Biophysical Journal, Volume *116*

Supplemental Information

Nucleosome Crowding in Chromatin Slows the Diffusion but Can Promote Target Search of Proteins

Ryo Kanada, Tsuyoshi Terakawa, Hiroo Kenzaki, and Shoji Takada

Supporting Text S1

A. Energy functions and the time evolution of the system

In our target system for intra proteins that are composed of histone proteins and a nuclear protein, we used the AICG2+ model (1), in which each amino acid in proteins is represented by a CG particle located at the position of C_α-atom. The energy function is expressed as

$$
\begin{aligned}
V_{AICG2+} &= \sum_{ibd} K_{b,ibd}\left(b_{ibd} - b_{ibd,0}\right)^2 + V_{loc}^{flp} \\
&+ \sum_{j=i+2} \varepsilon_{loc,ij} \exp\left(-\frac{(r_{ij} - r_{ij0})^2}{2W_{ij}^2}\right) + \sum_{j=i+3} \varepsilon_{loc,ij} \exp\left(-\frac{(\phi_{ij} - \phi_{ij0})^2}{2W_{\phi,ij}^2}\right) \\
&+ \sum_{j>i+3}^{nat-contact} \varepsilon_{go,ij}\left[5\left(\frac{r_{ij0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij0}}{r_{ij}}\right)^{10}\right] + \sum_{j>i+3}^{non-native} \varepsilon_{exv}\left(\frac{\sigma}{r_{ij}}\right)^{12},
\end{aligned}
\tag{1}
$$

where the first, the third and fourth terms represent the structure-based virtual bond lengths b_{ibd}, virtual angle r_{ij} and virtual dihedral-angle ϕ_{ij} potentials, respectively. Each degree of freedom (b_{ibd}, r_{ij}, and ϕ_{ij}) is biased to the native structure ($b_{ibd,0}$, r_{ij0}, and ϕ_{ij0}), respectively (In the third and fourth the parameters *W* define widths of the Gaussian). The second term represents a generic flexible local potential (2), which is sequence dependent angle and dihedral-angle potentials. The fourth term is the nonlocal contact potential that stabilizes amino acid pairs, where the summation is restricted to amino acid pairs that are in contact at the native structure. The last term represents a generic excluded volume effect. The coefficients *k*, ε, and *W* except for the radius parameter $\sigma = 5.5\,\text{Å}$ related to the excluded volume term were kept as the CafeMol 2.2 default values.

For dsDNA, we used 3SPN.1 model (3) where each nucleotide in DNA contains three beads, each corresponding to base, sugar, and phosphate. The energy function is expressed as follows:

$$V_{3SPN.1} = V^{\mathrm{DNA}}_{\mathrm{bond}} + V^{\mathrm{DNA}}_{\mathrm{angle}} + V^{\mathrm{DNA}}_{\mathrm{dihedral}} + V^{\mathrm{DNA}}_{\mathrm{stuck}} + V^{\mathrm{DNA}}_{\mathrm{base}} + V^{\mathrm{DNA}}_{\mathrm{excluded}} + V^{\mathrm{DNA}}_{\mathrm{solv}} + V_{\mathrm{ele}}. \quad (2)$$

As mentioned below, in our work, all parameters except for ones related to virtual bond stretching $V^{\mathrm{DNA}}_{\mathrm{bond}}$ were kept as the 3SPN.1 default values. In $V_{3SPN.1}$, the first, second, and third term is the local-potential energy for virtual bond stretching, angle bending, and dihedral angle twisting, respectively. Concretely, the first terms $V^{\mathrm{DNA}}_{\mathrm{bond}}$ related to bond stretching is expressed as

$$V^{\mathrm{DNA}}_{\mathrm{bond}} = \sum_I k_1^{DNA}(r^I - r_0^I)^2 + k_2^{DNA}(r^I - r_0^I)^4, \quad (3)$$

where r^I is the *I*-th virtual bond length, and r_0^I is the length in the native structure, which is the canonical B-type dsDNA structure. In this work, the two parameters in 3SPN.1 related to the bond-stretching term are redefined ($K^{DNA}{}_1 = 1.839, K^{DNA}{}_2 = 0.0$). The parameters related to the remain parts ($V^{\mathrm{DNA}}_{\mathrm{stuck}}$, $V^{\mathrm{DNA}}_{\mathrm{base}}$, $V^{\mathrm{DNA}}_{\mathrm{excluded}}$, $V^{\mathrm{DNA}}_{\mathrm{solv}}$, and V_{ele}) which corresponds to the nonlocal potential energy for base stacking, Watson-Crick type base pairing, an excluded volume effect, solvation energy, and electrostatic interaction, respectively were kept as the 3SPN.1 default values as mentioned above.

Within each nucleosome, histone octamers and nucleosomal DNA maintain their native interactions via the structure-based potential as in Kenzaki et al (4). Concretely, the following Go-potential is applied:

$$V_{\text{histon octamers-nucleosomal DNA}} = \sum_{j>i+3}^{nat-contact} \varepsilon_{go,ij} \left[5\left(\frac{r_{ij0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij0}}{r_{ij}}\right)^{10} \right], \quad (4)$$

For the rest of protein-DNA interactions and protein-protein interactions are approximated by electrostatic interactions and excluded volume interactions. The electrostatic interactions for charged pairs take the Debye-Huckel formula with the ion-strength 0.21M as in Terakawa et al.

(5, 6). We assign +1e for ARG and LYS, -1e for GLU, ASP, and the phosphate in DNA, and -2 for the phosphorylated residues THR-185 and TYR-187 in the activation segment of ERK, unless otherwise noted. In the excluded volume term, we used the radius parameter $\sigma = 5.5\,\text{Å}$ as mentioned above. With the Debye-Huckel (DH) approximation for electrostatic interactions, we do not need to treat long-range interaction separately or to use the periodic boundary condition. We note that the condensed nucleosome array is highly charged, and so the DH approximation has limited accuracy.

The time evolution of our target system follows the under-damped Langevin equation:

$$m_i \frac{d^2 r_i}{dt^2} = -\frac{\partial V_{total}}{\partial r_i} - m_i \gamma_i \frac{dr_i}{dt} + m_i \xi_i \qquad (5)$$

Here, i is the coarse-grained particle ID, V is the total system potential, m is the mass for each particle, γ is the friction coefficient, ξ is white Gaussian noise that satisfies the following fluctuation-dissipation theorem:

$$\langle \xi_i(t) \rangle = 0,$$
$$\langle \xi_i(t) \xi_j(t') \rangle = \frac{2\gamma_i k_B T}{m_i} \delta(t - t') \delta_{i,j} \qquad (6)$$

where k_B is the Boltzmann constant, and T is temperature and set to 300K (room temperature). (We used the mass 10 and a low friction coefficient value 0.02 in CafeMol unit as in previous works (4, 7)).

B. Modeling of the full-length initial structures of HMGB1 and p53 tetramer.

The full-length structures of HMGB1 were prepared by adding unstructured regions to the folded region (PDB ID: 2YRQ) with MODELLER (https://salilab.org/modeller/). In the same way, the full-length structures of p53 were prepared by adding unstructured regions to the

folded region (PDB ID: 2XWR for the core domain and PDB ID: 1AIE for the TET).

In our simulations, unstructured regions for HMGB1 and p53 are treated explicitly by applying a flexible-local potential (2). For structured parts of HMGB1 and p53, we applied AICG2+ model. For p53, the native contact interactions of AICG2+ between different subunits were imposed for the TET domains based on 1AIE but not for the core domains based on 2XWR. The assumption that contact interactions between the core domains are negligible is based on the previous experimental result (8).

C. The strict definition of the characteristic times: t_1 and t_2.

To quantify the diffusion coefficient of nuclear proteins in finite chromatin environment carefully, we have to distinguish protein motions in the core and the surface/outside of the model chromatin. To this end, for each trajectory, we introduced a characteristic time when any part of the nuclear protein reaches the chromatin surface for the first time and another characteristic time when nearly all residues of the nuclear protein go out of chromatin. Specifically, the two characteristic times were defined by the first time that violates the following inequalities,

$$\left\| \vec{r}_{CM}^{TF}(t) - \vec{r}_{CM}^{Chro}(t) \right\| \le R_g^{Chro}(t) + (-1)^{\alpha} R_g^{TF}(t), \qquad (7)$$

in the trajectory. $\vec{r}_{CM}^{TF}(t)$ and $\vec{r}_{CM}^{Chro}(t)$ are the temporal coordinates of COM for the nuclear protein and for the model chromatin, respectively, while $R_g^{TF}(t)$ and $R_g^{Chro}(t)$ are the temporal radius of gyrations for the nuclear protein and chromatin, respectively. $\alpha = 1$ for t_1 and $\alpha = 2$ for t_2.

D. Maximum Δ time steps of MSD time trajectories in Fig. 4A

To compare the crowding effect by chromatin among three TFs easily, maximum Δ-time steps of MSD time trajectories in Fig. 4A (and Fig. S3) are unified at 20% of the approximate average time steps for each TF in free-solution to escape from the sphere of which radius is same as the radius of gyration for dense chromatin at initial. Concretely, the average escape time steps $T_{average}$ are expressed as

$$T_{average} = \frac{\left(R_g^{Chro}(t=0)\right)^2}{6\frac{D_1}{N_{TF}}dt}, \qquad (8)$$

where *dt* (=0.1) is the step size for time-integral of Langevin simulation, D_1 is the diffusion coefficient for one coarse-grained bead in free solution. D_1 is given by the Einstein law: $D_1 = k_B T / m\gamma$. Since N_{TF} is the system size of TF, D_1 / N_{TF} means the (approximate) diffusion coefficient for TF. Therefore maximum Δ-time steps of MSD time trajectories for TFs (ERK, HMGB1, and p53) are estimated as $8.5\text{x}10^5$, $5.0\text{x}10^5$, and $3.7\text{x}10^6$, respectively.

E. On the origin of the sub-diffusion of nuclear proteins in high density chromatin

We considered that the apparent sub-diffusion feature of protein dynamics in the crowded nucleosome in Fig. 4A may be explained by two underlying dynamical modes. Two-modes model for nuclear proteins is assumed to be expressed by the diffusional motion of a particle in the effective potential as shown by the conceptual picture in Fig. S5A. In Fig. S5A a blue particle corresponds to the COM of nuclear protein. In the effective potential which corresponds to the interaction between COM of nuclear protein and nucleosome, the spatial width characterized by L_F for each well may depend on the unoccupied space for the target proteins to move around in crowded nucleosome environment and the internal degree of freedom which is derived from the flexibility of target proteins. So, in short duration, the diffusion coefficient

characterized by D_F of COM of a target protein within a restricted area (defined by L_F of each well) is relatively large and close to one in free solution. On the other hand, in the long duration mode, the effective diffusion coefficient is smaller, because the proteins have to repeatedly dissociate from and associate to DNA. In Fig. S5A, this process is represented by a green dashed arrow with the diffusion coefficient D_S. The diffusion coefficient D_S is assumed to be smaller than D_F, because this diffusion effectively includes the hopping process between neighboring wells with finite free energy barriers.

Assuming that the two modes characterized (D_F, L_F) and D_S are mutually independent, the MSD of the COM of the protein can be expressed as:

$$MSD(t) = 6D_S t + MSD_F(t), \quad (9)$$

$$MSD_F(t) = \frac{L_F^2}{4} + 6\sum_{n=1}^{\infty} \exp\left(-D_F (n\pi/L_F)^2 t\right) \cos(n\pi/2)\left[\cos(n\pi)+1\right](L_F/n\pi)^2, \quad (10)$$

where $6D_S t$ and $MSD_F(t)$ are the MSD components derived from the slow diffusion and the fast (restricted) diffusion, respectively. We confirmed that, if $D_F \geq D_S$ is the case, the MSD curve as a function of time difference shows a convex shape, such as shown in Fig. S5B, with $L_F = 5.0$, $D_F = 1.0$, and $D_S = 0.1$. The gradient of the convex MSD curve at time zero becomes $6(D_F + D_S)$, while that in the long-time limit converges to $6D_S$, from which we can obtain the (pure) diffusion coefficient. This qualitative tendency is held for any effective length-scale L_F.

F. Exclusion for highly unwrapped nucleosome

We estimated the contact frequency between beads of the functional domains of TFs and each of beads of DNA-bp only for the wrapped nucleosomes. If the native contact fraction between histone-proteins and two terminus regions of the (one-quarter) nucleosomal DNA-bp (bp-ID [-73:-39] and [39:73]) is less than 0.05, the temporal state of the corresponding nucleosome is

un-wrapped and the contact between TF and DNA-bp in the corresponding nucleosome are not counted for contact-frequency.

Supporting Figures

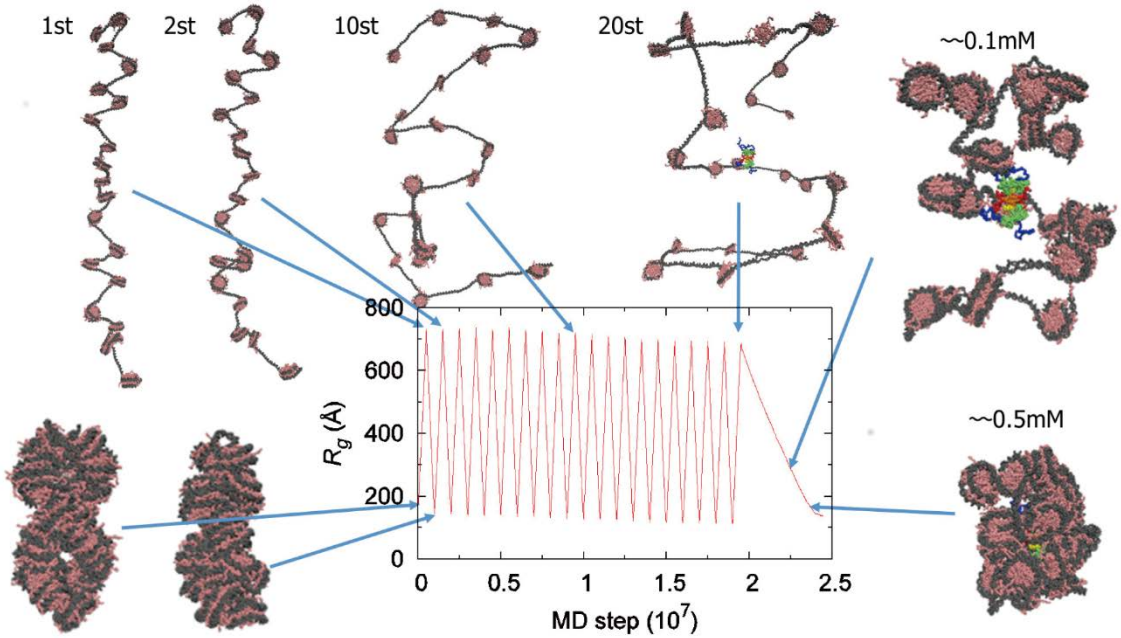

Fig. S1: Construction of the model chromatin structures. The time series of a radius of gyration, R_g, for the 20-nucleosome array, the model chromatin in the expand-and-shrink simulation are plotted together with representative snapshots. The 20-nucleosome array is enforced to expand-and-shrink with its cycle of 10^6 MD steps. In the last shrink, we embedded p53 tetramer into the 20-nucleosome array, and the radius of gyration R_g is more slowly shrunk. The last two sampled structures shown in this figure are used as the initial structure for the investigation of p53.

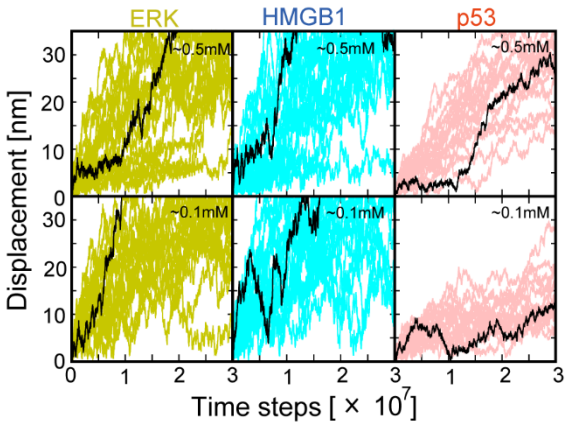

Fig. S2: Movements of nuclear proteins in chromatin environments. All time courses of displacements for centers of mass (COM) of nuclear proteins in the high-density (upper panel) and the low-density (lower panel) chromatins; ERK (yellow), HMGB1 (cyan), and p53 (pink). The black line in each panel shows the representative trajectory in Fig. 3A.

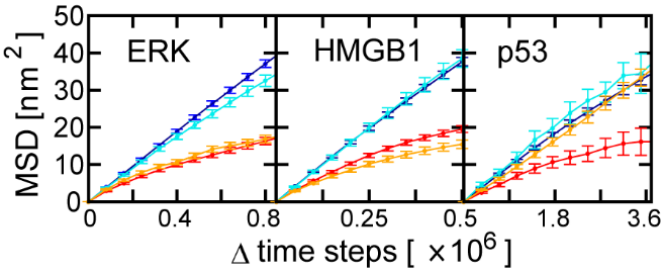

Fig. S3: The comparison of the mean square displacements (MSDs) of nuclear proteins between those inside chromatin and those around the surface of chromatin. The blue and red circles with error-bars are MSD plots within intra-chromatin, while the cyan and orange circles with error bars correspond to MSD plots around the surface of chromatin. The blue and cyan are ones in the low-density chromatin, while the red and orange correspond to ones in the high-density chromatin. The error bars are defined as the standard errors.

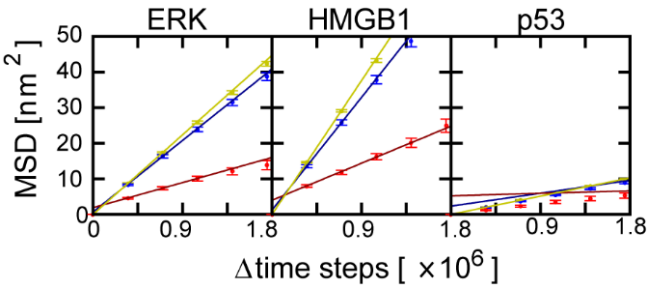

Fig S4: The mean square displacements (MSDs) of nuclear proteins at the same Δtime steps. Each color for circles with error-bars is same as Fig. 4. The straight lines represent the linear fit line which is also same as Fig. 4. The error bars are defined as the standard errors.

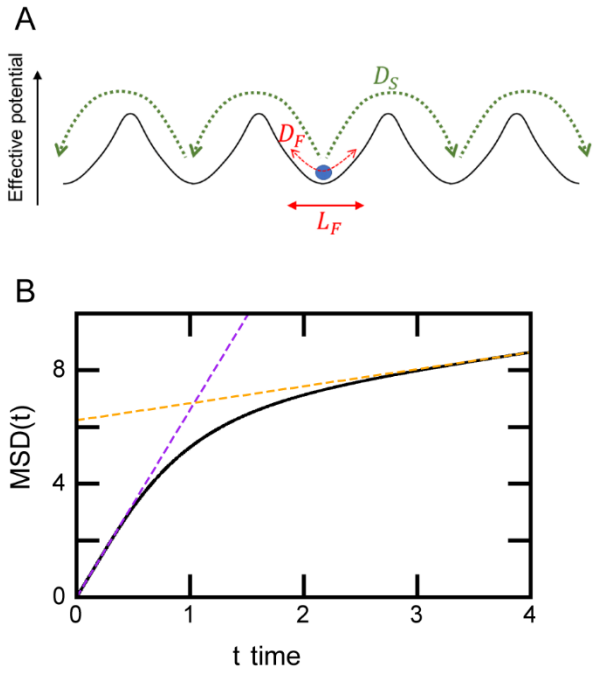

Fig. S5: (A) The schematic view of the two-mode model of TF diffusion in crowded space. The two-mode model may be expressed by the diffusional motion of a particle in effective potential. The blue particle represents the COM of the TF molecule. The width L_F is the effective distance for the restricted diffusion of the COM of the TF. The short-time diffusion around the trapping center, which includes the internal structural-fluctuation of the TF, is characterized by a larger diffusion coefficient D_F. The long-time diffusional movement of the COM of the TF includes hopping between neighboring wells (green dashed curves), which corresponds to a smaller diffusion coefficient D_S. (B) The analytical solution of MSD for the two-mode model, where the parameters are set as L_F=5.0, D_F=1.0, and D_S = 0.1. The black solid lines show the analytical solution of MSD expressed by Eq. 9 and Eq. 10. The purple (around t=0) and orange (around t=4) dashed-lines corresponds to the linearly asymptotic lines of which gradients are $6(D_F + D_S)$ and $6D_S$, respectively.

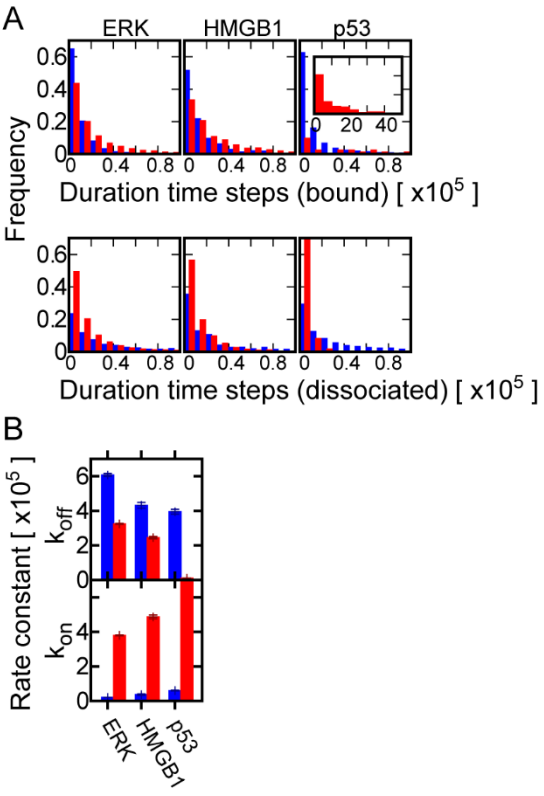

Fig. S6: The kinetics of protein-DNA binding. (A) The histogram of the duration time steps for the bound state (the upper panel) and the dissociated state (the lower panel) between proteins and DNA. The inset for p53 shows the distribution at the extended time steps. (B) The estimated rate constants k_{off} and k_{on} for the dissociation and the binding. In each panel, the red and blue indicate the results for the high- and low- density chromatin, respectively. The error bars are defined as the standard errors.

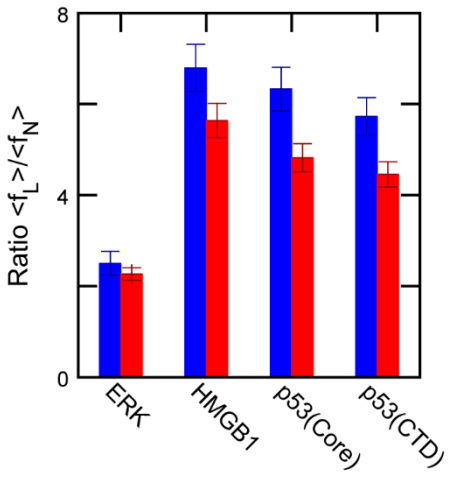

Fig. S7: The ratio of the average contact frequencies in the linker-DNA: $\langle f_L \rangle$ relative to those in nucleosomal-DNA: $\langle f_N \rangle$. The blue (red) bars correspond to the result for the low (high) density chromatin. The error bars are defined as the standard errors.

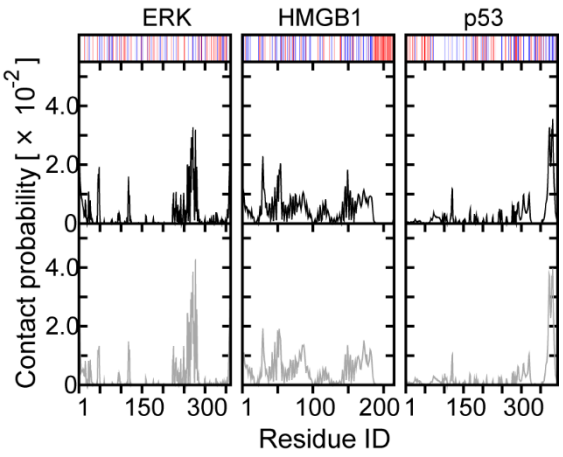

Fig. S8: Contact probabilities (= normalized frequencies) of protein residues with DNA. The black and gray curves correspond to the results for high-density and low-density chromatins, respectively. The upper panel with blue (+1) and red (-1) bars shows the charge distribution for each TF.

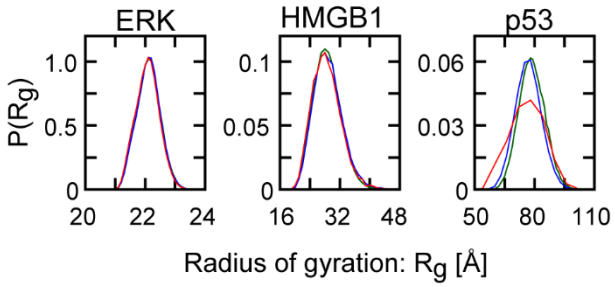

Fig. S9: The probability distribution of radius of gyration R_g for nuclear proteins in chromatins. (A) In each panel, the green, blue, and red curves indicate the distributions at [nucleosome] =0.0 mM, 0.1mM, and 0.5mM, respectively.

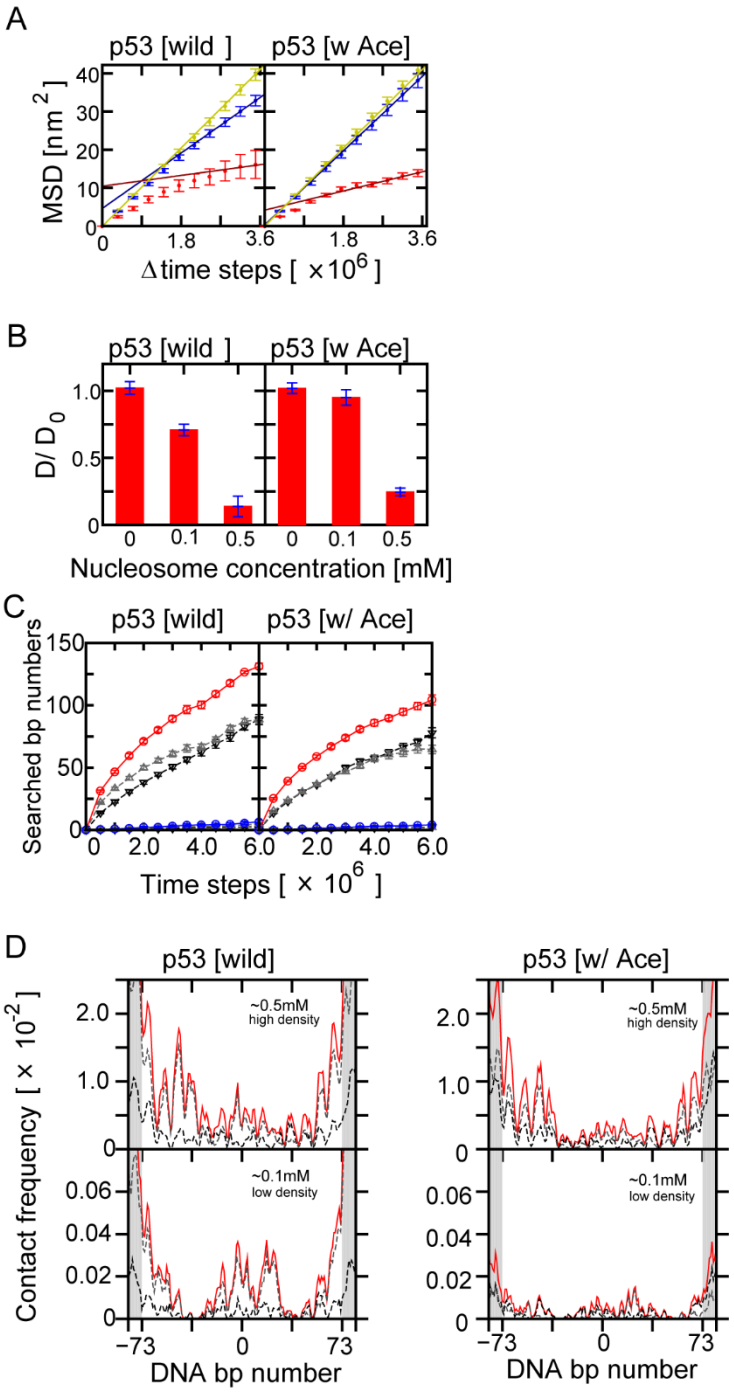

Fig. S10: The comparison between wild-type p53 [wild] and the partially acetylated p53 [w Ace]. (A) MSDs, (B) the diffusion coefficients, (C) DNA-bp searching speeds, and (D) the contact-frequency profiles between p53 and DNA. The coloring rules and the meanings of panels (upper/lower) for panels (A-B), and (D) are same as ones for Fig. 4 A-B, and Fig. 6B,

respectively. In panels (C), the numbers of base pairs associated with the core and CTD are shown by black and gray dashed curves, respectively, together with that by a monomer of p53, The color code in panels (C) is the same as Fig. 5. The error bars are defined as the standard errors.

Supporting Movie Legends

Movie S1-S6: Movie S1-S3 and Movie S4-S6 are representative movie files in low- and high-density chromatins, respectively. Movie S1 and S4 are for ERK, Movie S2 and S5 are for HMGB1, and Movie S3 and S6 are for p53.

Supporting References

1. Li, W.F., T. Terakawa, W. Wang, and S. Takada. 2012. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. Proc. Natl. Acad. Sci. U. S. A. 109: 17789–17794.
2. Terakawa, T., and S. Takada. 2011. Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: p53 N-Terminal Domain. Biophys. J. 101: 1450–1458.
3. Sambriski, E.J., D.C. Schwartz, and J.J. de Pablo. 2009. A Mesoscale Model of DNA and Its Renaturation. Biophys. J. 96: 1675–1690.
4. Kenzaki, H., and S. Takada. 2015. Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. PLOS Comput. Biol. 11: e1004443.
5. Terakawa, T., H. Kenzaki, and S. Takada. 2012. p53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains. J. Am. Chem. Soc. 134: 14555–14562.
6. Terakawa, T., and S. Takada. 2015. p53 dynamics upon response element recognition explored by molecular simulations. Sci. Rep. 5: 17107.
7. Takada, S., R. Kanada, C. Tan, T. Terakawa, W. Li, and H. Kenzaki. 2015. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. Acc. Chem. Res. 48: 3026–3035.
8. Tidow, H., R. Melero, E. Mylonas, S.M. V Freund, J.G. Grossmann, J.M. Carazo, D.I. Svergun, M. Valle, and A.R. Fersht. 2007. Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. Proc. Natl. Acad. Sci. U. S. A. 104: 12324–12329.