From Macroscopic to Mesoscopic Models of Chromatin Folding

Tamar Schlick[§]

Department of Chemistry and Courant

Institute of Mathematical Sciences

New York University, 251 Mercer Street

New York, New York 10012

Running title: Mesoscopic Chromatin Model

Keywords: multiscale modeling, macroscopic, mesoscopic, chromatin folding, supercoiled DNA, histone tails, nucleosome, DNA folding

[§]To whom correspondence should be addressed (Phone:212-998-3116; e-mail: schlick@nyu.edu; fax: 212-995-4152)

May 30, 2008

Abstract

An overview of the evolution of computer models for simulation of chromatin folding is presented. Chromatin is the protein/nucleic acid fiber that stores the genetic material in higher organisms. Many biological questions concerning the fiber structure and its dependence on internal and external factors remain a puzzle. Modeling and simulation can in theory provide molecular view for analysis, but the sheer size and range of spatial and temporal scales involved require tailored multiscale models. Our first-generation, macroscopic models ignored histone tail flexibility but generated insights into preferred zigzag configurations and folding/unfolding dynamics at univalent salt. The second-generation mesoscale models incorporated histone tail flexibility, linker histones, and the presence of divalent ions. Recent results reveal the profound compaction induced by linker histones and the polymorphic fiber structure at divalent salt environments, with a small fraction of the linker DNAs bent rather than straight for ultimate compaction. Our chromatin model can be extended further to study many important questions dealing with histone tail post-translational modifications, the effects of variations in linker DNA length and of histone variants on chromatin structure, and the nature of higher-order fiber structures.

1. Introduction

One of the current challenges in scientific computing is model development that entails bridging the resolution among different spatial and temporal scales. In biological applications, a wide range of spatial scales defines systems, from the quantum particles to atomic, molecular, cellular, organ, system and genome entities. Temporal scales range from sub-femtosecond for electronic motion to billions of years of evolutionary changes. As our computing power and algorithms have improved, problems of greater scientific significance can be addressed with enhanced confidence and accuracy. However, developing appropriate molecular models and simulation algorithms to answer specific biological questions that require bridging all-atom details with the macroscopic view of activity on the cellular level remains an ad-hoc endeavor which requires as much art as science.

As a special volume of SIAM's journal on *Multiscale Modeling and Simulation* illustrated [1], multiscale biology is being developed by many varied techniques and applied to a variety of problems, such as involving protein and RNA three-dimensional structures, DNA supercoiling, ribosomal motions, DNA packaging in viruses, heart muscle motion, RNA translation, or fruit fly circadian rhythm. For these applications, techniques involve hierarchical methods that transform fast, low-resolution to slower, higher-resolution models; dynamics propagation using projection of standard molecular dynamics to longer timescales using master-equation methods; rigid-body dynamics; elastic or normal-mode models; and coarse-grained studies of slow, large-scale motions and features using differential equations for global properties or statistical methods.

This chapter describes the evolution of our group's models from macroscopic to mesoscopic scales developed to study chromatic folding; the focus is on presenting an overview of the models and results rather than simulation details which can be found in the individual papers. The chromatin fiber is the protein/DNA complex in eukaryotes that stores the genetic material [2]. With recent discoveries that point to a "second code" in DNA, a nucleosome-positioning code [3], understanding chromatin structure and dynamics becomes central, since transcription regulation and hence the most fundamental biological processes depend on chromatin architecture. This is because such fundamental cellular processes that are DNA-template directed require direct access to the DNA material, which must be unraveled from its compact folded state with the cellular protein matrix [2].

Following a brief introduction into the biology of chromatin, we describe development of physical models, energy functions, and simulation algorithms (Monte Carlo, Brownian Dynamics) for chromatin that gradually incorporated greater molecular complexity. We then mention how the models were validated against experiments and describe the biological findings concerning low-salt/high-salt unfolding/folding dynamics and the structural aspects of the chromatin fiber as a function of the ionic environment (ion charge and type) and the presence of linker histones, as well as the stabilizing role of the histone tails. We conclude by mentioning future applications and required model developments.

2. Chromatin Structure

Superimposed upon the canonical right-handed B-DNA helix is a left-handed

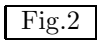
Fig.1

superhelical (or supercoiled) structure that facilitates fundamental template-directed biological process like transcription and replication. In higher organisms, this supercoiled DNA is wrapped around proteins, much like a long yarn around many spools (Fig. 1, taken from [4]). These proteins anchors consist of a core of 8 histone proteins – two copies each of H2A, H2B, H3, and H4 – whose flexible positively-charged tails extrude from the cylindrical-like core to help shield the negatively charged DNA chain. These proteins, along with the wrapped DNA, form the basic building block of chromatin, the nucleosome [5]. Additional, linker-histone proteins are needed for compacting the chromatin fiber into condensed states [6, 7, 8, 9, 10, 11]. The diameter of the chromatin fiber – 30nm – is widely quoted as the dimension at physiological ionic environments, but its detailed structure is unknown. However, this dimension represents only a DNA compaction factor of about 40 (Fig. 1). Very condensed states at certain stages of the cell cycle likely involve fiber/fiber interactions and further compaction by structural proteins [12, 13, 14, 15]. Thus, to pack the genomic material into the cell whose diameter of $\sim 5\mu$m is smaller by more than five orders of magnitude than the linear length of the DNA stored, a severe folding problem must be solved. This multiscale "DNA folding problem" is a challenge to both experimentalists and modelers because the resolution of the respective techniques is limited [16].

X-ray crystallography has produced very detailed structures of the nucleosome building block ($\sim$150bp of DNA wrapped around the histone octamer) [17, 5], including the recent tetranucleosome [18], while electron microscopy can provide macroscopic fiber views and dynamic techniques can measure polymer properties like diffusion properties or sedimentation coefficients [19]. Modeling on the level of the nucleosome is already too challenging for all-atom approaches while macroscopic elastic models for long supercoiled DNA (e.g., [20, 21]) are inappropriate for the molecular complexity of the fiber. Hence modeling fiber structure and motion requires specialized models that treat the system's electrostatics as accurately as possible – since these features are thought to be crucial for chromatin organization – while approximating others so as to make possible studies of nucleosome arrays (or oligonucleosomes) with sufficient configurational sampling (e.g., 100 million configurations) or simulation times of milliseconds and longer to be biologically relevant. In recent years, a variety of computational models have been developed (e.g., [22, 23, 24, 25, 26, 27, 28]).

A key question for investigation is what is the precise organization of the chromatin fiber at various physiological conditions (e.g., ionic concentrations). In particular, how do these structures depend on the presence of linker histones and variations in the linker DNA length? What is the role of the histone tails in fiber compaction? And what are the energetics of folding/unfolding?

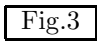
Already, many experimental and theoretical studies have proposed various structural possibilities (e.g., see [29]). Two broad classes are the *solenoid*, a



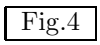
helical arrangement in which the linker DNA is bent (e.g., [9, 30]), and the *zigzag*, in which the linker DNA is mostly straight [18, 31]. Figure 2 illustrates some possible

chromatin architectures with associated internucleosome patterns. Besides the classic solenoid and 2-start zigzag models, an interdigitated solenoid [9] is also shown. Many structural studies converge upon the irregular two-start *zigzag* which brings each nucleosome into closest contact with its second nearest neighbor (i±2**).** *Solenoid*-like structures, however, have dominant interactions between each successive pair of nucleosome (i±1) and/or involving (i±5) or (i±6) [32] depending on the repeated patterns (see Fig. 2). Both monovalent and divalent ions as well as linker histones are known to affect these patterns significantly since, through electrostatic shielding, they can allow fiber segments to come into closer contact. In addition, it is likely that fiber structure is polymorphic [33, 34], so that several morphologies exist and interchange depending on detailed conditions in the cell. Chromatin modeling and simulation, thus, has the potential to add important insights into these questions and provide detailed structures for analysis.

3. The First-Generation Macroscopic Chromatin Models and Results

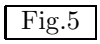


Our first-generation "macroscopic" models treated the nucleosome and the wound DNA according to general mechanical and electrostatic properties, with the histone tails approximated as rigid bodies and linker histones neglected (Fig. 3) [24, 35, 36, 37, 38].

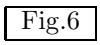


As Fig. 3 shows, the nucleosome core is represented as a large regular disk and a short slender disk for part of the H3 histone tail resolved in the 1997 crystal structure [17] with discrete charges determined to approximate the Poisson-Boltzmann solution of the electric field using our program DiSCO (Discrete Surface Charge Optimization) (see Fig.4) [35]; the minimization is performed efficiently using our truncated Newton code

TNPACK [39]. The linker DNA is denoted by charged beads using the well-known wormlike/chain model for supercoiled DNA developed by Allison and coworkers (as applied in [20]) using the Stigter charged cylinder electrostatics formulation. For dynamics, variables r_i are defined on each core and each linker DNA bead, each associated with a local Euler coordinate frame (Fig. 5).

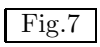


The advantage of this model is its relative simplicity and the fact that most associated energy parameters are taken directly from experiment. Fig. 5 shows typical stretching, bending, twisting, electrostatic Debye-Hückel, and excluded volume terms used for the model. For Brownian dynamics, we use complete hydrodynamics to treat both the translation and rotation of nucleosome cores and linker DNA beads (Fig. 5, right), making possible nanosecond to microsecond simulations to capture folding/unfolding events of short polymers.



Specifically, translational diffusion constants (D_t) were verified against experimental values for dimers and trimers, with a gentle linker-DNA bending noted for the latter, explaining the experimentally-noted sharper increase of D_t with the salt concentration [24]. The accordion-like opening at low salt of a polynucleosome can already be captured with this simple model (Fig. 6). Moreover, a 30nm helical *zigzag*

naturally emerges from the trimer geometry with a packing ratio of 4 nucleosomes per 11nm (Fig. 6).



When the new crystal structure with full tails was resolved crystallographically in 2002 [40], we extended the DiSCO optimization approach to handle irregular surfaces (Fig. 7) [37]. With this new model, the salt-dependent folding of nucleosomal arrays was captured, along with explanations for the energetics involved (Fig. 7). Namely, at low

salt, linker DNA repulsion triggers array unfolding while, at high salt, internucleosomal attraction from histone tail regions of the core triggers folding, with the H3 tails dominating internucleosomal attraction.

The importance of these histone tail interactions signaled to us that histone tails flexibility must be incorporated for further resolution and studies of chromatin structure and dynamics. Indeed, the histone tails are known to be crucial for compacting chromatin and, via biochemical modifications such as acetylation, methylation, and phosphorylation, can affect signaling pathways and transcriptional regulation. Detailed structures and physical insights are needed to interpret such observations.

4. The Second-Generation Mesoscopic Chromatin Model and Results

To incorporate histone tail flexibility as well as linker histones, we applied coarse graining to add components compatible with the nucleosome and linker DNA units (Fig. 8) [41]. Namely, starting from the amino-acid/subunit model

Fig.8

of Warshel and Levitt [42], for each histone tail (where each residue is a bead), we simulated Brownian dynamics of the tails and further coarse-grained the polymers to obtain parameterized protein beads (charges, excluded volume, harmonic stretching and bending) that reproduced configurational properties of the subunit model.

To model the linker histone, we use the model of the rat H1d linker histone [43, 44] and represent the globular domain (76 residues) by one bead and the C-terminal domain (110 residues) by two beads, rigidly attached to the nucleosome core along the dyad as suggested experimentally (Fig. 8); we neglect the shorter (33-residue) N-terminal domain for simplicity.

The resulting oligonucleosome simulation model is extended accordingly, so that each independent variable (one per core, each linker DNA, linker histone, and histone tail bead) is associated with an Euler coordinate frame.

With this more complex, "mesoscopic" model, Brownian dynamics simulations were too computationally intensive. We therefore sampled oligonucleosome configuration space by Monte Carlo simulations with tailored local moves for translation and rotation of nucleosome core and linker DNA, global pivots of the polymer, and histone tail-regrowth moves; a new, end-transfer configurational bias method was developed to regrow tails at apposite ends so as to scale quadratically with polymer size rather than exponentially as in traditional configurational-bias MC [45].

Fig.9

The new chromatin model was validated against experiments as well as against the rigid-histone-tail model for configuration-dependent measurements such as sedimentation coefficients, $S_{20,W}$ (which measures how fast a polymer sediments in a fluid), translational diffusion constants (D_t), and tail-to-tail mononucleosome diameters (D_{max}) (see Fig. 9) [41]. Not only did the refined model results compare well with experiment; results for the flexible-tail model improved upon results for the earlier, rigid-tail model.

Fig.10

Oligonucleosome dynamics (Fig. 10) reveal the dynamic nature of the histone tails, the strong repulsion among linker DNA at low salt, and the significant internucleosomal attraction mediated by the histone tails at high salt conditions. Note that *zigzag* topologies dominate at monovalent ion conditions, with fiber axis bending leading to fiber/fiber cross interactions (Fig. 10). In fact, we found that each histone tail has a unique role in fiber organization

because of its size and location [28]: the H4 tails mediate the strongest internucleosomal interactions; the H3 tails bind strongly to parental DNA to screen the negatively-charged linker DNAs; and the H2A and H2B tails, located at the periphery of the nucleosome, mediate the most fiber/fiber interactions.

The most interesting aspect of our recent studies involves the dramatic effect of linker histone and divalent ions on fiber organization. The latter was modeled as a first approximation by setting the Debye parameter κ to the inverse DNA diameter $\kappa = 1/(2.5)$ nm, to allow the DNA beads to almost touch one another. Moreover, the linker DNA bending persistence length was reduced from 50 to 30nm to mimick magnesium effects [46].

Fig.11

As Fig. 11 shows [47], without linker histones and a monovalent concentration of $c_s = 0.15$M, the fiber organizes as a classic open *zigzag* with a sedimentation coefficient $S_{20,W}$ of 39±1.2S (compared to 37S experimentally) and a packing ratio of 4 nucleosomes per 11nm. When the linker histone is added, sandwiching the entering and exiting linker DNA from each nucleosome, a rigid stem forms to allow closer contact; the fiber is markedly more compact ($S_{20,W}$=49±1.7S, compared to 55.6S experimentally, and packing ratio of 7 nucleosomes per 11nm). The linker DNAs remain relatively straight as in the classic *zigzag* (see below). When divalent counter ions are incorporated (to mimick 1mM Mg^{2+}), further compaction follows: $S_{20,W}$=54±2.2S (compared to 60S experimentally, and 8 nucleosomes per 11nm). Moreover, this further compaction is made possible by a tendency to bend a small proportion (e.g., 15%) of the linker DNAs rather than maintain all straight orientations. The dynamics associated with these fibers at these different conditions (Fig.10) emphasize the greater flexibility and open structure without linker histones versus the more rigid and compact structures with linker histones.

Fig.12

This surprising result, of a compact fiber architecture at divalent ionic condition involving both straight and bent linker DNAs, was analyzed further using internucleosomal interaction patterns (Fig. 12). The entries of these interaction matrices, I'(i,j) measure the intensity of tail-mediated interactions between nucleosomes; i and j (i.e., fraction of configurations over a 200-million MC ensemble where the two nucleosome cores are within 80% of the van der Waals radii). The one-dimensional plot I(k) decomposes the interactions (i, i±k) for k=1,2,3....

Fig. 12 shows that (i±2) and (i±3) interactions dominate without linker histone at

monovalent salt (left column), reflecting *zigzag* configuration with some third-neighbor interactions. With the linker histone, (i±2) interactions domi-

nate, producing the classic *zigzag* architecture (compare to Fig.2). In this structure, the nucleosome/nucleosome distances (d_{12}) and triplet angles (θ) formed by 3 consecutive nucleosome narrows considerably (Fig. 12, bottom). When, in addition, divalent ions are included (right column), (i$\pm$1) interactions indicative of *solenoid* configurations appear, with a bimodal θ distribution reflecting a mixture of mostly straight and some bent linker DNA.

This intriguing suggestion for a hybrid compact fiber in divalent ion conditions was recently verified by a new experimental technique termed EMANIC which uses formaldehyde cross-linking followed by unfolding and EM visualization to capture internucleosomal patterns [47]. Such an ensemble of interchanging configurations with straight and bent linker DNAs is energetically advantageous since linker DNA bending can minimize repulsion at the fiber axis. Moreover, it merges both classic models of *zigzag* and *solenoid* for optimal compaction!

5. Future Perspective

Elucidating the structural details of the chromatin fiber and its dynamics at various physiological conditions and chemical compositions will undoubtedly remain an ongoing challenge for both modelers and experimental scientists. This is because of the magnitude of the folding problem – from the open "beads-on-a-string" – like model at low salt of width 10 nm to very condensed states of mitotic chromosomes at "silent" stages of the cell cycle [2]. In addition, chromatin structure varies depending on the presence of specific linker histones (e.g., H1, H5), the length of the linker DNA (which roughly varies from 20 to 80bp), the histone proteins themselves (which have well known variants), and the modified states of the tails. Four specific problems for modeling involving these components are described below, along with associated modeling development issues.

Linker DNA Length Effects. Though many experiments suggest typical two-start helix models for fiber organization based on *zigzag* arrangement of nucleosome as revealed by electron microscopy, *in vivo* studies of linker-length variations have not been systematic. Recently, Robinson et al. [9] have prepared in vitro assays of oligonucleosome with linker lengths varying from 10 to 70 bp. Surprisingly, they found two distinct fiber structures: a diameter of 33nm and repeat pattern of 11 nucleosome per 11nm for 10-40 bp linker lengths, and a diameter of 44nm and 15 nucleosomes per 11nm for the longer linker lengths (50-70bp). Moreover, based on model building, they argue that a one-start helical model with nucleosome inter-digitation and bent linker DNA explains these observations. However, this solenoid form differs from the favored zigzag configuration observed widely by X-ray crystallography [18].

These contradictory findings are puzzling, particularly since Robinson et al. [9] suggest that the linker histones alter the chromatin geometry to reconcile the findings. But our simulations and the hand-in-hand experiments by the Grigoryev group [47] show that linker histones favor the *zigzag* structure and that instead divalent ions lead to some bending in linker lengths (Fig. 11). Thus, simulations with varied linker lengths are needed to reconcile these observations and provide atomic views and energetic estimates to interpret the linker-length

effects. This application of linker-length variations is easily performed with our current mesoscale model of the chromatin fiber.

Histone Variant Effects. The naturally occurring histone variant H2A.Z of H2A has been associated with assembly of specialized compacted forms of chromatin [4]. Remarkably, H2A.Z can alter the equilibrium dynamics of the fiber compared to H2A. It has been hypothesized that electrostatics differences between the two proteins are responsible for these profoundly different structural effects, and that different residues on H2A/H2A.Z interact differently with the H4 tails to impart these global effects. Other experiments reveal that a different variant, namely H2.Bbd has an opposite effect, promoting unfolding of condensed chromatin [48]. These problems form excellent questions for modeling. However, successful studies will likely require model enhancements since a 20-40% sequence homology difference of the variants compared to wildtype H2A may require more sensitive accounts than our current united-residue model. In fact, atomic-level modeling of the various histones H2A may be required to obtain potentials of mean force for each one and then incorporate these averaged force fields into the mesoscale oligonucleosome model. With this modification, varying the concentration of H2A and its variants (e.g., 60% H2A, 40% H2A.Z) and their distribution along the polymer chain is readily amenable for simulation.

Histone Tail Modifications. The histone-code hypothesis, in which different histone proteins are chemically modified to regulate transcription [49], is one of the fundamental tenets in modern biology. Because the N-termini of the histone tails are flexible and disordered in the crystal state, post-translational modifications can trigger fiber changes and specific biological functions. However, how these chemical modifications contribute to fiber packaging and folding/unfolding events is not well understood. For example, acetylation (adding $-COCH_3$ units to arginine and lysine residues) can prevent condensation at high salt (e.g., when applied to lysine 16 of H4), while methylation (adding $-CH_3$ moieties to arginine and lysine residues) can cause folding when applied to lysine 9 of H3 but unfolding when applied to lysine 4 of H3.

In theory, modeling and simulation can systematically test these changes to interpret the structural effects. However, as above, a detailed residue model with accurate electrostatics may be needed on a fine level and integrated with a coarser model of oligonucleosome chains. A finer level of modeling would also require more careful parameterization of the non-electrostatic terms.

Higher-Order Chromatin Organization. Silent phases of the cell cycle are associated with heterochromatin – a highly condensed, highly ordered state of nucleosomal arrays, possibly compacted through clusters of repetitive DNA elements found at centromeres and telomeres [50]. Understanding heterochromatin formation and transformations is thought to be key to deciphering epigenetic control of the genome. Crucial to formation of heterochromatin are the binding of many chromatin-regulating (including linker) proteins as well as histone tail modifications, two elements of which may act in concert. From studies of mammalian nuclei, fibers of diameter $\sim$500-700nm have been suggested. This high level of compaction must involve a decrease of the vacant spaces in the

∼30nm fiber through inter-digitation of nucleosomes, a generalization of single fiber inter-digitation, as shown in Fig. 2. Fiber cross-linking must be enhanced by linker histones, chromatin binding proteins like MeCP2 [51], and divalent ions. Modeling can in theory explore these potential inter-digitated fiber states and study the dependencies on the above factors.

To a first approximation the mesoscale model described here could be expanded to account for multiple oligonucleosome chains with attractive electrostatic interactions that mimick bridging effects of protein cofactors, linker histones, and divalent ions. Such studies could examine the feasibility of inter-digitation of various solenoid, zigzag, or hybrid fiber models to exclude certain geometric possibilities and suggest favorable topologies. More detailed studies would require construction de novo of a new bio-molecular network model appropriate for such complex investigations of multiple fiber/protein interactions.

6. Conclusions

As demonstrated in this overview, understanding chromatin structure, function, and dynamics is a challenging enterprise amenable to innovative multiscale models. As new developments in biology, modeling, and scientific computing are made available, more intricate models can be designed with more challenging questions addressed. In the not-too-distant future, I envision an all-atom simulation of an oligonucleosome (millions of atoms) and, within a few decades, a realistic simulation of the entire cell cycle of chromatin organization and associated epigenetic control! Biologists, chemists, and mathematics/physical scientists should continue to combine their expertise to achieve these exciting goals.

Acknowledgements

This work was supported by NSF award MCB-0316771 and NIH award R01 GM55164 to T. Schlick. Acknowledgment is also made to the donors of the American Chemical Society Petroleum Research Fund for support of this research (Award PRF39225-AC4 to T. Schlick) and to Philip Morris USA and Philip Morris International. I thank Dr. Sergei Grigoryev for valuable discussions, Yoon Ha Cha and Trang Nguyen for technical assistance with manuscript preparation, and Sean McGuffee for preparing three figures.

Figure captions

Figure 1: Eukaryotic DNA in the cell (from [4]). In eukaryotes, the DNA is organized in the chromatin fiber, a complex made of DNA wrapped around core of 8 proteins, 2 copies each of H2A, H2B, H3, H4. The chromatin building block is the nucleosome. At low salt, the fiber forms "beads-on-a-string" like models and, at physiological salt concentrations, it compacts into the 30nm fiber, whose detailed structure is unknown. Much more condensed fiber structures are formed during transcription-silent phases of the cell cycle. In each level of organization shown, the pink unit represents the system from the prior level, to emphasize the compaction ratio involved. Note also that at lengths much less than the DNA persistence length (which is ∼150 bp), the DNA is relatively straight but, at much longer lengths, the DNA is very flexible.

Figure 2: Three hypothesized chromatin arrays are shown from the top and side with black nucleosome cores. In the classical zigzag model [18], DNA is shown in orange and blue, and nucleosomes i interact with i $\pm$ 2. In the classical

solenoid model [30], DNA is shown in pink and blue, and nucleosomes i interact with i $\pm$ 1 and i $\pm$ 6. In the interdigitated solenoid model [9], DNA is shown in pink and blue, and nucleosomes i interact on the flat sides with i $\pm$ 5, i $\pm$ 6, and on the narrow edges at i $\pm$ 11. Yellow arrows point to visible nucleosomes, and grey arrows point behind the chromatin to hidden nucleosomes.

Images at bottom were reprinted with permission from: (left) Macmillan Publishers Ltd. (Fig.3 of T. Schlach et al., *Nature* **436**(7047): 138-141, 2005, license 1943250688673); (middle) Elsevier (Fig.5 of J.D. McGhee et al., *Cell* **33**(3): 831-841, 1983, license 1982084); and (right) National Academy of Sciences (Fig.3 of P.J. Robinson et al., *PNAS USA* **103**(17): 6506-6511, 2006, copyright 2006 National Academy of Sciences U.S.A).

Figure 3: First-generation chromatin model. The nucleosome core (left) is represented by charged bodies denoting the nucleosomes with linker DNA represented as connecting beads (right).

Figure 4: DiSCO optimization for the nucleosome core. To optimize the charges on the surface representing the histone octamer with wrapped DNA, we distribute charges homogeneously on the surface and choose charges to approximate the electric field of the atomistic core by a Debye-Hückel approximation in a region V', separated from the surface by a distance d, where the Debye-Hückel (linear) approximation is valid. Optimization is performed efficiently using our truncated Newton method TNPACK [39]. We found that 277 particles yielded an error of <10% are a large range of monovalent salt [35].

Figure 5: Model for chromatin dynamics and energetics. On top, the basic unit along with position vector and coordinates for core and linker DNA beads are shown [24]. On bottom, the governing energies and Brownian dynamics protocol are described.

Figure 6: Nucleosome array trajectory snapshots from 10ns Brownian dynamics simulations with the macroscopic model for dinucleosome, trinucleosome, and 12-nucleosome systems [35]. The linker DNA from the wormlike chain model is shown in red, and the nucleosomes are the cylindrically shaped objects with electrostatic charges colored on a scale of red as negative to blue as positive. The dinucleosome and trinucleosome systems reflect compact structures under high univalent salt concentrations (50mM). The 12-nucleosome array expands as the negatively charged linker DNAs repel one another at low univalent salt concentrations (10mM). At the bottom, the 30nm fiber consisting of 48-nucleosomes was refined using Monte Carlo methods from a solenoid starting structure that was constructed using the dinucleosome folding motif.

Figure 7: The irregular DiSCO algorithm and associated results. The algorithms create designed to create an irregular surface for the nucleosome so that tail geometries (though rigid) are captured (top) [37]. This model captured the salt-dependent folding and unfolding of chromatin (bottom) at different univalent salt environment [38].

Figure 8: Mesoscopic oligonucleosome model. The irregular DiSCO model for the nucleosome without tails is combined with a coarse-grained representation of the histone tails and the linker histones [41].

Figure 9: Experimental validation of flexible-tail model for sedimentation

coefficients, diffusion constants, and mono nucleosome extension [41]. Results show good agreement with experiment and better reproduction of the data composed to the rigid-tail model.

Figure 10: Nucleosome arrays with flexible histone tails simulated using the mesoscale chromatin model by Monte Carlo. The tails are colored as: H2A—yellow, H2B—red, H3—blue, and H4—green. The leftmost two columns illustrate behavior of 12-nucleosome arrays under low (10mM) and high (200mM) univalent salt concentrations. At low salt, the negatively charged linker DNAs (red) repel one another and expand the array, while at high salt, the chromatin fiber condenses. The other columns show the behavior of 48-nucleosome arrays at high (200mM) monovalent salt concentrations with linker DNA colored grey and nucleosome cores colored black at three compositions: without linker histones without magnesium (–LH–Mg); with linker histone H1 (cyan) without magnesium(+LH–Mg) ; and with linker histone H1with magnesium (+LH+Mg). Note the compaction introduced by linker histone as well as divalent ions.

Figure 11: Representative fiber architectures without linker histone, without magnesium ions (top); with linker histone but without magnesium ions (middle); and with both linker histones and divalent ions (bottom). The histone tail analyses at bottom reveal the important role of the tails in stabilizing structures, especially when linker histone is present.

Figure 12: Internucleosomal nteraction patterns (averaged over 100,000 configurations from 200 million MC steps) for the three systems as in Fig. 11 (LH=linker histone, Mg=magnesium ions). The pairwise pattern show the zigzag preference of the fiber without magnesium ions, and the introduction of solenoid-like (bent-DNA) features when divalent ions are introduced. The fiber compaction trends are also evident from analyses of the respective internucleosome distances, triplet and quadroplet angles for successive nucleosomes [47].

References

[1] Schlick, T. and Dill, K., Editors. 2006. Multiscale modeling in biology. *Special issue of SIAM J. Mult. Mod. and Sim.* **5**(4):1174–1366.

[2] Felsenfeld, G. and Groudine, M. 2003. Controlling the double helix. *Nature* **421**:448–453.

[3] Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, IK., Wang, JP. and Widom, J. 2006. A genomic code for nucleosome positioning. *Nature* **442**(7104):772–778.

[4] Schlick, T. 2002. *Molecular Modeling: An Interdisciplinary Guide*, Springer-Verlag, NY.

[5] Richmond, T. J. and Davey, C. A. 2003. The structure of DNA in the nucleosome core. *Nature* **423**(6396):145–150.

[6] Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J. and Woodcock, C. L. 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. USA* **95**(24):14173–14178.

[7] Woodcock, C. L. and Dimitrov, S. 2001. Higher-order structure of chromatin and chromosomes. *Curr. Opin. Gen. Dev.* **11**(2):130–135.

[8] Huynh, V. A., Robinson, P. J. and Rhodes, D. 2005. A method for the in vitro reconstitution of a defined "30nm" chromatin fiber containing stoichiometric amounts of the linker histone. *J. Mol. Biol.* **345**(5):957–968.

[9] Robinson, P. J., Fairall, L., Huynh, V. A. and Rhodes, D. 2006. EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. USA* **103**(17):6506–6511.

[10] Grigoryev S. A. 2004. Keeping fingers crossed: heterochromatin spreading through interdigitation of nucleosome arrays. *FEBS Lett.* **564**(1-2):4–8.

[11] Widom, J. and Klug, A. 1985. Structure of the 300A chromatin filament – X-ray diffraction from oriented samples. *Cell* **43**(1):207–213.

[12] Bystricky, K., Heun, P., Gehlen, L., Langowski, J. and Gasser, S. M. 2004. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. USA* **101**(47):16495–16500.

[13] Caravaca, J. M., Cano, S., Gallego, I. and Daban, J. R. 2005. Structural elements of bulk chromatin within metaphase chromosomes. *Chrom. Res.* **13**(7):725–743.

[14] Luger, K. and Hansen, J. C. 2005. Nucleosome and chromatin fiber dynamics. *Curr. Opin. Struct. Biol.* **15**(2):188–196.

[15] Tremethick, D. J. 2007. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**(4):651–654.

[16] van Holde, K. and Zlatanova, J. 2007. Chromatin fiber structure: Where is the problem now? *Semin. Cell Dev. Biol.* **18**(5):651–658.

[17] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 Angstrom resolution. *Nature* **389**(6648):251–260.

[18] Schalch, T., Duda, S., Sargent, D. F. and Richmond, T. J. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fiber. *Nature* **436**(7047):138–141.

[19] Hansen, J. C. 2002. Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms and functions. *Ann. Rev. Biophys. Biomol. Str.* **31**:361–392.

[20] Jian, H., Vologodskii, A. and Schlick, T. 1997. A combined wormlike chain and bead model for dynamic simulations of long DNA. *J. Comp. Phys.* **136**:168–179.

[21] Schlick, T. Olson, W. K. 1992. Computer simulations of supercoiled DNA energetics and dynamics. *J. Mol. Biol.* **223**:1089–1119.

[22] Leuba, S. H,. Yang, G. L., Robert, C., Samori, B., van Holde, K., Zlatanova, J. and Bustamante, C. 1994. 3-Dimensional Structure of Extended Chromatin Fibers as Revealed by Tapping-Mode Scanning Force Microscopy. *Proc. Natl. Acad. Sci. USA* **91**(24):11621–11625.

[23] Katritch, V., Bustamante, C. and Olson, V. K. 2000. Pulling chromatin fibers: Computer simulations of direct physical micromanipulations. *J. Mol.*

Biol. **295**(1):29–40.

[24] Beard, D.A. and Schlick, T. 2001. Computational modeling predicts the structure and dynamics of chromain fiber. *Structure* **9**(2):105–114.

[25] Wedemann, G. and Langowski, J. 2002. Computer simulation of the 30-nanometer chromatin fiber. *Biophys. J.* **82**(6):2847–2859.

[26] Sun, J., Zhang, Q. and Schlick, T. 2005. Electrostatic mechanism of nucleosomal array folding revealed by computer simulation. *Proc. Natl. Acad. Sci. USA* **102**(23):8180-8185.

[27] Mozziconacci, J., Lavelle, C., Barbi, M., Lesne, A. and Victor, J. M. 2006. A physical model for the condensation and decondensation of eukaryotic chromosomes. *Febs Lett.* **580**(2):368-372.

[28] Arya, G. and Schlick, T. 2006. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. USA.* **103**:16236–1624.

[29] Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R. and Richmond, T. J. 2004. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **306**(5701):1571-1573.

[30] McGhee, J. D., Nickol, J. M., Felsenfeld, G. and Rau, D. C. 1983. Higher order structure of chromatin: Orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length. *Cell* **33**(3):831–841.

[31] Woodcock, C. L., L. L. Frado and J.B. Rattner. 1984. The higher-order structure of chromatin – evidence for a helical Ribbon Arrangement. *J. Cell. Biol.* **99**(1):42-52.

[32] Thoma, F., Koller, T. and Klug, A. 1979. Involvement of histone-H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin." *J. Cell Biol.* **83**(2):403-427.

[33] Wong, H., Victor, J. -M. and Mozziconacci, J. 2007. An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. *PLoS ONE* **2**, e877:1-8.

[34] Gordon, F., Luger, K. and Hansen, J.C. 2005. The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. *J. Biol. Chem.* **280**:33701-3706.

[35] Beard, D. A. and Schlick, T. 2001. Modeling salt-mediated electrostatics of macromolecules: The algorithm DiSCO (Discrete Surface Charge Optimization) and its application to the nucleosome, *Biopolymers* **58**:106–115.

[36] Beard, D. A. and Schlick, T. 2003. Unbiased rotational moves for rigid-body dynamics. *Biophys. J.* **85**:2973–2976.

[37] Zhang, Q., Beard, D. A. and Schlick, T. 2003. Constructing irregular surfaces to enclose macromolecular complexes for mesoscale modeling using the discrete surface charge optimization (disco) algorithm. *J. Comp. Chem.* **24**:2063–2074.

[38] Sun, J., Zhang, Q. and Schlick, T. 2005. Electrostatic mechanism of nucleosomal array folding revealed by computer simulation. *Proc. Natl. Acad. Sci. USA* **102**(23): 8180–8185.

[39] Schlick, T. and Fogelson, A. 1992. TNPACK – a truncated Newton minimization package for large-scale problems: I. algorithm and usage. *ACM Trans. Math. Softw.* **18**:46–70.

[40] Davey, C.A., Sargent, D.F., Luger, K., Mader, A.W. and Richmond, T.J. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 ? resolution. *J. Mol. Biol.* **391**:1097–113.

[41] Arya, G., Zhang, Q. and Schlick, T. 2006. Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys. J.* **91**(1):133–150.

[42] Warshel, A. and Levitt, M. 1975. Computer simulation of protein folding. *Nature* **253**(5494):694–698.

[43] Bharath, M. M., Chandra, N. R., and Rao, M. R. 2002. Prediction of an HMG-box fold in the C-terminal domain of histone H1: insights into its role in DNA condensation. *Proteins* **49**:71–81.

[44] Bharath, M. M. S., Chandra, N. R. and Rao, M. R. S. 2003. Molecular modeling of the chromatosome particle. *Nuc. Acids Res.* **31**:4264–4274.

[45] Arya, G. and Schlick, T. 2007. Efficient global biopolymer sampling with end-transfer configurational bias Monte Carlo. *J. Chem. Phys.* **126**(4):044107.

[46] Baumann, C. G., Smith, S. B., Bloomfield, V. A. and Bustamante, C. 1997. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. USA* **94**:6185–6190.

[47] Grigoryev, S. A., Arya, G., Correl, S., Woodcock, C. and Schlick, T. 2008. Nucleosome packing and interactions in higher-order chromatin fibers. In Preparation

[48] Fan, J., Rangasamy, D., Luger, K. and Tremethick, D. 2004. H2A.Z alters the nucleosome surface to promote HP1 alpha-mediated chromatin fiber folding. *Mol. Cell* **16**(4):655–661.

[49] Strahl, B. D. and Allis, C. D. 2000. The language of covalent histone modifications. *Nature* **403**:41–45.

[50] Shiv, G. and Jia, S. 2007. Heterochromatin revisited. *Nat. Rev. Gen.* **8**:35–46.

[51] McBryant, S. M., Adam, V. and Hansen, J. C. 2006. Chromatin architectural proteins. *Chrom.e Res.* **14**:39–51.

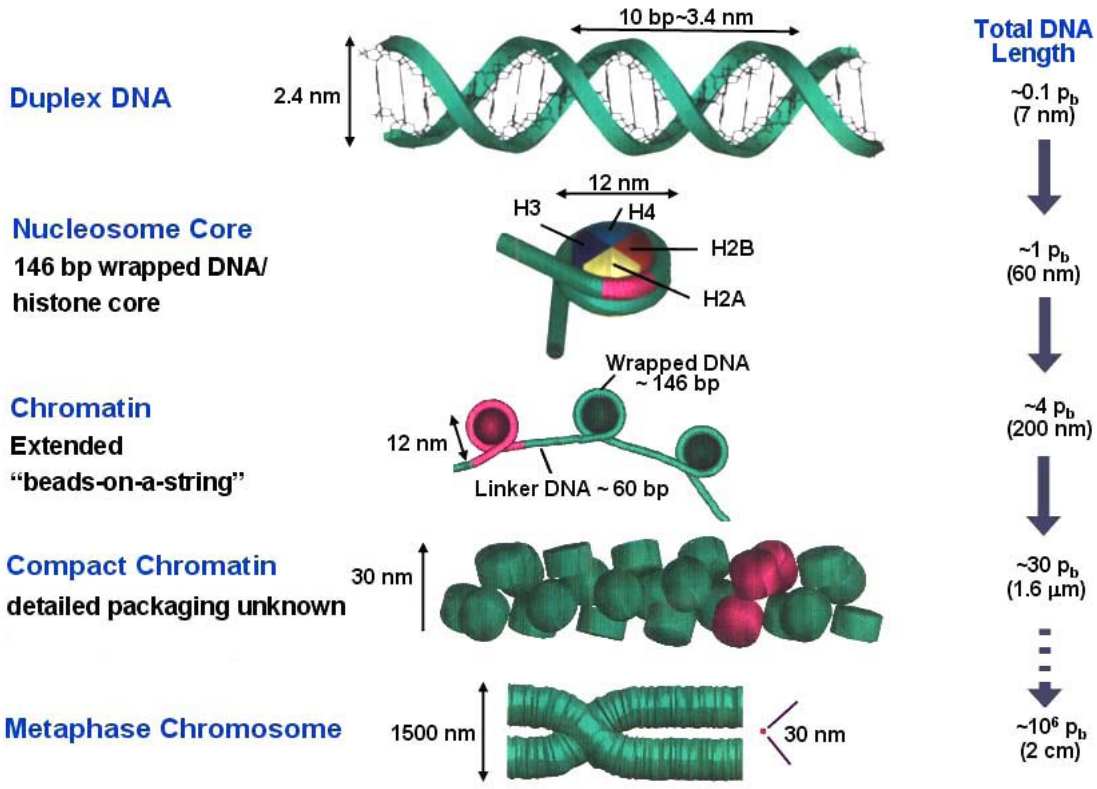

Duplex DNA
10 bp~3.4 nm
2.4 nm
Nucleosome Core
146 bp wrapped DNA/
histone core
12 nm
H3
H4
H2B
H2A
Chromatin
Extended
"beads-on-a-string"
Wrapped DNA
~146 bp
12 nm
Linker DNA ~60 bp
Compact Chromatin
detailed packaging unknown
30 nm
Metaphase Chromosome
1500 nm
30 nm
Total DNA Length
~0.1 pb
(7 nm)
~1 pb
(60 nm)
~4 pb
(200 nm)
~30 pb
(1.6 μm)
~10^6 pb
(2 cm)

Figure 1

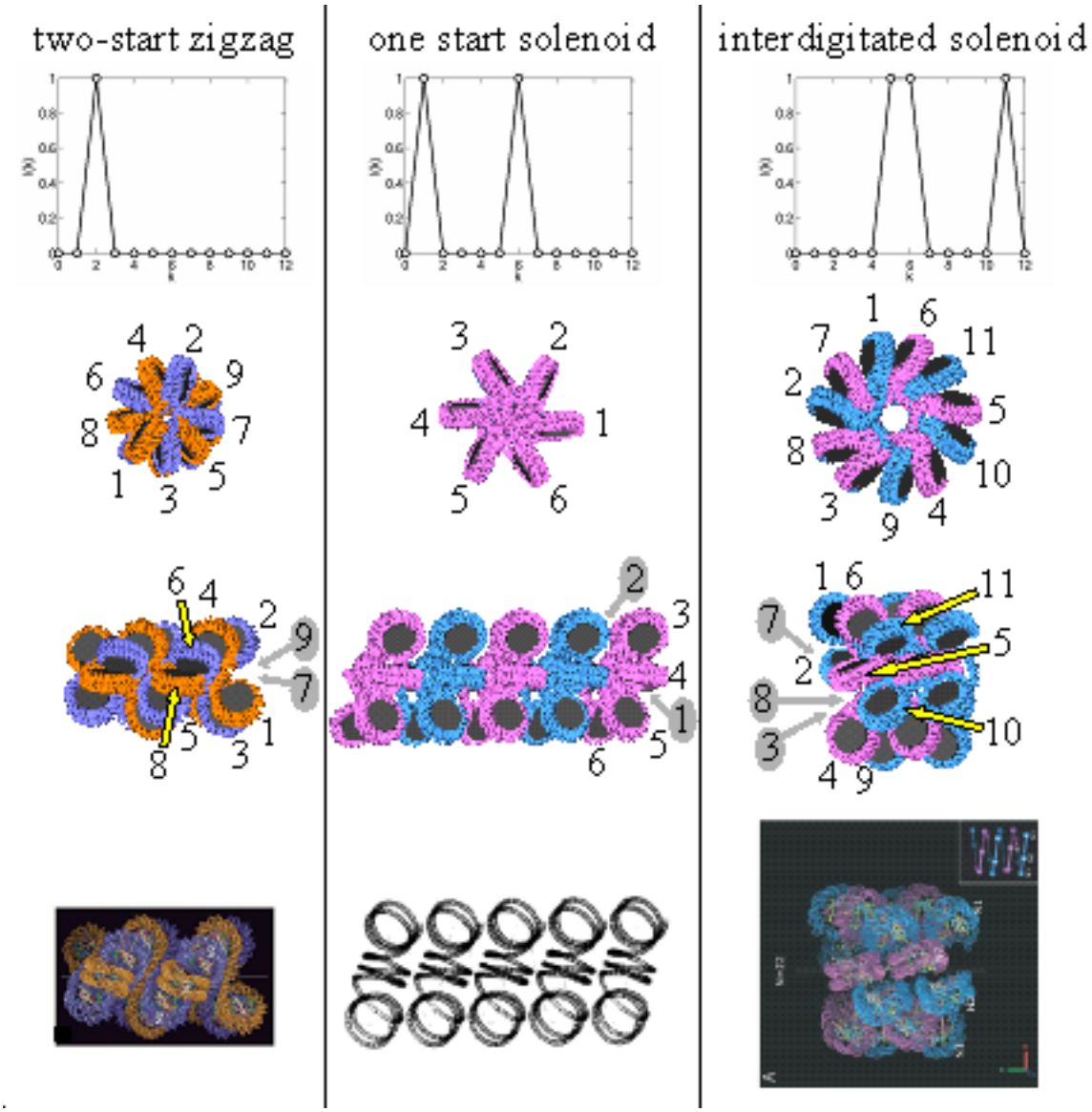

two-start zigzag
one start solenoid
interdigitated solenoid

Figure 2

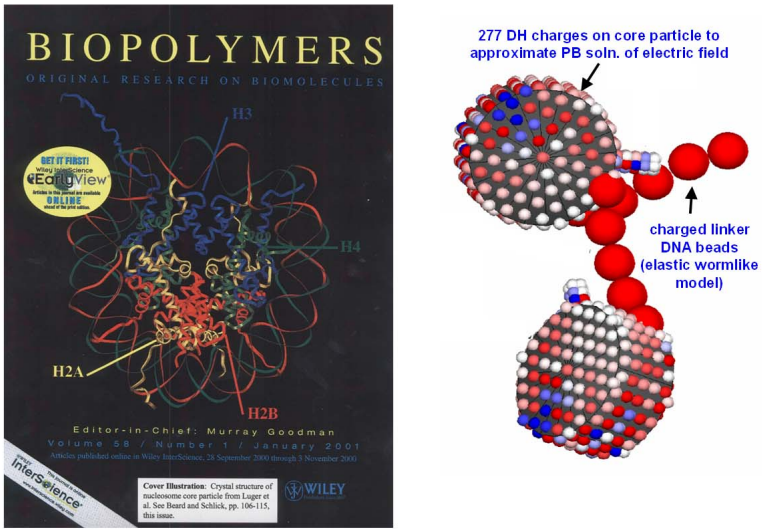
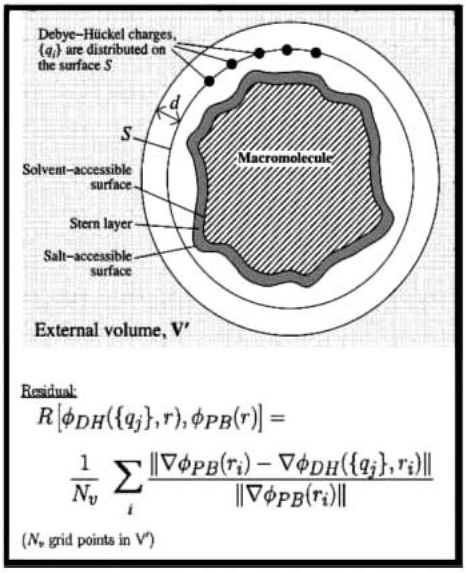
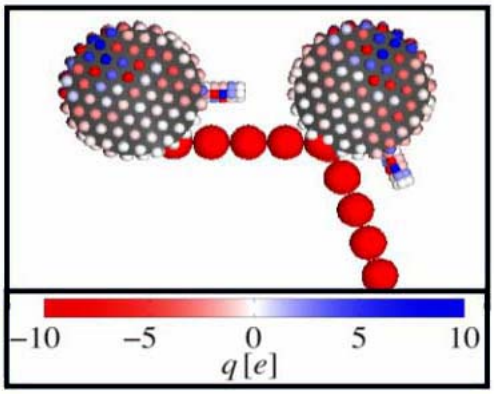

BIOPOLYMERS
ORIGINAL RESEARCH ON BIOMOLECULES
H3
H4
H2A
H2B
Editor-in-Chief: Murray Goodman
Volume 58 / Number 1 / January 2001
Articles published online in Wiley InterScience, 28 September 2000 through 3 November 2000
Cover Illustration: Crystal structure of nucleosome core particle from Luger et al. See Beard and Schlick, pp. 106-115, this issue.
WILEY
InterScience
277 DH charges on core particle to approximate PB soln. of electric field
charged linker DNA beads (elastic wormlike model)

Figure 3





Debye-Hückel electrostatics:

$$\phi_{DH} = \sum_{j>i} \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}}$$

κ^{-1}: salt-dependent Debye length

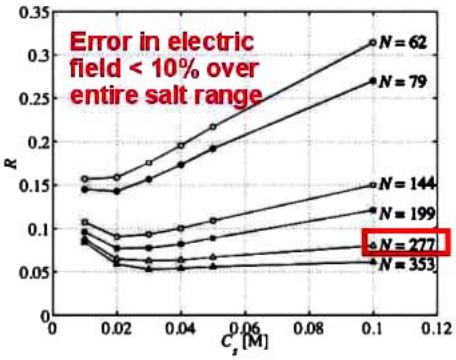
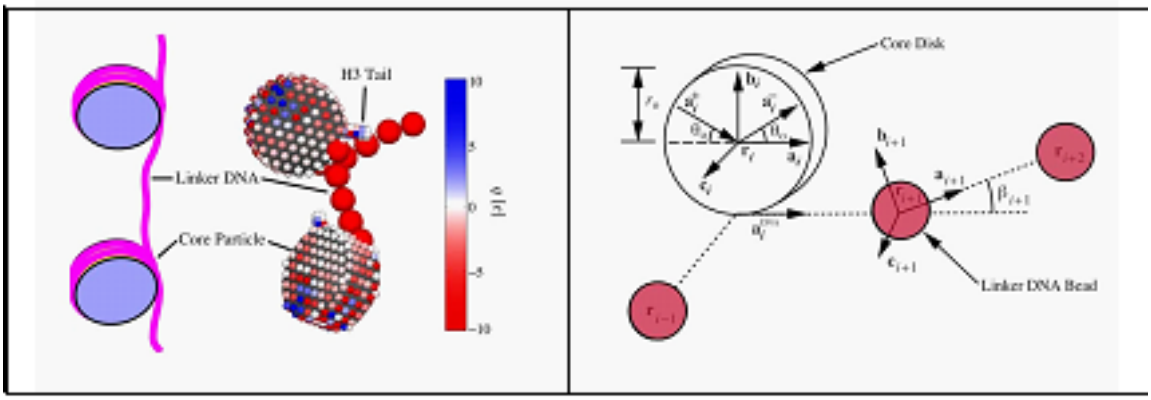


Figure 4

Model for Chromatin Dynamics

Basic unit involving electrostatic parameterization

Position vectors and coordinates for core and linker DNA beads



Chromatin Energetics and Dynamics

Energetics

Stretching: $E_S = \frac{h}{2}\sum_{i=1}^{N-1}(l_i - l_0)^2$ $l_0 = 3$ nm (~10 bp/segment)

Bending: $E_B = \frac{p_0 k_B T}{2l_0}\sum_{i=1}^{N-1}(\beta_i)^2$ $p_0 = 50$ nm

Torsion: $E_T = \frac{C}{2l_0}\sum_{i=1}^{N-1}(\phi_i + \gamma_i)^2$ $C = 2\times10^{-19}$ erg.cm

Electrostatics: $E_C = \sum_{j>i}\frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}}$ salt dependent coefficients

Excluded volume: $E_V = k_{mn}\sum_{j>i}\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}$ component dependent coefficients

Brownian dynamics with hydrodynamics

Translation: $r^{n+1} = r^n + \left(\frac{\Delta t}{k_B T}\right)D^n \cdot F^n + R^n$, where $\langle (R^n)(R^m)^T \rangle = 2\Delta t D^n \delta_{nm}$

$F^n = -\nabla E$, $D^n = k_B T H$, and H = Rotne-Prager Tensor

Rotation: $\Delta\Omega^{n+1} = \Delta\Omega^n + \left(\frac{\Delta t}{\xi}\right)\Gamma^n + \omega^n$, where $\langle (\omega^n)(\omega^m)^T \rangle = (2k_B T/\xi)\delta_{nm}$

Γ^n = torque, ξ = rotational friction coefficient

$\Delta t = 2$ ps

Figure 5

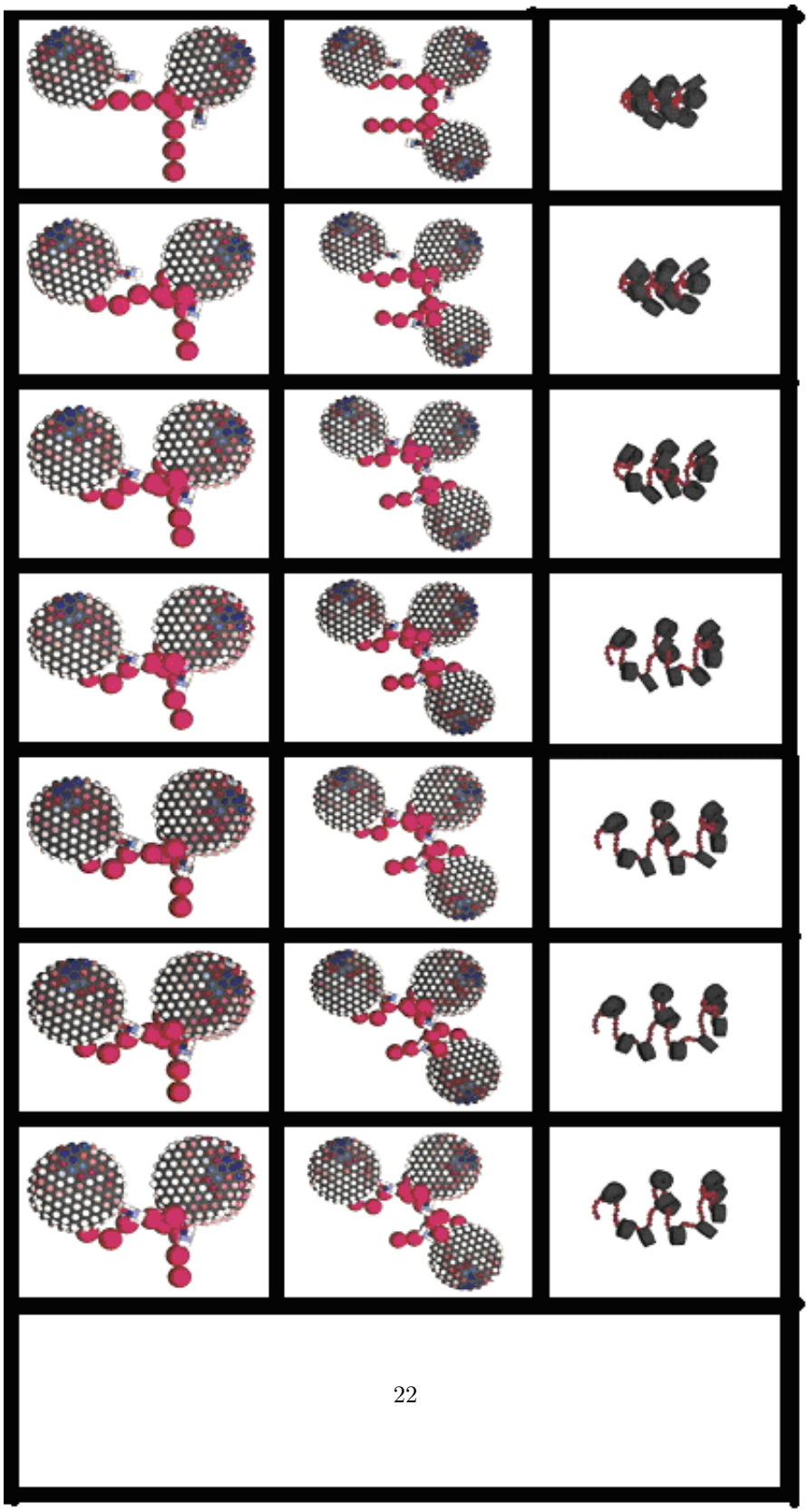


Figure 6

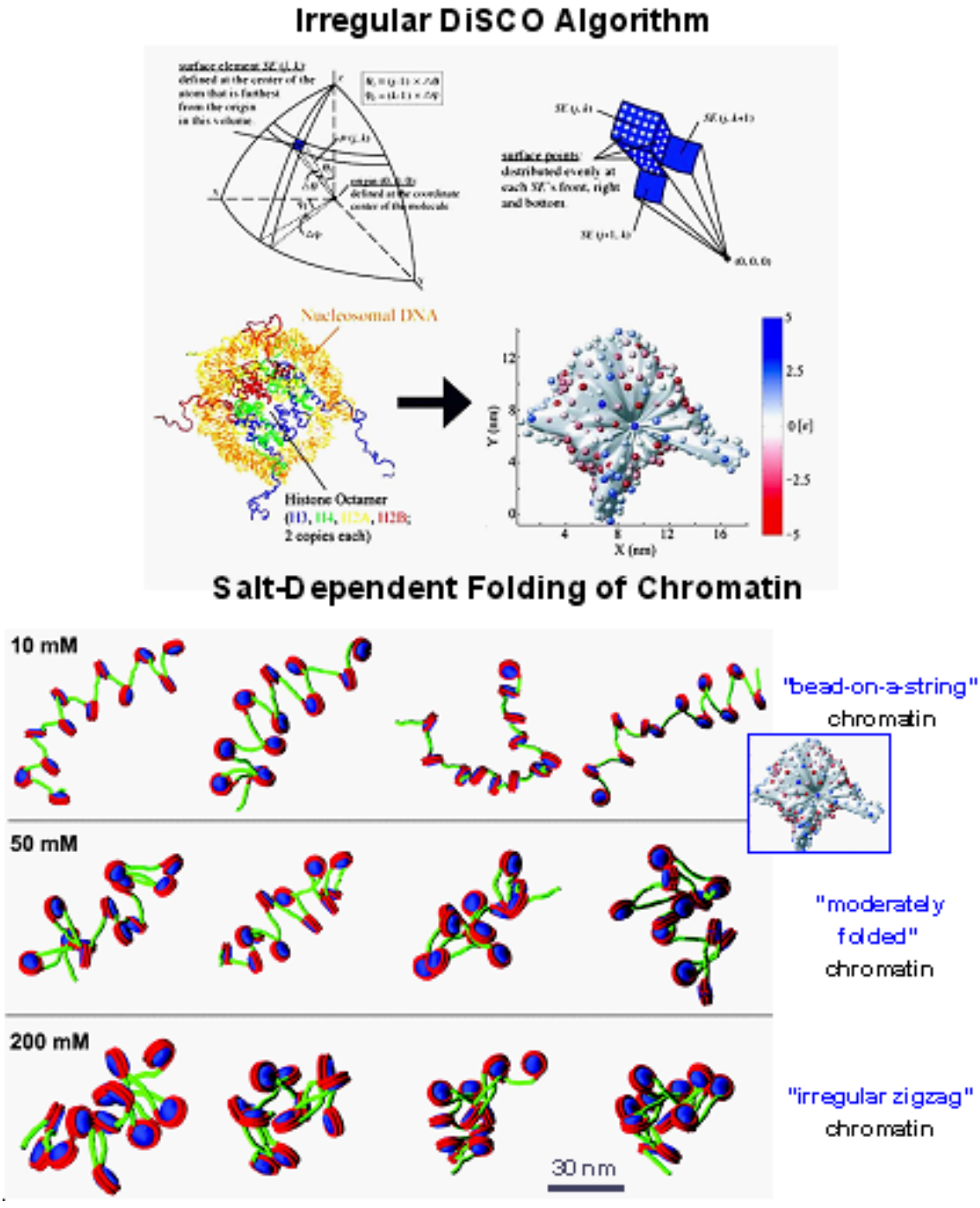

Irregular DiSCO Algorithm
surface element SE (i, j, k)
defined at the center of the
atom that is furthest
from the origin
in this volume
surface points
distributed evenly at
each SE's front, right
and bottom
Nucleosomal DNA
Histone Octamer
(H3, H4, H2A, H2B;
2 copies each)
Y (nm)
X (nm)
Salt-Dependent Folding of Chromatin
10 mM
50 mM
200 mM
"bead-on-a-string"
chromatin
"moderately
folded"
chromatin
"irregular zigzag"
chromatin
30 nm

Figure 7

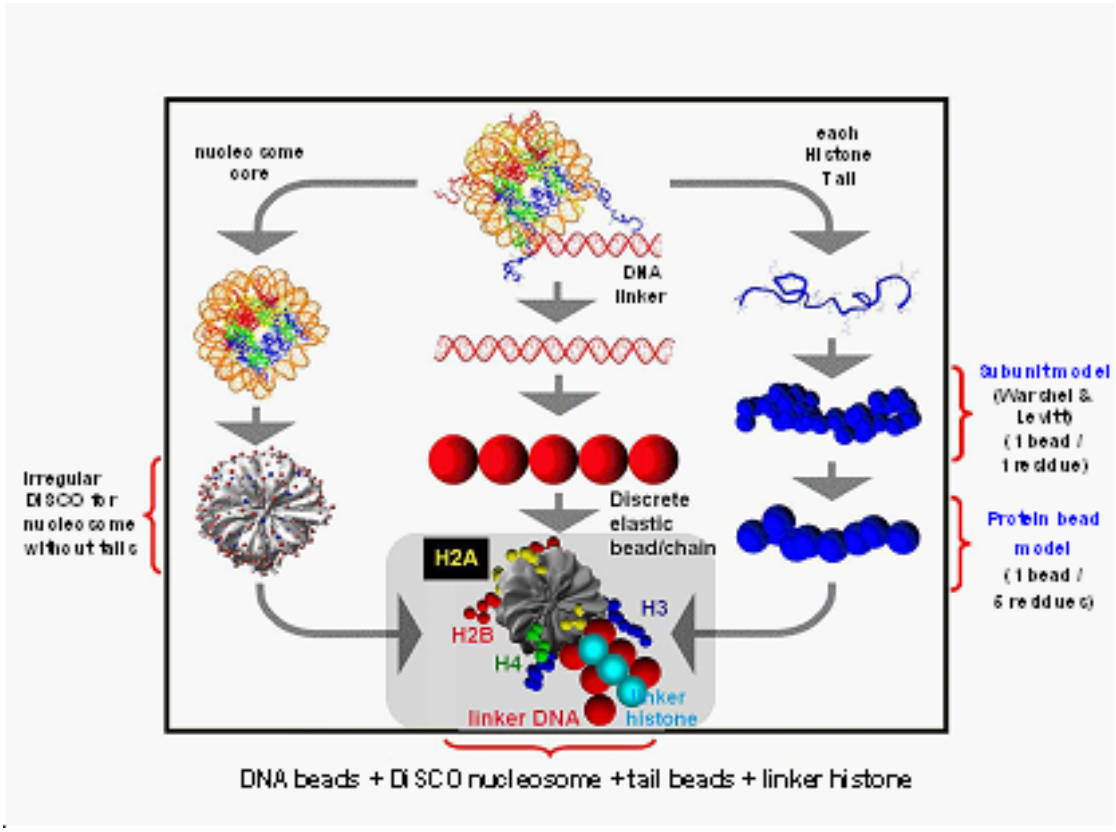

nucleosome core
each Histone Tail
DNA linker
Irregular DiSCO for nucleosome without tails
Discrete elastic bead/chain
Subunit model (Wardiel & Levitt) (1 bead / 1 residue)
Protein bead model (1 bead / 6 residues)
H2A
H2B
H3
H4
linker DNA
linker histone
DNA beads + DiSCO nucleosome +tail beads + linker histone

Figure 8

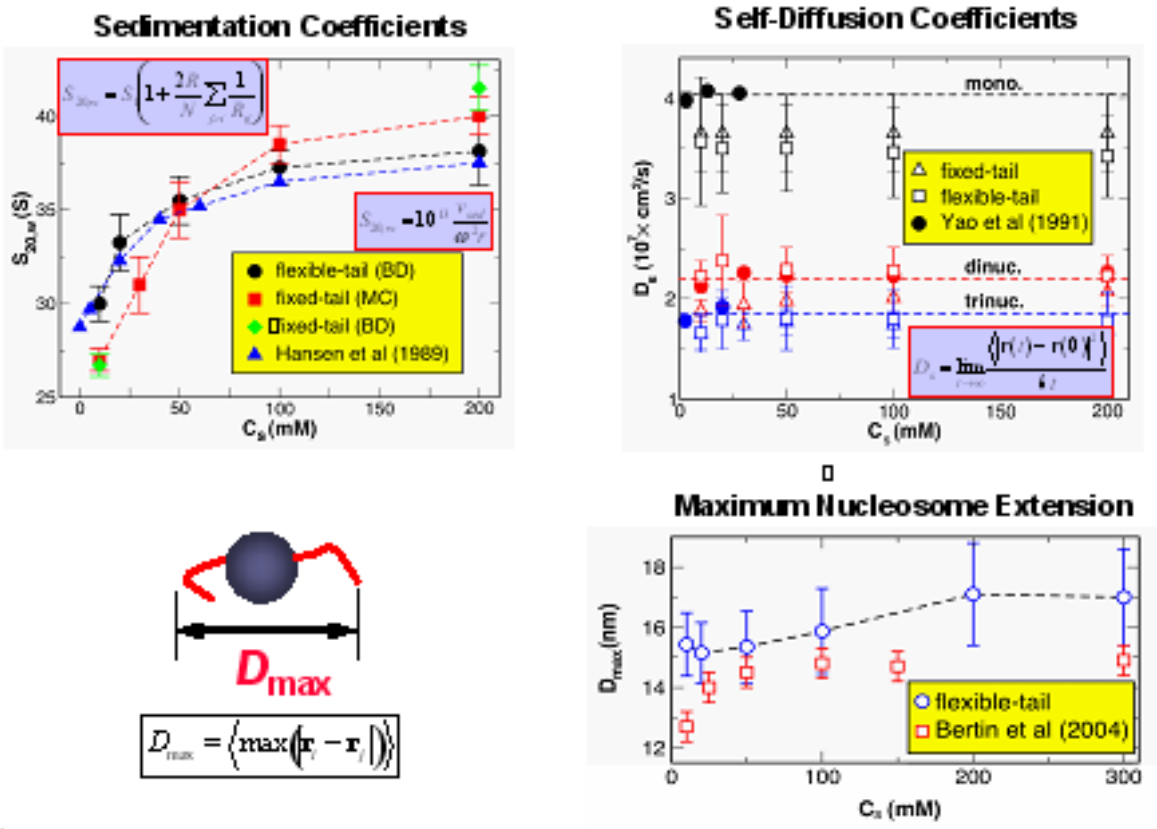

Sedimentation Coefficients
flexible-tail (BD)
fixed-tail (MC)
fixed-tail (BD)
Hansen et al (1989)
C_s (mM)
Self-Diffusion Coefficients
mono.
fixed-tail
flexible-tail
Yao et al (1991)
dinuc.
trinuc.
C_s (mM)
D_{max}
Maximum Nucleosome Extension
flexible-tail
Bertin et al (2004)
C_s (mM)

Figure 9

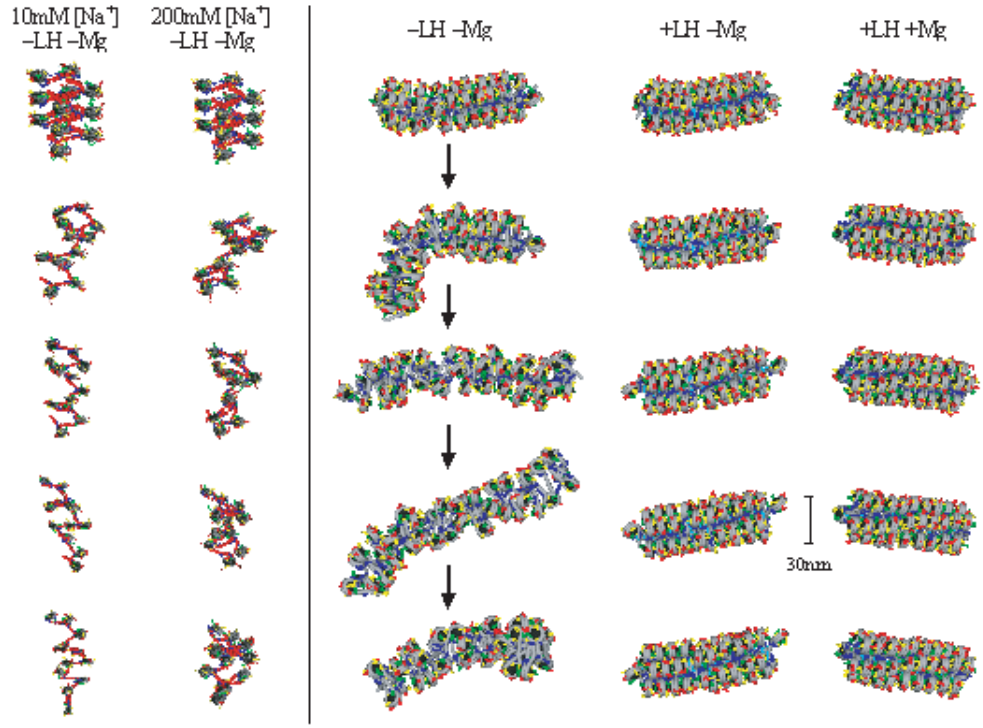


Figure 10

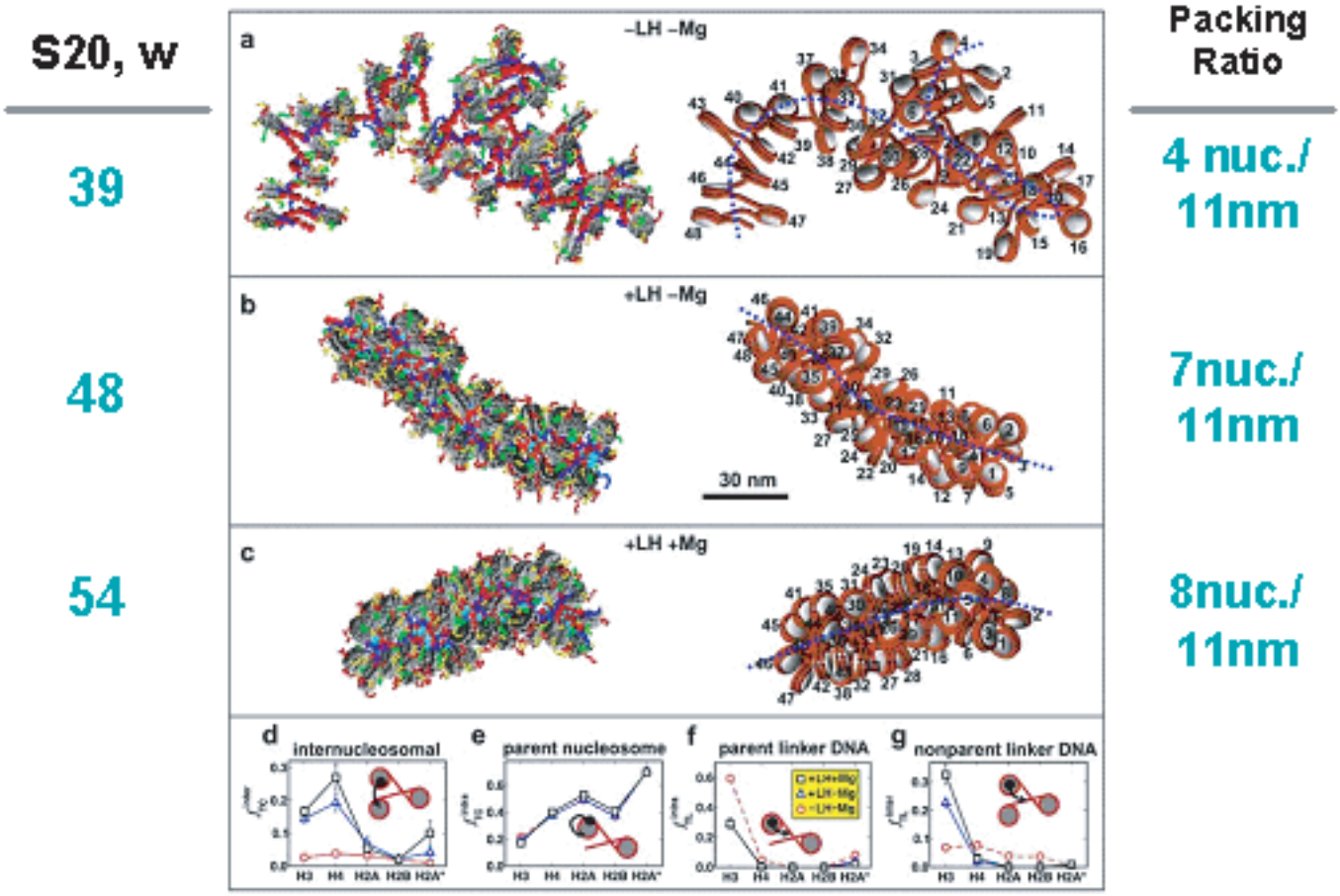

S20, w
39
48
54
a
–LH –Mg
b
+LH –Mg
30 nm
c
+LH +Mg
d
internucleosomal
e
parent nucleosome
f
parent linker DNA
g
nonparent linker DNA
H3
H4
H2A
H2B
H2A*
Packing Ratio
4 nuc./ 11nm
7nuc./ 11nm
8nuc./ 11nm

Figure 11

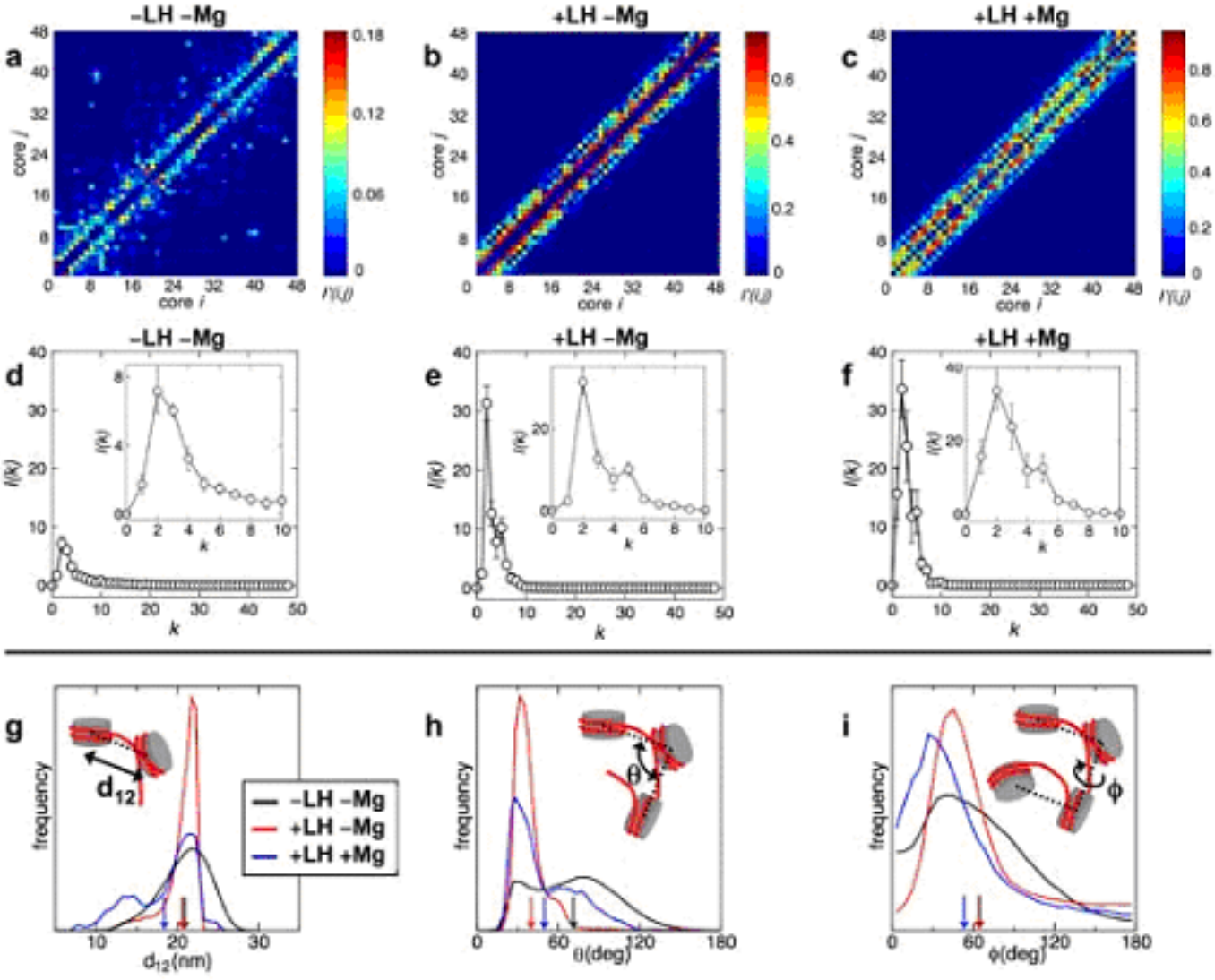

–LH –Mg
+LH –Mg
+LH +Mg
a
b
c
core j
core i
I'(i,j)
d
e
f
I(k)
k
g
h
i
frequency
d12
d12(nm)
θ
θ(deg)
ϕ
ϕ(deg)
–LH –Mg
+LH –Mg
+LH +Mg

Figure 12