Dynamic condensation of linker histone C-terminal domain regulates chromatin structure

Supplementary Information

Antoni Luque, Rosana Collepardo-Guevara, Sergei A. Grigoryev, and Tamar Schlick

Refined LH model

Linker histone globular head

We model the LH globular head (GH) by coarse-graining the globular head of H1.4 (76 residues) from the all-atom linker histone model of Bharath *et al.* [1, 2]. After adding hydrogen atoms, we compute the positions and sizes of the rigid beads by applying the coarse graining shaped-based method[1] developed in the Schulten Lab [4–7]. Our models involve $N_{gh} = 1$ to 10 spherical beads. The charges are adjusted using DiSCO [8, 9] at different monovalent salt concentrations at room temperature (293.15K); see Table S.1. As expected, increasing the number of beads in the model reduces the relative error of the Debye-Hückel charges with respect the all-atom Poisson-Boltzmann electric field. The $N_{gh} = 6$ case shows a pronounced drop in the relative error and a good structural matching with the all-atom structure. Since models with $N_{gh} > 6$ provide similar accuracy, we coarse-grain the globular head using six beads, computationally more efficient. We assume a symmetric positioning of the GH in accord with recent high resolution experiments [10] and orient the GH in the dyad based on the predicted full-atom model for GH and CTD of H1 [1, 2] (see Figs. 1 and S.1). Table S.2 gives the positions of the GH beads in the nucleosome reference system. The excluded volume of the GH is adjusted to be compatible with the steric effects of the other chromatin elements [9] (see Table S.2).

[1] This technique is integrated in VMD [3].

Bead	q [e] (5mM)	q [e] (15mM)	q [e] (80mM)	q [e] (150mM)
1	−4.19	−4.12	−3.64	−3.29
2	2.08	2.13	3.11	4.22
3	6.90	6.83	7.47	8.48
4	−0.28	−0.18	−0.09	0.28
5	1.10	1.38	1.89	2.08
6	1.21	1.53	2.53	3.27

Table S.1: Coarse-grained charges for the $N_{gh} = 6$ globular head model with DiSCO at 5, 15, 80, and 150 mM NaCl.

Bead	x [nm]	y [nm]	z [nm]	σ_{gg} [nm]	σ_{gn} [nm]	σ_{gl} [nm]	σ_{gt} [nm]
1	5.2268	−3.8609	1.0762	1.4360	1.7134	2.2180	1.6180
2	4.8939	−3.0429	−0.6935	1.4720	1.7368	2.2360	1.6360
3	5.7043	−3.2130	0.5458	1.4460	1.7199	2.2230	1.6230
4	4.5038	−3.7282	0.3817	1.5380	1.7797	2.2690	1.6690
5	4.1897	−3.9769	−0.8150	1.6180	1.8317	2.3090	1.7090
6	5.7580	−4.1316	−0.4953	1.5280	1.7732	2.2640	1.6640

Table S.2: Coarse-grained parameters for globular head model with 6 beads. The x, y, and z are the coordinates of the beads in the nucleosome reference system. The intrinsic excluded volume of each GH bead is σ_{gg}. The other parameters correspond to the excluded volume of GH-core (σ_{gn}), GH-linker (σ_{gl}), and GH-tail (σ_{gt}).

Linker histone C-terminal domain

The LH C-terminal domain (CTD) is a highly basic and intrinsically unstructured polypeptide similar to the core histone tails (Fig. S.1). As a first approximation, we coarse-grain the unfolded LH C-terminal with parameters used in our core histone tail model. We use 5 amino acids per bead, i.e., $N_C = 22$ beads to represent the 111 residue rat H1.4 CTD (109-219). The atomistic charges are replaced by effective Debye-Hückel charges adjusted at different salt concentrations; the effective charge of a bead is the total charge of its associated amino-acids multiplied by the prefactors 1.1063 (5mM monovalent salt), 1.1644 (15mM), 1.4815 (80mM), and 1.6800 (150mM), which were derived from the core histone tail prefactors [12]. We also include a 3/2 prefactor adjusted to reproduce the mesoscopic structural data of chromatin fibers. For each chromatin element, the CTD beads have a specific excluded volume: $\sigma_{cc} = 1.8$ nm (CTD-CTD)), $\sigma_{cn} = 1.8$ nm (CTD-core), $\sigma_{cl} = 2.7$ nm (CTD-linker), $\sigma_{ct} = 1.8$ nm (CTD-tail), and $\sigma_{cg}(i) = (\sigma_{cc} + \sigma_{gg}(i))/2$ (CTD-GH), where i is the index of the GH bead. Based on the core histone tail model, we derive a uniform elastic force field for the CTD: $k_s \sim 10$ kcal/mol nm^2 (stretching constant), $k_b \sim 1$ kcal/mol rad^2 (bending constant),

$l_{eq} \sim 1.5$ nm (stretching equilibrium length), and $\beta_{eq} \sim 110°$ (bending equilibrium angle) [12]. Because the unfolded equilibrium configuration of the CTD is long, $\sim$ 25 nm, we also design a more compressed but unfolded CTD starting configuration: it has length $\sim$ 10 nm and is oriented through the dyad axis to fit with the initial medium-linker DNA fibers (see Fig. S.1B).

Linker histone interaction

Each CTD has associated intramolecular force-field composed of elastic and excluded volume terms. The intramolecular stretching and bending components of the elastic term are, respectively,

$$E_s = \sum_{i=1}^{N_C-1} \frac{1}{2} k_s (l_i - l_{eq})^2 \tag{S.1}$$

and

$$E_b = \sum_{i=1}^{N_C-2} \frac{1}{2} k_b (\beta_i - \beta_{eq})^2 \quad , \tag{S.2}$$

where l_i is the distance between two consecutive beads, and β_i is the angle defined by three consecutive beads. The excluded volume term avoids the overlapping of non-consecutive beads:

$$E_{ev} = \sum_{i=1}^{N_C-2} \sum_{j>i+1}^{N_C} k_{ev} \left[\left(\frac{\sigma_{cc}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{cc}}{r_{ij}} \right)^{6} \right] \quad . \tag{S.3}$$

Here r_{ij} is the distance between beads i and j, and the prefactor k_{ev} is 0.001 $k_B T$ as in the rest of the model [9]. We assume that the internal electrostatic interaction between CTD beads is compensated during condensation with other molecular interactions, like hydrogen bonding, hydrophobic interactions, and salt-bridges. Despite this approximation cannot capture the folding of specific secondary motifs, it is very efficient computationally and captures the global structural trend of the CTD observed experimentally (see Fig. 3).

As in our prior model [9], each LH can interact with the entry and exit linker DNAs, non-parental linker DNAs and nucleosome cores, all core histone tails, and other LHs. For each bead, all pair interactions with these elements include an excluded volume term,

$$E_{ev} = \sum_{i=1}^{N_C-2} \sum_{j>i+1}^{N_C} k_{ev} \left[\left(\frac{\sigma_{cc}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{cc}}{r_{ij}} \right)^{6} \right] \quad , \tag{S.4}$$

and a Debye-Hückel potential,

$$E_c(i,j) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} exp(-\kappa r_{ij}) \quad . \tag{S.5}$$

Here i is a LH bead and j is an effective bead of an interacting chromatin element; σ_{ij} is the specific excluded volume parameter as described above; q_i and q_j are the effective charges; κ is the inverse Debye length; ϵ_0 is the electric permittivity of vacuum, and ϵ_r is the relative permittivity of the medium (set to 80).

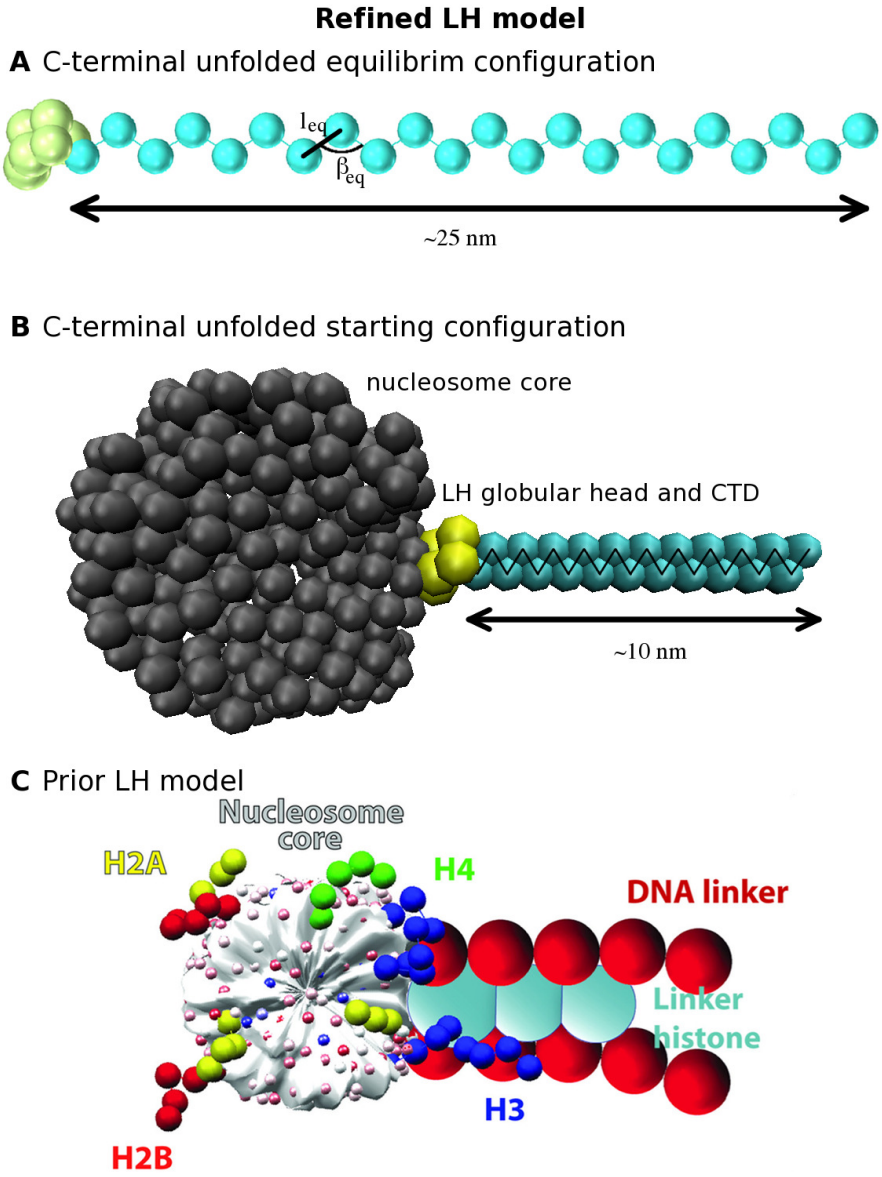

Figure S.1: Refined LH model. (A) Equilibrium LH model. The globular head (gold) is made of 6 beads with coarse-grain parameters derived from atomistic calculations. The unfolded C-terminal (cyan) lies on the dyad axis and is made of 22 beads that are coarse-grained adapting the core histone tail model (5 amino acid resolution). The parameter $l_{eq} = 1.5$ nm is the stretching equilibrium length, and $\beta_{eq} = 110°$ is the bending equilibrium angle of the elastic force field. The equilibrium unfolded C-terminal is relatively large, ~ 25 nm; (B) A compressed configuration of the CTD of ~ 10 nm using a bong length $l = l_{eq}/1.69$ and a bending angle $\beta = \beta_{eq}/1.69$. (C) Prior LH model inserted in the nucleosome [9, 11]. The LH is made of three beads with charges adjusted from the full atom structure predicted by Bharath *et al.* [1, 2].

Structural properties

Calculation of the sedimentation coefficient

We approximate the sedimentation coefficient, $s_{20,w}$, adapting the method developed by Bloomfield *et al.* in refs. [13, 14] to the case of oligonucleosomes using inter-core distances [15, 16]. In this approximation, the N cores of the fiber dominate the sedimentation velocity of the structure, i.e., $s_{20,w} \approx S_N$, where

$$\frac{S_N}{S_1} = 1 + \frac{R_1}{N_1} \sum_i \sum_j \frac{1}{R_{ij}} \quad . \tag{S.6}$$

Here $R_1 = 5.5$ nm is the effective spherical radius of the cores; R_{ij} is the distance between the centers of the cores i and j, and S_1 is the sedimentation coefficient of a single nucleosome, which is $S_{-LH} = 11.1$ S (Svedberg) without LH [16] and $S_{+LH} = 12.0$ S with LH [17]. This approach is compatible with a more sophisticated method using the program HYDRO [9, 18].

Packing ratio

The fiber packing ratio, pr, represents the number of nucleosomes per 11 nm of fiber length. In each simulation frame, we calculate the fiber length, L_{fiber}, by defining a curve through the center of the fiber. The curve is parametrized as a quadratic polynomial in each of the x, y, and z directions. The least-square distance with the center of the cores in each direction determines the polynomial's parameters. The fiber length is the sum of the consecutive distances of nucleosomes projected in this fiber axis. For a fiber of N nucleosomes, the packing ratio is then

$$pr = \frac{N}{L_{fiber}} \times 11\,\text{nm} \quad , \tag{S.7}$$

where L_{fiber} is given in nm. See ref. [9] for a detailed derivation of this method.

Core-core interaction pattern

To determine the frequency of interaction between nucleosomes, we have developed a method comparable to cross-linking experiments in chromatin fibers [11]. In our approach, nucleosome i is considered in contact

with nucleosome j when the distance between a tail bead of i and a tail bead or core charge of j is smaller than the correpsonding tail-tail or tail-core excluded volume distance [12]. For a given frame M, such pairwise distances define a symmetric, binary interaction matrix for all nucleosome pairs, $\delta_{i,j}(M)$. The average of the matrix $\delta_{i,j}$ over sampled configurations gives the internucleosome interaction matrix $I'(i,j)$. To determine the number of interactions for all elements separated by k neighbors, we define a function of k as follows

$$I(k) = \sum_{i=1}^{N} I'(i, i \pm k)/(N-1) \quad , \tag{S.8}$$

where N is the total number of nucleosomes in the fiber [9].

Gyration tensor and structure of the LH CTD

To characterize the structural properties of the flexible LH terminal, we compute the gyration tensor,

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)} \tag{S.9}$$

where $r_m^{(i)}$ is the m^{th} Cartesian coordinate of the i^{th} particle with origin in the center of mass of the LH CTD. This tensor is symmetric and can be diagonalized,

$$S = \begin{bmatrix} \lambda_x^2 & & \\ & \lambda_y^2 & \\ & & \lambda_z^2 \end{bmatrix} , \tag{S.10}$$

where the eigenvalues (principal moments of the gyration tensor) follow the relation $\lambda_x^2 \leq \lambda_y^2 \leq \lambda_z^2$. In this way we can compute the squared radius of gyration

$$R_g^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2 \, , \tag{S.11}$$

which is equivalent to the standard definition $R_g^2 = \sum_i (\vec{r}_m - \langle \vec{r} \rangle)^2 / N$, the asphericity,

$$b = \lambda_z^2 - \frac{1}{2}(\lambda_x^2 + \lambda_y^2) \, , \tag{S.12}$$

the acylindricity,

$$c = \lambda_y^2 - \lambda_x^2 \,, \tag{S.13}$$

and the relative shape anisotropy,

$$\kappa^2 = (b^2 + {}^3\!/_4\, c^2)/R_g^4 \,, \tag{S.14}$$

which is bounded between 0 (spherical) and 1 (cylindrical).

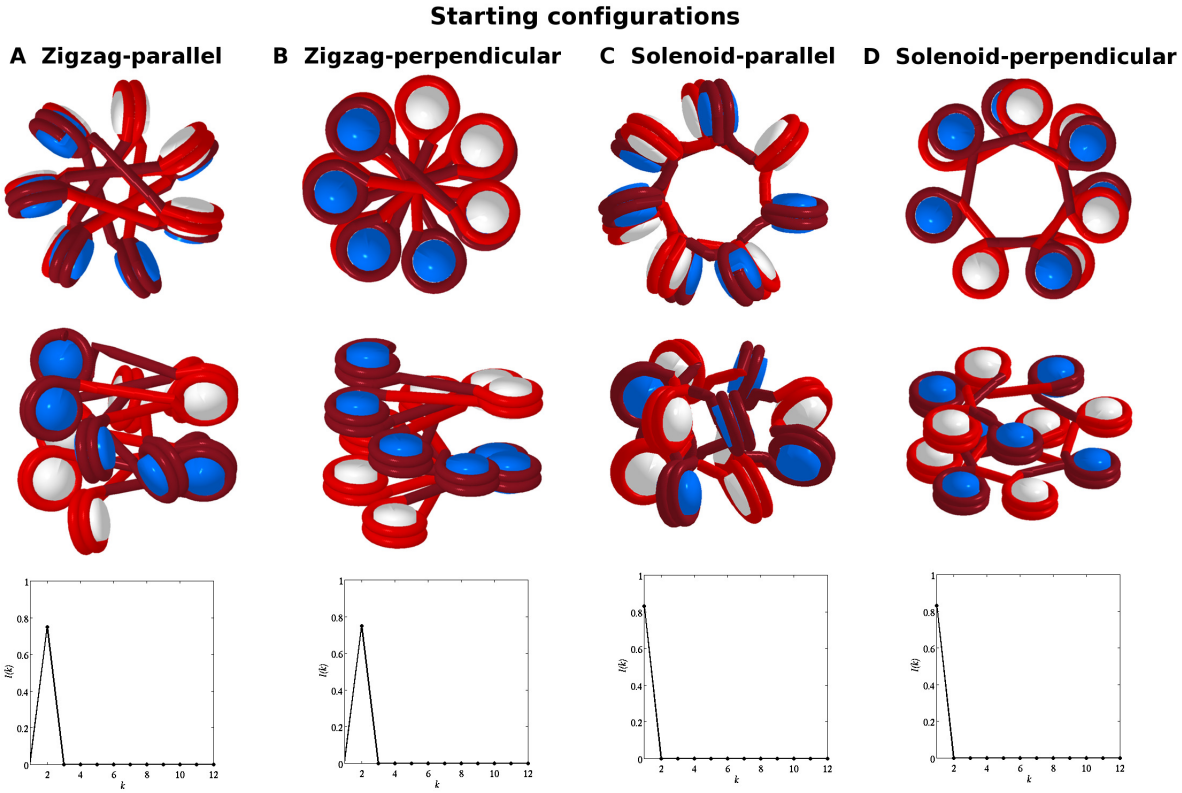

Figure S.2: Starting configurations for 12x209bp fibers: (A) zigzag-parallel, (B) zigzag-perpendicular, (C) solenoid-parallel, and (D) solenoid-perpendicular. Each panel includes a top and a side view of the structure as well as the internucleosome interaction pattern.

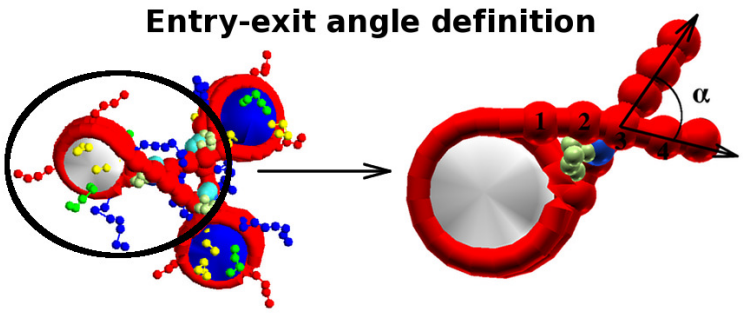

Figure S.3: The entry-exit angle of a nucleosome is determined by the directions of the entry and exit linker DNAs. These directions are defined by the vector connecting the third and fourth DNA beads in each linker DNA, counting from the parent core. (The directions defined by the first two beads are dominated by the GH and do not capture the variations in the entry-exit angle.) The angle is not defined for extremal nucleosomes; thus, for a trinucleosome, only the central core contributes to the entry-exit angle analysis.

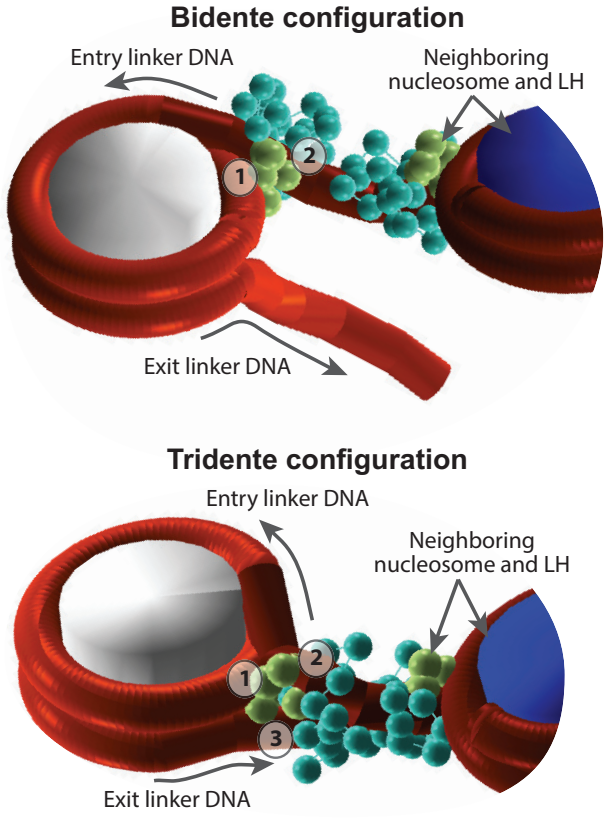

Figure S.4: Bidente and tridente chromatosome configurations extracted from a chromatin fiber at 15mM and 150mM NaCl, respectively. The numbers within circles and arrows highlight the main properties for each configuration.

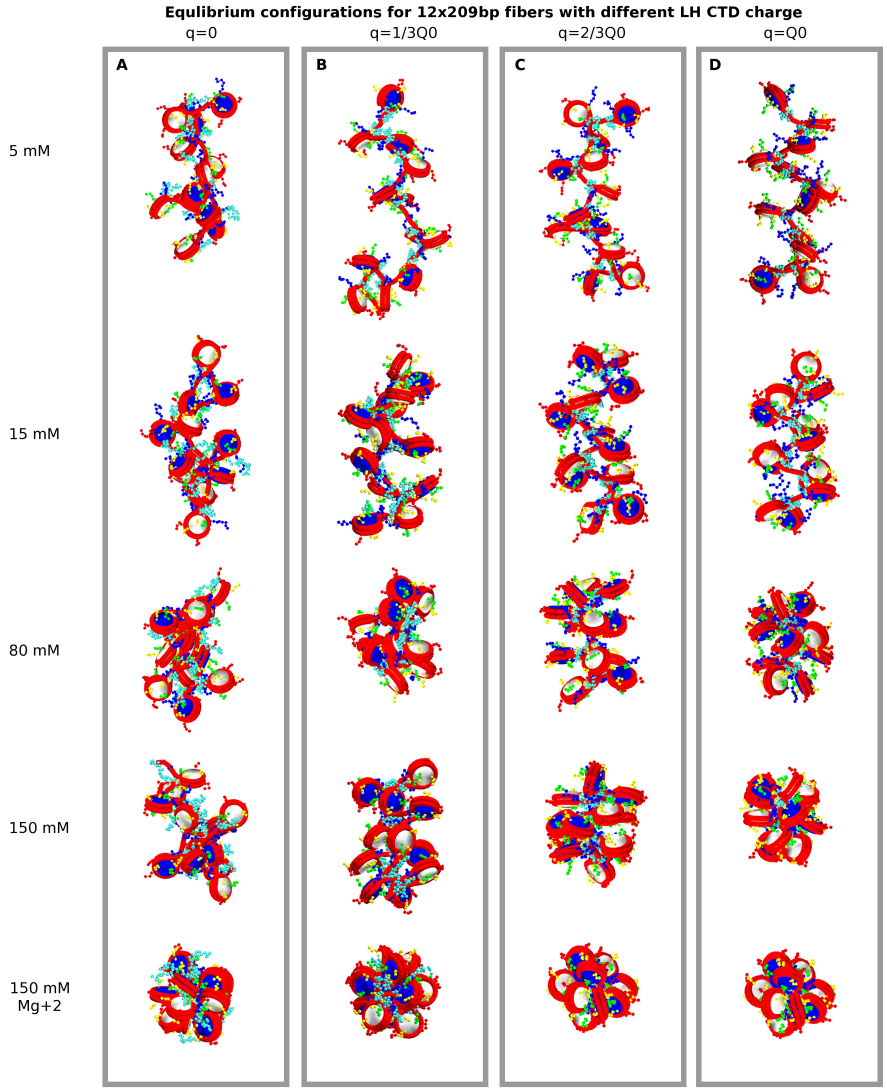

Figure S.5: Equilibrium structures for 12x209bp fibers for different CTD charged states: neutral (A), 1/3 charged (B), 2/3 charged (C), and fully charged (D). Rows are associated to the 5 salt conditions explored: NaCl concentration at 5mM, 15mM, 80mM, 150mM, and 150mM with Mg^{2+}.

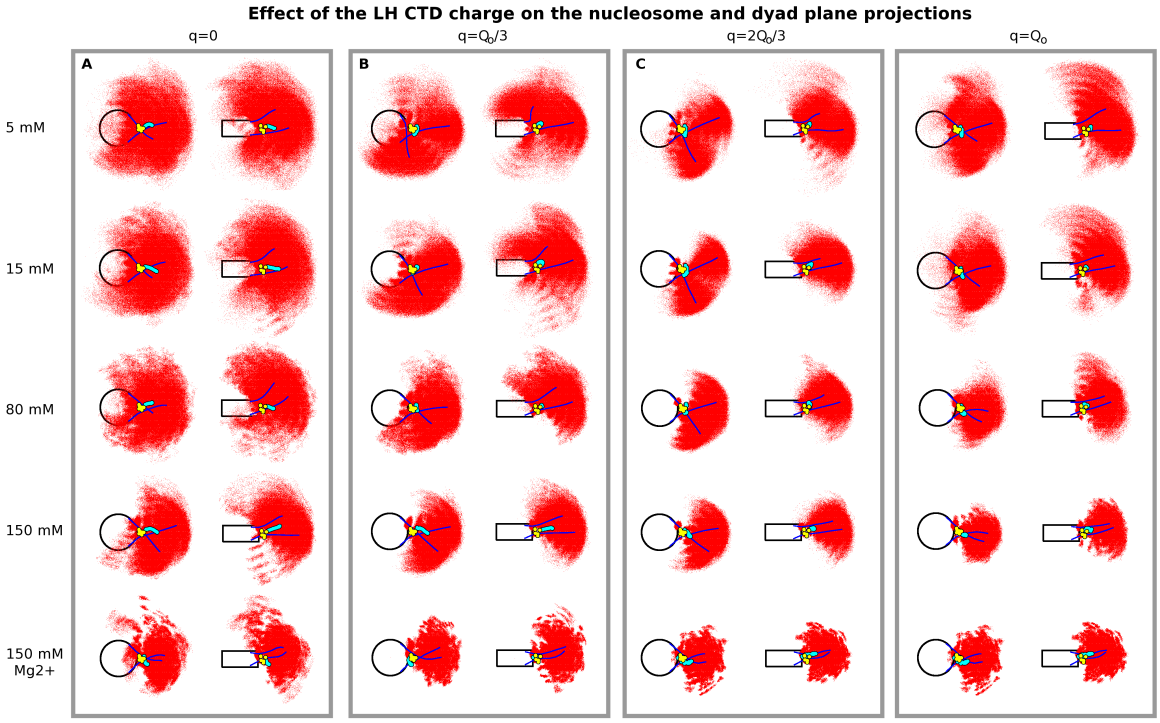

Figure S.6: Average chromatosome projections for 12x209bp fibers for different CTD charged states: neutral (A), 1/3 charged (B), 2/3 charged (C), and fully charged (D). Rows are associated to the 5 salt conditions explored: NaCl concentration at 5mM, 15mM, 80mM, 150mM, and 150mM with Mg^{2+}. Red dots are the accumulated projections of the entry and exit linker DNAs, and the blue lines are the average position of the linker DNAs. The cyan beads are the average CTD positions, and the Gh beads are in yellow.

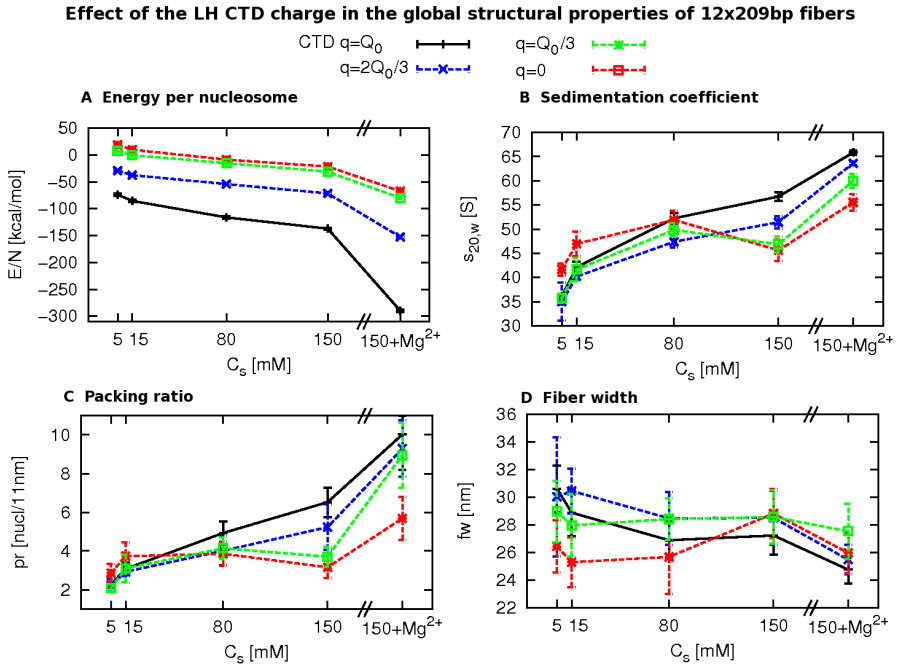

Figure S.7: Global properties for 12x209bp fibers for different CTD charged states: neutral (A), 1/3 charged (B), 2/3 charged (C), and fully charged (D). (A) Energy per nucleosome, (B) sedimentation coefficient, (C) packing ratio, and (D) fiber width. The error bars correspond to the standard deviation.

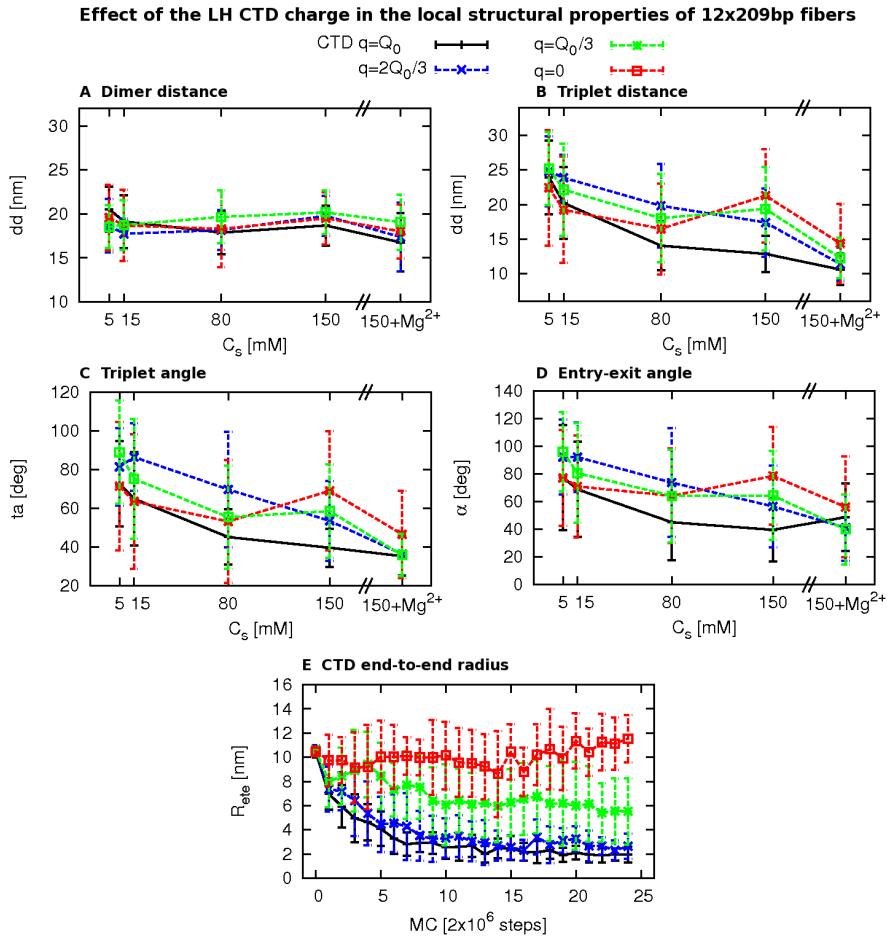

Figure S.8: Local properties for 12x209bp fibers for different CTD charged states: neutral (A), 1/3 charged (B), 2/3 charged (C), and fully charged (D). (A) Dimer distance, (B) triplet distance, (C) triplet angle, and (D) entry-exit angle. (E) CTD condensation at 150 mM NaCl illustrated by the end-to-end radius for single trajectories, where each point is the average R_{ete} overall LHs in a fiber. All error bars correspond to the standard deviation.

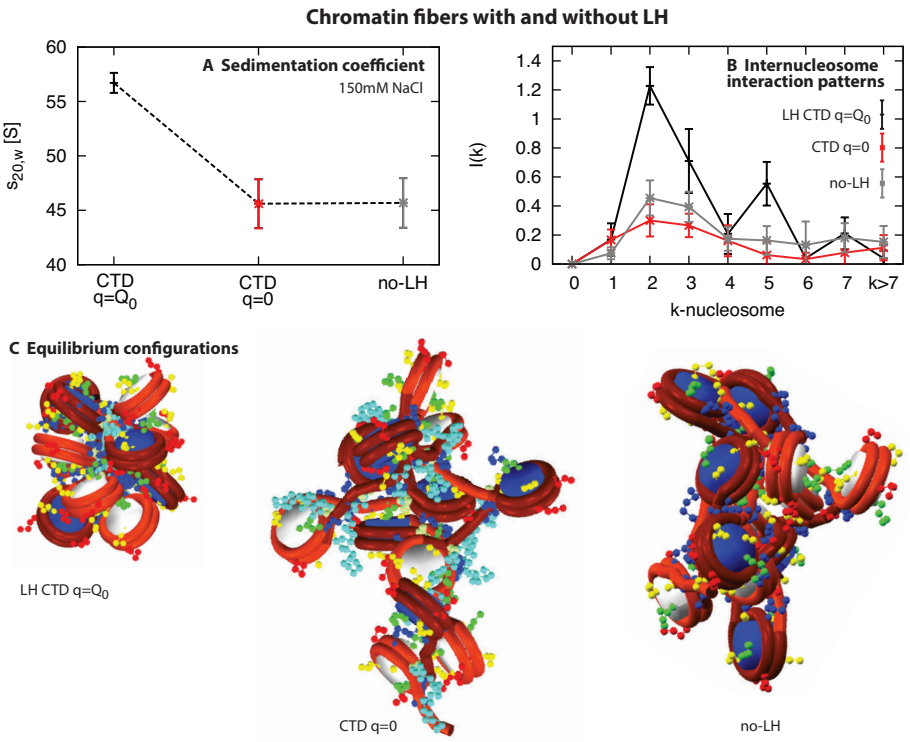

Figure S.9: (A) Sedimentation coefficient, (B) internucleosome interaction pattern, and (C) equilibrium snapshots for fibers with LH CTD q=Q_0 (black), LH CTD q=0 (red), and without LH (gray).

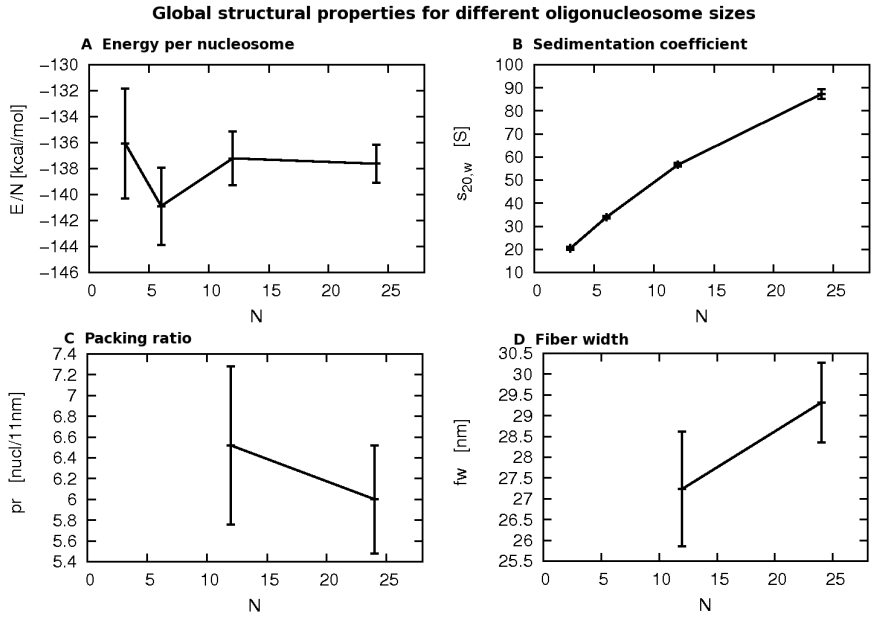

Figure S.10: Global structural properties for 3-, 6-, 12-, and 24-oligonucleosomes with 209bp NRL. (A) Energy per nucleosome, (B) sedimentation coefficient, (C) packing ratio, and (D) fiber width. The algorithm to compute the packing ratio and fiber width is defined for $N > 6$ oligonucleosomes.

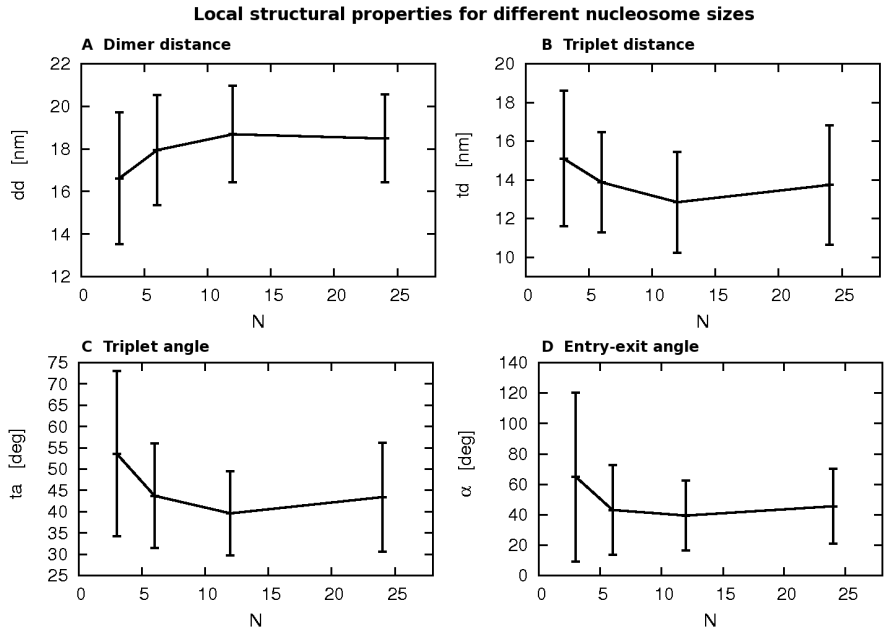

Figure S.11: Local structural properties for 3-, 6-, 12-, and 24-oligonucleosomes with 209bp NRL. (A) Dimer distance, (B) triplet angle, and (E) entry-exit angle (see definition in Fig. S.3).

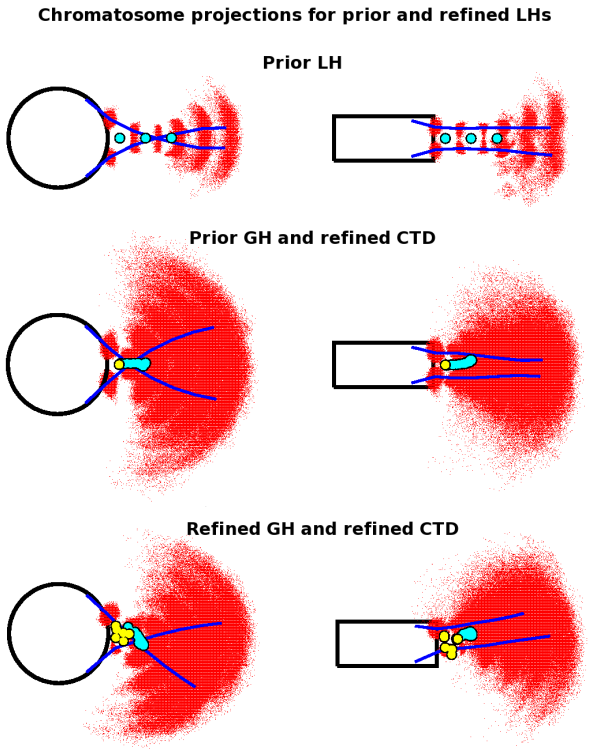

Figure S.12: Chromatosome projections for 12x209bp fibers containing the prior and refined LH domain models. The color code is described in Fig. **??**. Model simulated at 150mM NaCl.

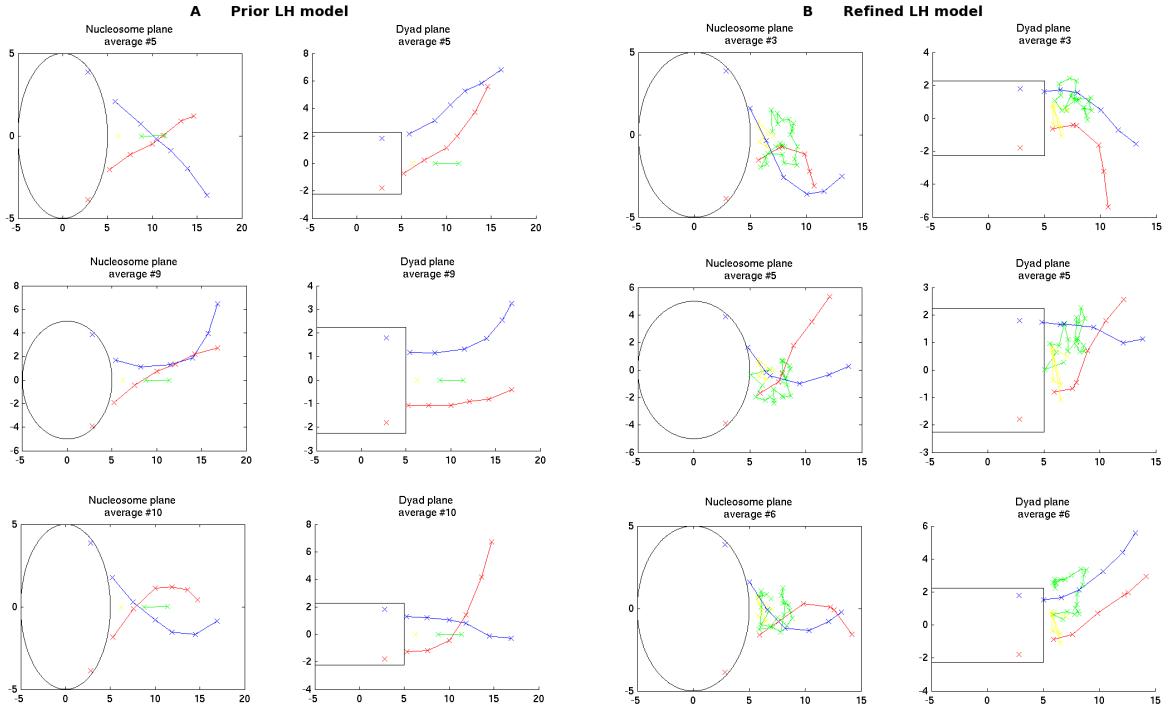

Figure S.13: Individual chromatosome projections for 12x209bp fibers containing the old (A) and new LH (B) models. For each case we choose three representative nucleosomes and plot the projection average in the nucleosome plane (left) and dyad plane (right): Entry linker DNA (red), exit linker DNA (blue), GH (yellow), and CTD (green). Model simulated at 150mM NaCl.

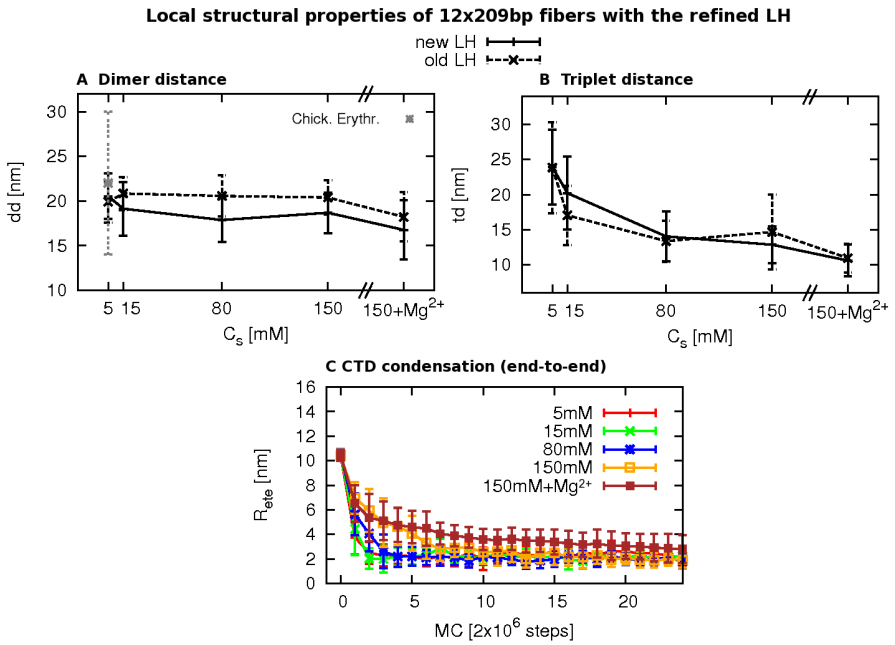

Figure S.14: Local structural properties for 12x209bp fibers with the refined LH (solid line) and the prior LH (dashed line) at different monovalent ionic strengths (C_s): (A) Dimer distance and (B) triplet angle. The experimental values correspond to native chromatin from chicken erythrocytes [19, 20]. (C) Condensation trajectories for CTD end-to-end distances, R_{ete}, at different concentrations of NaCl and Mg^{2+}.

References

[1] Bharath, M., Chandra, N., and Rao, M. (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.

[2] Bharath, M., Chandra, N., and Rao, M. (2003) Molecular modeling of the chromatosome particle *Nucleic Acids Res.* **31**, 4264–4274.

[3] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - Visual Molecular Dynamics *J. Molec. Graphics* **14**, 33–38.

[4] Arkhipov, A., Freddolino, P., and Schulten, K. (2006) Stability and dynamics of virus capsids described by coarse-grained modeling *Structure* **14**, 1767–1777.

[5] Arkhipov, A., Freddolino, P., Imada, K., Namba, K., and Schulten, K. (2006) Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum *Biophys. J.* **91**, 4589–4597.

[6] Arkhipov, A., Yin, Y., and Schulten, K. (2008) Four-scale description of membrane sculpting by BAR domains *Biophys. J.* **95**, 2806–2821.

[7] Yin, Y., Arkhipov, A., and Schulten, K. (2009) Simulations of membrane tubulation by lattices of amphiphysin N-BAR domains *Structure* **17**, 882–892.

[8] Zhang, Q., Beard, D., and Schlick, T. (2003) Constructing Irregular Surfaces to Enclose Macromolecular Complexes for Mesoscale Modeling Using the Discrete Surface Charge Optimization (DiSCO) Algorithm *J. Comput. Chem.* **24**, 2063–2074.

[9] Perišić, O., Collepardo-Guevara, R., and Schlick, T. (2010) Modeling studies of chromatin fiber structure as a function of DNA linker length. *J. Mol. Biol.* **403**, 777–802.

[10] Meyer, S., Becker, N., Syed, S., Goutte-Gattat, D., Shukla, M., Hayes, J., Angelov, D., Bednar, J., Dimitrov, S., and Everaers, R. (2011) From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem *Nucleic Acids Res.* **39**, 9139–9154.

[11] Grigoryev, S., Arya, G., Correll, S., Woodcock, C., and Schlick, T. (2009) Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions *Proc. Natl. Acad. Sci. USA* **106**, 13317–13322.

[12] Arya, G., Zhang, Q., and Schlick, T. (2006) Flexible Histone Tails in a New Mesoscopic Oligonucleosome Model *Bioph. J.* **91**, 133–150.

[13] Bloomfield, V., Dalton, W., and vanHolde, K. (1967) Frictional coefficients of multisubunit structures *I. Theory. Biopolymers* **5**, 135–148.

[14] Kirkwood, J. (1954) The general theory of irreversible processes in solutions of macromolecules *J. Polym. Sci.* **12**, 1–14.

[15] Hansen, J., Lu, X., Ross, E., and Woody, R. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* **281**, 1853–6.

[16] Garcia-Ramirez, M., Dong, F., and Ausio, J. (1992) Role of the histone "tails" in the folding of oligonucleosomes depleted of histone H1. *J. Biol. Chem.* **267**, 19587–95.

[17] Butler, P. and Thomas, J. (1998) Dinucleosomes Show Compaction by Ionic Strength, Consistent with Bending of Linker DNA *J. Mol. Biol.* **281**, 401–407.

[18] Garcia de laTorre, J., Navarro, S., Lopez Martinez, M. C., Diazand, F. G., and Lopez Cascales, J. J. (1994) HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules *Biophys. J.* **67**, 530–1.

[19] Scheffer, M., Eltsov, M., Bednar, J., and Frangakis, A. (2012) Nucleosomes stacked with aligned dyad axes are found in native compact chromatin in vitro. *Journal of structural biology* **178**, 207–214.

[20] Bednar, J., Horowitz, R., Grigoryev, S., Carruthers, L., Hansen, J., Koster, A., and Woodcock, C. (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**, 14173–8.

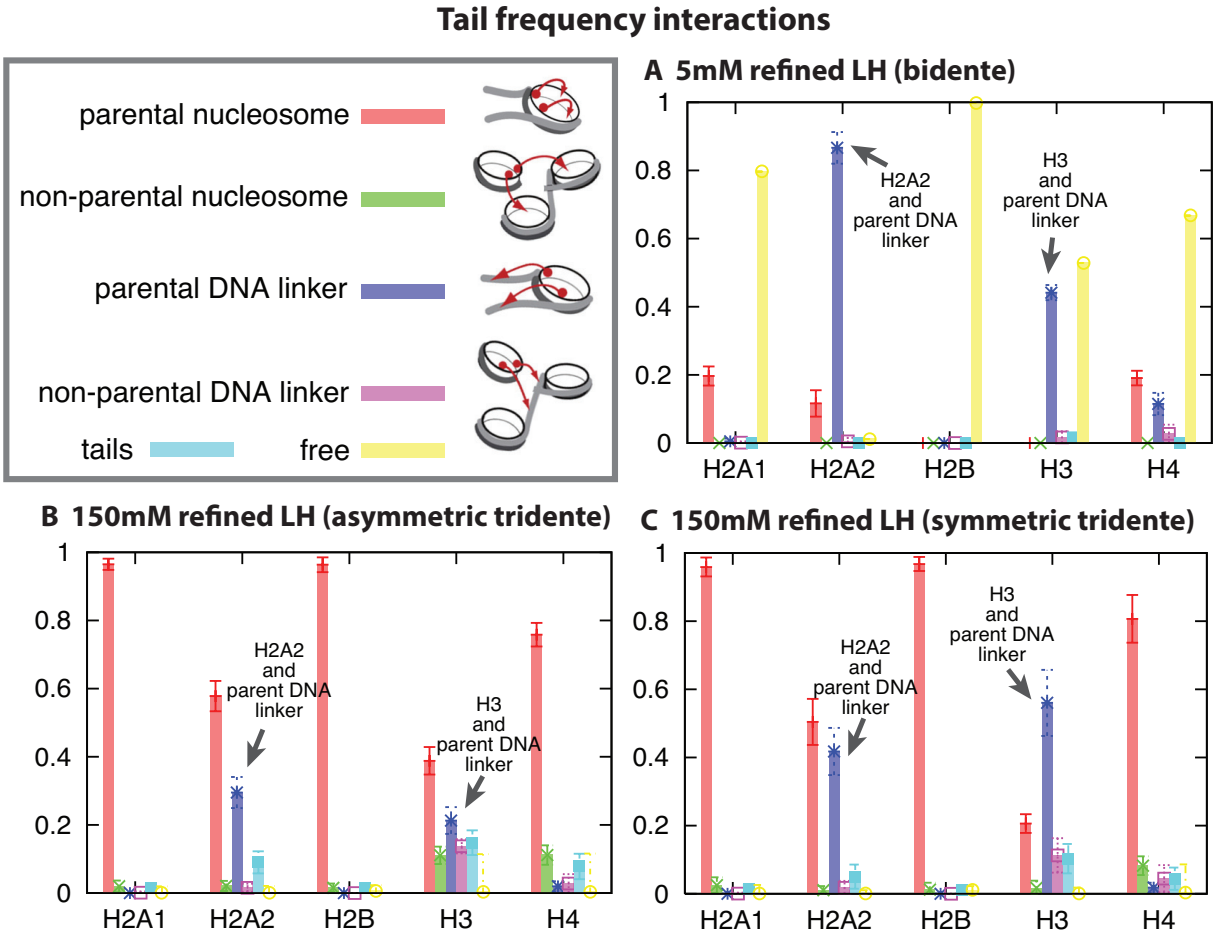

Figure S.15: Tail frequency interactions at (A) 5mM and (B) 150mM NaCl for the refined LH model as well as (C) 150mM NaCl for the prior LH model. We compute the interactions between core histone tails (see Fig. **??**) with different elements in the nucleosome: parental nucleosome (red), non-parental nucleosome (green), parental DNA linker (blue), non-parental DNA linker (magenta), other tails (cyan), and free (yellow). For the refined LH model, at low salt concentrations the chromatosome adopts a bidente configuration (see Fig. S.4); most tails are free of interactions although H2A2 strongly interacts with the parental DNA linker followed by H3. At higher salt concentrations an asymmetric tridente configuration emerges (see Fig. S.4); tails interact with the parental nucleosome rather than being free; the frequency of interaction with the parental linker DNA is lower, but H2A2 is still leading this interaction. For the prior LH model, at high salt concentrations the chromatosome adopts a symmetric tridente configuration (see Fig. S.12), where H3 interacts more strongly with the parental linker DNA than H2A2.