokat and K. Luger, *Nat. Struct. Mol. Biol.*, 2008, **15**, 1122–1124.
126. J. S. Choy, S. Wei, J. Y. Lee, S. Tan, S. Chu and T.-H. Lee, *J. Am. Chem. Soc.*, 2010, **132**, 1782–1783.
127. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251–260.
128. C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond, *J. Mol. Biol.*, 2002, **319**, 1097–1113.
129. P. Y. Kan and J. J. Hayes, *Methods*, 2007, **41**, 278–285.
130. C. Zheng, X. Lu, J. C. Hansen and J. J. Hayes, *J. Biol. Chem.*, 2005, **280**, 33552–33557.
131. R. K. Suto, M. J. Clarkson, D. J. Tremethick and K. Luger, *Nat. Struct. Biol.*, 2000, **7**, 1121–1124.
132. J. Zhou, J. Y. Fan, D. Rangasamy and D. J. Tremethick, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1070–1076.
133. J. Y. Fan, F. Gordon, K. Luger, J. C. Hansen and D. J. Tremethick, *Nat. Struct. Biol.*, 2002, **9**, 172–176.
134. B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder and T. J. Richmond, *Science*, 2004, **306**, 1571–1573.
135. R. T. Simpson, *Biochemistry*, 1978, **17**, 5524–5531.
136. M. Noll and R. D. Kornberg, *J. Mol. Biol.*, 1977, **109**, 393–404.
137. W. An, S. H. Leuba, K. van Holde and J. Zlatanova, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 3396–3401.
138. V. Graziano, S. E. Gerchman, D. K. Schneider and V. Ramakrishnan, *Nature*, 1994, **368**, 351–354.
139. Y. B. Zhou, S. E. Gerchman, V. Ramakrishnan, A. Travers and S. Muyldermans, *Nature*, 1998, **395**, 402–405.
140. V. Ramakrishnan, *Crit. Rev. Eukaryot. Gene Expr.*, 1997, **7**, 215–230.

141. J. Zlatanova and K. v. Holde, *Prog. Nucl. Acid Res. Mol. Biol.*, 1996, **52**, 217–259.
142. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee and R. M. Sweet, *Nature*, 1993, **362**, 219–223.
143. D. J. Clark and T. Kimura, *J. Mol. Biol.*, 1990, **211**, 883–896.
144. L. M. Carruthers, J. Bednar, C. L. Woodcock and J. C. Hansen, *Biochemistry*, 1998, **37**, 14776–14787.
145. S. Pennings, G. Meersseman and E. M. Bradbury, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 10275–10279.
146. A. Travers, *Trends Biochem. Sci.*, 1999, **24**, 4–7.
147. M. M. Bharath, N. R. Chandra and M. R. Rao, *Nucleic Acids Res.*, 2003, **31**, 4264–4274.
148. D. T. Brown, T. Izard and T. Misteli, *Nat. Struct. Mol. Biol.*, 2006, **13**, 250–255.
149. L. Fan and V. A. Roberts, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 8384–8389.
150. S. H. Syed, D. Goutte-Gattat, N. Becker, S. Meyer, M. S. Shukla, J. J. Hayes, R. Everaers, D. Angelov, J. Bednar and S. Dimitrov, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 9620–9625.
151. G. V. Pachov, R. R. Gabdoulline and R. C. Wade, *Nucleic Acids Res.*, 2011.
152. K. Tóth, N. Brun and J. Langowski, *Biochemistry*, 2001, **40**, 6921–6928.
153. J. F. Kepert, K. Fejes Tóth, M. Caudron, N. Mücke, J. Langowski and K. Rippe, *Biophys. J.*, 2003, **85**, 4012–4022.
154. J. Zlatanova, C. Seebart and M. Tomschik, *Trends. Biochem. Sci.*, 2008, **33**, 247–253.
155. J. Zlatanova, P. Caiafa and K. Van Holde, *FASEB J.*, 2000, **14**, 1697–1704.
156. S. McBryant, V. Adams and J. Hansen, *Chromosome Res.*, 2006, **14**, 39–51.
157. P. J. Verschure, I. van der Kraan, W. de Leeuw, J. van der Vlag, A. E. Carpenter, A. S. Belmont and R. van Driel, *Mol. Cell. Biol.*, 2005, **25**, 4552–4564.
158. C. Maison and G. Almouzni, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 296–304.
159. S. I. Grewal and S. Jia, *Nat. Rev. Genet.*, 2007, **8**, 35–46.
160. S. H. Kwon and J. L. Workman, *Bioessays*, 2011, **33**, 280–289.
161. S. A. Jacobs and S. Khorasanizadeh, *Science*, 2002, **295**, 2080–2083.
162. W. Fischle, Y. Wang, S. A. Jacobs, Y. Kim, C. D. Allis and S. Khorasanizadeh, *Genes Dev.*, 2003, **17**, 1870–1881.
163. P. R. Nielsen, D. Nietlispach, H. R. Mott, J. Callaghan, A. Bannister, T. Kouzarides, A. G. Murzin, N. V. Murzina and E. D. Laue, *Nature*, 2002, **416**, 103–107.
164. R. Sgarra, S. Zammitti, A. Lo Sardo, E. Maurizio, L. Arnoldo, S. Pegoraro, V. Giancotti and G. Manfioletti, *Biochim. Biophys. Acta*, 2010, **1799**, 37–47.
165. M. Stros, *Biochim. Biophys. Acta*, 2010, **1799**, 101–113.

166. Y. Postnikov and M. Bustin, *Biochim. Biophys. Acta*, 2009, **1799**, 62–68.
167. M. Rochman, C. Malicet and M. Bustin, *Biochim. Biophys. Acta*, 2010, **1799**, 86–92.
168. F. Catez and R. Hock, *Biochim. Biophys. Acta*, 2010, **1799**, 15–27.
169. N. L. Adkins and P. T. Georgel, *Biochem. Cell Biol.*, 2011, **89**, 1–11.
170. P. T. Georgel, R. A. Horowitz-Scherer, N. Adkins, C. L. Woodcock, P. A. Wade and J. C. Hansen, *J. Biol. Chem.*, 2003, **278**, 32181–32188.
171. T. Nikitina, X. Shi, R. P. Ghosh, R. A. Horowitz-Scherer, J. C. Hansen and C. L. Woodcock, *Mol. Cell. Biol.*, 2006, **27**, 864–877.
172. J. E. Phillips and V. G. Corces, *Cell*, 2009, **137**, 1194–1211.
173. L. Handoko, H. Xu, G. Li, C. Y. Ngan, E. Chew, M. Schnapp, C. W. Lee, C. Ye, J. L. Ping, F. Mulawadi, E. Wong, J. Sheng, Y. Zhang, T. Poh, C. S. Chan, G. Kunarso, A. Shahab, G. Bourque, V. Cacheux-Rataboul, W. K. Sung, Y. Ruan and C. L. Wei, *Nat. Genet.*, 2011, **43**, 630–638.
174. N. J. Francis, R. E. Kingston and C. L. Woodcock, *Science*, 2004, **306**, 1574–1577.
175. A. J. Wood, A. F. Severson and B. J. Meyer, *Nat. Rev. Genet.*, 2010, **11**, 391–404.
176. K. Nasmyth and C. H. Haering, *Annu. Rev. Genet.*, 2009, **43**, 525–558.
177. R. Stehr, PhD thesis, University of Heidelberg, 2010.
178. S. Mangenot, A. Leforestier, D. Durand and F. Livolant, *J. Mol. Biol.*, 2003, **333**, 907–916.
179. S. Mangenot, A. Leforestier, D. Durand and F. Livolant, *Biophys. J.*, 2003, **84**, 2570–2584.
180. S. Mangenot, A. Leforestier, P. Vachette, D. Durand and F. Livolant, *Biophys. J.*, 2002, **82**, 345–356.
181. M. Kruithof, F. Chien, M. de Jager and J. van Noort, *Biophys. J.*, 2008, **94**, 2343–2348.
182. Y. Cui and C. Bustamante, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 127–132.
183. M. Kruithof, F.-T. Chien, A. Routh, C. Logie, D. Rhodes and J. van Noort, *Nat. Struct. Mol. Biol.*, 2009, **16**, 534–540.
184. B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson and M. D. Wang, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 1960–1965.
185. B. Brower-Toland, D. A. Wacker, R. M. Fulbright, J. T. Lis, W. L. Kraus and M. D. Wang, *J. Mol. Biol.*, 2005, **346**, 135–146.
186. A. Bancaud, N. Conde e Silva, M. Barbi, G. Wagner, J.-F. Allemand, J. Mozziconacci, C. Lavelle, V. Croquette, J.-M. Victor, A. Prunell and J.-L. Viovy, *Nat. Struct. Mol. Biol.*, 2006, **13**, 444–450.
187. L. Jen-Jacobson, L. E. Engler and L. A. Jacobson, *Structure*, 2000, **8**, 1015–1023.
188. M. A. Hall, A. Shundrovsky, L. Bai, R. M. Fulbright, J. T. Lis and M. D. Wang, *Nat. Struct. Mol. Biol.*, 2009, **16**, 124–129.
189. B. Brower-Toland and M. D. Wang, *Methods Enzymol.*, 2004, **376**, 62–72.
190. S. Mihardja, A. J. Spakowitz, Y. Zhang and C. Bustamante, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 15871–15876.
191. J. D. Anderson and J. Widom, *J. Mol. Biol.*, 2000, **296**, 979–987.

192. C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev and C. Bustamante, *Science*, 2009, **325**, 626–628.
193. Y. J. Park, P. N. Dyer, D. J. Tremethick and K. Luger, *J. Biol. Chem.*, 2004, **279**, 24274–24282.
194. G. Li, M. Levitus, C. Bustamante and J. Widom, *Nat. Struct. Mol. Biol.*, 2005, **12**, 46–53.
195. M. Tomschik, H. Zheng, K. van Holde, J. Zlatanova and S. H. Leuba, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 3278–3283.
196. W. J. Koopmans, R. Buning, T. Schmidt and J. van Noort, *Biophys. J.*, 2009, **97**, 195–204.
197. M. L. Bennink, S. H. Leuba, G. H. Leno, J. Zlatanova, B. G. de Grooth and J. Greve, *Nat. Struct. Mol. Biol.*, 2001, **8**, 606–610.
198. L. H. Pope, M. L. Bennink, K. A. van Leijenhorst-Groener, D. Nikova, J. Greve and J. F. Marko, *Biophys. J.*, 2005, **88**, 3572–3583.
199. I. M. Kulić and H. Schiessel, *Phys. Rev. Lett.*, 2004, **92**, 228101.
200. M. Kruithof and J. van Noort, *Biophys. J.*, 2009, **96**, 3708–3715.
201. F. Battistini, C. A. Hunter, E. J. Gardiner and M. J. Packer, *J. Mol. Biol.*, 2010, **396**, 264–279.
202. T. J. Richmond and C. A. Davey, *Nature*, 2003, **423**, 145–150.
203. T. Wocjan, K. Klenin and J. Langowski, *J. Phys. Chem. B*, 2009, **113**, 2639–2646.
204. N. Korolev, O. V. Vorontsova and L. Nordenskiöld, *Prog. Biophys. Mol. Biol.*, 2007, **95**, 23–49.
205. C. L. Woodcock, S. A. Grigoryev, R. A. Horowitz and N. Whitaker, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 9021–9025.
206. J. R. Daban and A. Bermudez, *Biochemistry*, 1998, **37**, 4299–4304.
207. S. A. Grigoryev, *FEBS Lett.*, 2004, **564**, 4–8.
208. M. Engelhardt, *Nucleic Acids Res.*, 2007, **35**, e106.
209. T. Schalch, S. Duda, D. F. Sargent and T. J. Richmond, *Nature*, 2005, **436**, 138–141.
210. J. D. McGhee, J. M. Nickol, G. Felsenfeld and D. C. Rau, *Cell*, 1983, **33**, 831–841.
211. B. D. Athey, M. F. Smith, D. A. Rankert, S. P. Williams and J. P. Langmore, *J. Cell. Biol.*, 1990, **111**, 795–806.
212. 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.
213. C. L. Woodcock and S. Dimitrov, *Curr. Opin. Genet. Dev.*, 2001, **11**, 130–135.
214. D. Roccatano, A. Barthel and M. Zacharias, *Biopolymers*, 2007, **85**, 407–421.
215. J. Ruscio and A. Onufriev, *Biophys. J.*, 2006, **91**, 4121–4132.
216. L. Ehrlich, C. Munkel, G. Chirico and J. Langowski, *Comput. Appl. Biosci.*, 1997, **13**, 271–279.
217. R. Collepardo-Guevara and T. Schlick, *Biophys. J.*, 2011, **101**, 1670–1680.
218. M. Tark-Dame, R. van Driel and D. W. Heermann, *J. Cell Sci.*, 2011, **124**, 839–845.
219. G. Chirico and J. Langowski, *Biophys. J.*, 1996, **71**, 955–971.

220. K. Klenin, H. Merlitz and J. Langowski, *Biophys. J.*, 1998, **74**, 780–788.
221. C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann and R. Seidel, *Phys. Rev. Lett.*, 2010, **105**, 158101.
222. T. Schlick and O. Perisic, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10729–10737.
223. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, UK, 1987.
224. S. B. Smith, Y. Cui and C. Bustamante, *Science*, 1996, **271**, 795–799.
225. J. Schellman and D. Stigter, *Biopolymers*, 1977, **16**, 1415–1434.
226. D. Stigter, *Biopolymers*, 1977, **16**, 1435–1448.
227. S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock and T. Schlick, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 13317–13322.
228. A. J. Stone, *Mol. Phys.*, 1978, **36**, 241–256.
229. H. Zewdie, *Phys. Rev. E*, 1998, **57**, 1793–1805.
230. D. A. Beard and T. Schlick, *Biopolymers*, 2001, **58**, 106–115.
231. J. J. Freire and A. Horta, *J. Chem. Phys.*, 1976, **65**, 4049–4054.
232. N. Madras and A. D. Sokal, *J. Stat. Phys.*, 1988, **50**, 109–186.
233. P. H. Verdier and W. H. Stockmayer, *J. Chem. Phys.*, 1962, **36**, 227.
234. A. Baumgärtner and K. Binder, *J. Chem. Phys.*, 1979, **71**, 2541–2545.
235. A. V. Vologodskii and J. F. Marko, *Biophys. J.*, 1997, **73**, 123–132.
236. F. A. Escobedo and J. J. d. Pablo, *J. Chem. Phys.*, 1995, **102**, 2636–2652.
237. G. Arya and T. Schlick, *J. Chem. Phys.*, 2007, **126**, 044107.
238. H. G. Katzgraber, S. Trebst, D. A. Huse and M. Troyer, *J. Stat. Mech.*, 2006, **2006**, P03018.
239. K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.*, 1996, **65**, 1604–1608.
240. U. H. E. Hansmann, *Chem. Phys. Lett.*, 1997, **281**, 140–150.
241. C. L. White, R. K. Suto and K. Luger, *EMBO J.*, 2001, **20**, 5207–5218.
242. F. Livolant, S. Mangenot, A. Leforestier, A. Bertin, M. Frutos, E. Raspaud and D. Durand, *Philos. Transact. A Math. Phys. Eng. Sci.*, 2006, **364**, 2615–2633.
243. C. Woodcock, L. Frado and J. Rattner, *J. Cell Biol.*, 1984, **99**, 42–52.
244. J. G. Yodh, N. Woodbury, L. S. Shlyakhtenko, Y. L. Lyubchenko and D. Lohr, *Biochemistry*, 2002, **41**, 3565.
245. T. D. Frouws, H. G. Patterton and B. T. Sewell, *Biophys. J.*, 2009, **96**, 3363–3371.
246. A. Leforestier, J. Dubochet and F. Livolant, *Biophys. J.*, 2001, **81**, 2414–2421.
247. J. P. Langmore and J. R. Paulson, *J. Cell Biol.*, 1983, **96**, 1120–1131.
248. D. Yang and G. Arya, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2911–2921.
249. N. Korolev, A. P. Lyubartsev and L. Nordenskiold, *Biophys. J.*, 2006, **90**, 4305–4316.
250. S. Mangenot, E. Raspaud, C. Tribet, L. Belloni and F. Livolant, *Eur. Phys. J. E.*, 2002, **7**, 221–231.
251. H. R. Widlund, J. M. Vitolo, C. Thiriet and J. J. Hayes, *Biochemistry*, 2000, **39**, 3835–3841.
252. F. Mühlbacher, H. Schiessel and C. Holm, *Phys. Rev. E*, 2006, **74**, 031919.

253. M. J. Blacketer, S. J. Feely and M. A. Shogren-Knaak, *J. Biol. Chem.*, 2010, **285**, 34597–34607.
254. V. Makarov, S. Dimitrov, V. Smirnov and I. Pashev, *FEBS Lett.*, 1985, **181**, 357–361.
255. K. Rippe, *Curr. Opin. Genet. Dev.*, 2007, **17**, 373–380.
256. G. J. Filion, J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker and B. van Steensel, *Cell*, 2010, **143**, 212–224.
257. P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. P. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen and P. J. Park, *Nature*, 2011, **471**, 480–485.
258. J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein, *Nature*, 2011, **473**, 43–49.