Biophysical Journal
Article

Biophysical Society

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

Ryo Kanada,[1,2] Tsuyoshi Terakawa,[1] Hiroo Kenzaki,[1,3] and Shoji Takada[1,*]
[1]Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto, Japan; [2]Compass to Healthy Life Research Complex Program, Cluster for Science, Technology and Innovation Hub, RIKEN, Kobe, Japan; and [3]Information Systems Division, Head Office for Information Systems and Cybersecurity, RIKEN, Saitama, Japan

ABSTRACT Dynamics of nuclear proteins in crowded chromatin has only been poorly understood. Here, we address the diffusion, target search, and structural dynamics of three proteins in a model chromatin using coarse-grained molecular simulations run on the K computer. We prepared two structures of chromatin made of 20 nucleosomes with different nucleosome densities and investigated dynamics of two transcription factors, HMGB1 and p53, and one signaling protein, ERK, embedded in the chromatin. We found fast and normal diffusion of the nuclear proteins in the low-density chromatins and slow and subdiffusional movements in the high-density chromatin. The diffusion of the largest transcription factor, p53, is slowed by high-density chromatin most markedly. The on rates and off rates for DNA binding are increased and decreased, respectively, in the high-density chromatin. To our surprise, the DNA sequence search was faster in chromatin with high nucleosome density, though the diffusion is slower. We also found that the three nuclear proteins preferred to bind on the linker DNA and the entry and exit regions of nucleosomal DNA. In addition to these regions, HMGB1 and p53 also bound to the dyad.

SIGNIFICANCE Chromatin is a highly crowded molecular environment, in which DNA-binding proteins, such as transcription factors, need to search for their recognition sequence from genomic DNA. This is apparently a formidable task for proteins. Thus, how nuclear proteins can move in crowded chromatin and search DNA sequence is an interesting question, but not much is known at molecular level. The study addresses this issue by coarse-grained molecular dynamics simulations. We investigated diffusional motions and target DNA search dynamics of nuclear proteins in small chromatin environment. To our surprise, we found that the crowded nucleosome environment slows the diffusional motion of proteins but can speed up the target search process.

## INTRODUCTION

Chromatin accessibility regulates transcriptional activity in eukaryotic cells (1,2). Often, chromatin accessibility is characterized at a genomic scale by DNase assays (3). However, these assays do not directly measure chromatin structures or diffusional dynamics of nuclear proteins. Recently, the diffusional dynamics of EGFPs and transcription factors (TFs) in interphase nuclei and mitotic chromosomes were observed via fluorescence correlation spectroscopy and live-cell imaging approaches (4) together with Monte Carlo simulation, elucidating the fluctuation of individual nucleosomes that enhances chromatin accessibility for TFs (5,6). In these studies, both experiments and simulations have limited spatial resolutions, which precludes the precise characterization of the nuclear protein movements. For example, in the Monte Carlo simulation, nucleosomes and TFs were treated as hard spheres without intramolecular flexibility. The purpose of this study is to address the diffusional and the target search dynamics of nuclear proteins in chromatin using molecular dynamics (MD) simulations with intramolecular flexibility.

To investigate the dynamics of proteins in chromatin, we need chromatin models with a suitably large number of nucleosomes. Using modern supercomputers, we can perform MD simulations of increasingly large systems, such as an entire cytoplasm, with all of the 103 million atoms explicitly treated (7). However, the time that can be simulated is highly limited and not enough for the TFs and chromatin

*Submitted December 31, 2018, and accepted for publication May 1, 2019.*
*Correspondence: takada@biophys.kyoto-u.ac.jp
Ryo Kanada and Tsuyoshi Terakawa contributed equally to this work.
Editor: Tamar Schlick.
https://doi.org/10.1016/j.bpj.2019.05.007

components containing disordered segments with long equilibration time. To overcome this issue, we performed coarse-grained MD simulations without explicitly treating all the atoms. Many coarse-grained MD simulations of chromatin have been performed using coarse-grained models with different resolutions (8–31).

In this study, we used coarse-grained models in CafeMol (32,33), in which one particle represents each amino acid in proteins, whereas three particles represent sugar, phosphate, and base units of each nucleotide in DNA (32,33). Using these models, one can approximate physicochemical interactions accurately in the flexible molecules and simulate long time-scale dynamics within reasonable computation time. For proteins, we used the AICG2 + model, which incorporates interactions between atoms in the native structure into energy functions of the coarse-grained model (34). DNA was modeled by the 3SPN.1 model (19). This combination of coarse-grained models has been used for several protein-DNA complexes, including p53 dynamics along double-stranded DNA (22,35), DNA unwrapping from the nucleosome (20), trinucleosome folding (23), and DnaA oligomerization in the bacterial DNA replication initiation complex (25). Notably, even with the coarse-grained model, the simulation of the protein dynamics in chromatin is challenging. In this study, using the K computer—a massively parallelized supercomputer—we simulated nuclear protein dynamics in chromatin that contains 20 nucleosomes (molecular weight ~3199 kDa). The K computer involves more than 80,000 nodes and achieved over 10 peta-FLOPS in LINPACK benchmark in 2012, at which it was the top record (36). CafeMol is implemented with single instruction, multiple data to optimize for the K computer. By using CafeMol, the users can easily perform coarse-grained simulations of proteins and nucleic acids without detailed knowledge of models or coding.

In this work, we focus on the dynamics of three nuclear proteins: 1) a signaling protein ERK as an example of a non-TF; 2) a smaller TF, HMGB1; and 3) a larger TF, homotetrameric p53. Recent chromatin electron microscopy tomography (chromEMT) data elucidated that chromatin does not have a regular 30-nm fiber but takes an irregular folding structure (37), as was anticipated before (38,39). As a model chromatin environment, simulating a 20-nucleosome array, we constructed irregular folding structures that mimic the chromEMT image. Simulating three nuclear proteins in the model chromatin environment, we found that the diffusion of the largest TF, p53, is slowed by high-density nucleosomes in a most dramatic way, but the DNA sequence search of it is faster.

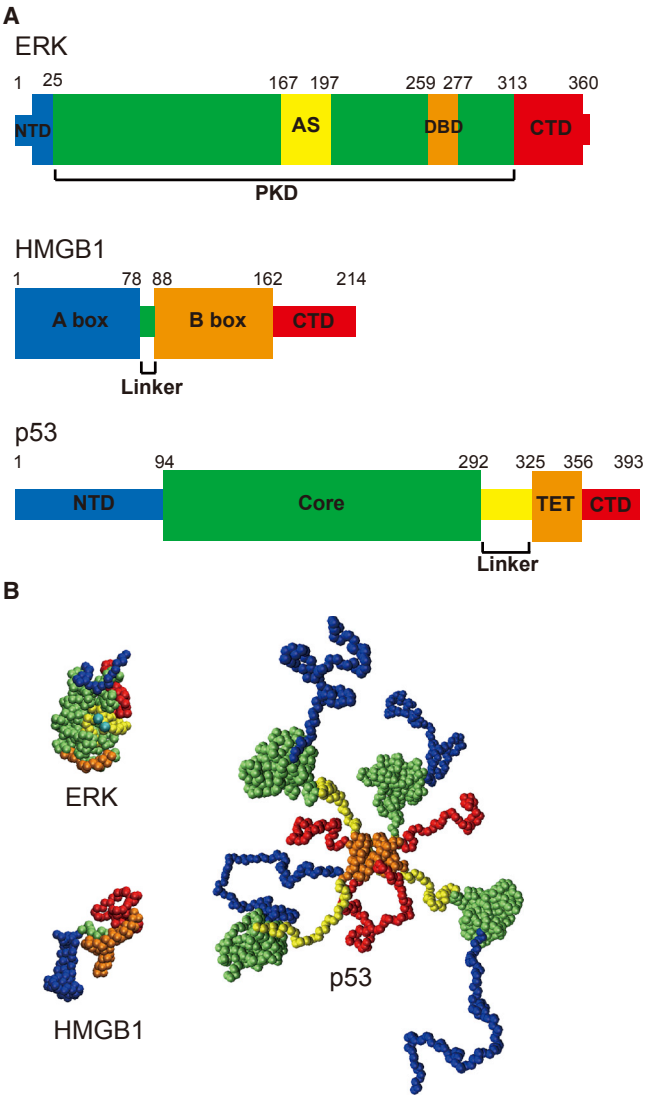


FIGURE 1 The three nuclear proteins studied, ERK, HMGB1, and p53. (*A*) The domain compositions are shown. The thick bands stand for globular domains, and the thin bands correspond to linker and/or intrinsically disordered regions. (*B*) Snapshots of nuclear proteins sampled by coarse-grained simulations are given. Note that p53 is a homotetramer. The colors in the snapshots correspond to the colors in (*A*). Three snapshots are drawn with the same scales. NTD, N-terminal domain; CTD, C-terminal domain; PKD, protein kinase domain; AS, activation segment; DBD, DNA-binding domain; TET, tetramerization domain. To see this figure in color, go online.

## MATERIALS AND METHODS

### Three nuclear proteins studied

The domain maps and the molecular structures of our target TFs are shown in Fig. 1. The signaling protein ERK (Fig. 1 *A*: *top*, *B*: *left-top*) is a globular protein composed of three domains: NTD (N-terminal domain), PKD (protein kinase domain), and CTD (C-terminal domain) (40). PKD is the enzymatic domain containing the activation segment (AS) in which THR185 and TYR187 are phosphorylated by MEK protein (40). The phosphorylation of ASs makes ERK active. Then, the activated ERK moves into the nucleus and phosphorylate the downstream TFs such as c-Fos and Elk-1 (40). In our simulation, the charges for THR185 and TYR187 are set to −2, mimicking the activated state. Although, in this study, ERK is merely an example of non-TF proteins, the recent experiment (41) suggested that PKD in ERK has DNA-binding domains (DBDs): residues 259–277.

HMGB1 (Fig. 1 *A*: *middle*, Fig. 1 *B*: *left-bottom*) is an example of a small TF. It contains 214 amino acids. In HMGB1, two structured domains—the

A-box and B-box, which are responsible for DNA-binding—are connected by a flexible linker. Also, it has a disordered NTD that includes an acidic tail (42).

p53 (Fig. 1 *A*: *bottom*, Fig. 1 *B*: *right*) is a famous tumor suppressor that normally forms a homo-tetramer in living cells (43). p53 is the largest (1572 amino acids as the homotetramer) in our targets. Each monomer of p53 consists of four domains: NTD, core domain, tetramerization domain (TET), and CTD (44). Because the folded domains, core and TET, are connected by the intrinsically disordered regions, the entire p53 molecule stochastically takes various conformations (45). The core domain of p53 is responsible for specific DNA binding (46,47), whereas CTD is responsible for nonspecific DNA-binding (48).

## Model chromatin

The model chromatin system contains 20 nucleosomes connected by linker DNA strands of 20 basepairs. Each nucleosome is that from *Xenopus laevis*, of which the crystal structure is taken from the Protein Data Bank (PDB) entry of 1KX5. Each nucleosome is flanked by 10-basepair linker DNA: 5′-ACTTACATGC-[nucleosomal DNA]-ACTTACATGC-3′. This linker DNA length and sequence are from that in the tetranucleosome crystal structure (49). Twenty copies of this fragment are tandemly connected to form the 20-nucleosome array.

## Coarse-grained models and energy functions

We used CafeMol and applied a coarse-grained model to investigate the dynamical behavior of the nuclear proteins in these nucleosomes. For proteins, each amino acid is represented by one bead, whereas each nucleotide in DNA is approximated by three beads, each representing sugar, phosphate, and base. Within each nucleosome, the globular part of histone octamers and nucleosomal DNA maintain their native interactions via the native-structure-based potential as in (20) (see Supporting Materials and Methods for more details). Interactions within globular proteins are residue-dependent and are mapped from atomic interactions in the reference nucleosome structure. Histone tails are treated as flexible polymers with sequence-based statistical potential for bond and torsion angles (50). The DNA model includes sequence-dependent basepair interactions, but the sequence dependence of the bending rigidity is not fine-tuned (19). The sum of electrostatic and excluded volume interactions approximated the other protein-DNA and protein-protein interactions. The electrostatic interactions for charged pairs take the Debye-Huckel formula with an ion strength of 0.21 M as in (22,35). We assign $+1e$ for ARG and LYS, $-1e$ for GLU, ASP, and the phosphate in DNA, and $-2e$ for the phosphorylated residues THR185 and TYR187 in the activation segment of ERK. For the acetylated LYS residues in p53, we set their charges to 0. In the excluded volume term, we used the radius parameter $\sigma = 5.5$ Å. Other main parameters were kept as the CafeMol 2.2 default values.

The reference structures for the globular domains of ERK, HMGB1, p53 core domain, and p53 TED were taken from PDB entries 2GPH, 2YRQ, 2XWR, and 1AIE, respectively. (See Supporting Materials and Methods for how to get full-length initial structures of HMBG1 and p53 tetramer.) For nucleosomes, we used 1KX5 as the reference structure.

## Construction of the chromatin models and initial placement of nuclear proteins

Recent experiments clarified that chromatin does not form regular 30-nm-fiber structures but folds into irregular structures (37,39). Here, we tried to model such irregular chromatin structures. First, we prepared a two-start helix configuration, one model of the 30-nm-fiber structure, by repeatedly connecting the first two nucleosome structures of the tetranucleosome (PDB: 1ZBB). Then, by biasing the radius of gyration of the two-start helix configuration, we conducted an expansion-and-shrink simulation. In this simulation, we alternately added the following biasing potential,

$$\pm k_g R_g^{\mathrm{Chro}}, \tag{1}$$

where $R_g^{\mathrm{Chro}}(t)$ is the radius of gyration of the chromatin system. Negative- and positive-signed potentials with $k_g = 100$ are alternately applied to the system 19 times to expand and shrink the model chromatin, respectively. Then, we applied the negative-signed potential to the chromatin once more to obtain an expanded structure. The time course and the snapshots of this simulation are depicted in Fig. S1. During the repeated expansion and shrink, the chromatin gradually lost the regular two-start helix configuration.

To investigate the dynamics of nuclear proteins in chromatin, we need to place the nuclear proteins inside chromatin. For this purpose, we put p53 near the COM of the chromatin in the most expanded structure after the expansion-and-shrink simulation. Subsequently, we slowly decreased the radius of gyration of the chromatin with a smaller positive potential ($k_g = 10$) for $5 \times 10^6$ time steps, leading to the structures in which p53 is embedded into the chromatin with various nucleosome density. We used them as the initial structure of the p53 simulations. For the other targets, p53 in these structures was replaced with HMGB1 and ERK, respectively. By replacement, there could be some additional space. However, we found that contacts between the replaced proteins and DNAs are immediately recovered so that this replacement would not affect the subsequent simulations.

The nucleosome density in chromatin depends on genomic loci, phases of the cell cycle, and other regulatory factors. As chromatin models, we put 20 nucleosomes into the model with two distinct densities. For the high-density (nucleosome density ~0.5 mM) structure, we set the maximal radius of the chromatin, $R_{max}$, to 261 Å based on the assumption that 20 nucleosomes uniformly distribute. This density roughly corresponds to that in compact clusters in the mitotic chromosome and likely corresponds to interphase heterochromatin (5). For the low-density (0.1 mM) structure, we set $R_{max}$ to 458 Å. We chose 0.1 mM as a low density because this has been used in previous work (5). A rough estimate of the percentage occupied volumes for high and low densities is 10 and 2%, respectively. From the p53-embedded slow-shrink simulation, we picked up the structures with the target densities (Fig. 2).

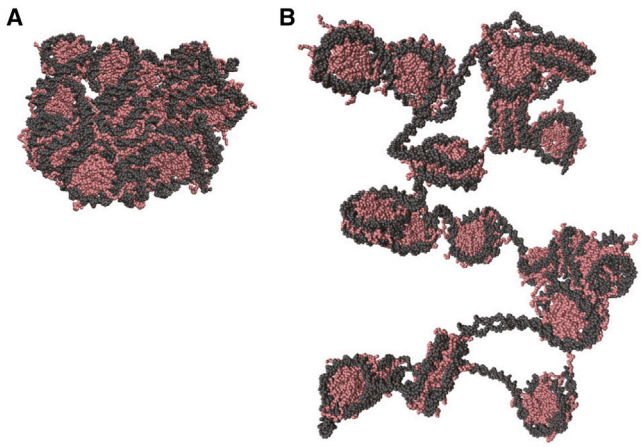


FIGURE 2 Two chromatin states used as an environment. Snapshots of the 20-nucleosome array in the high-density (*A*) and the low-density (*B*) chromatin environments are given. The DNA and the histone proteins are in gray and pink, respectively. (*A*) The high-density state with [nucleosome] ~0.5 mM and (*B*) the low-density state with [nucleosome] ~0.1 mM are shown. To see this figure in color, go online.

### Time evolution by coarse-grained simulations

Here, we focus on the dynamics of three nuclear proteins, ERK, HMGB1, and p53, in chromatin with the high- and low-density nucleosomes. For each of these six setups, we conducted 20 Langevin dynamics simulations for $3 \times 10^7$ time steps at a temperature of 300 K. We used a low friction coefficient value of 0.02 in CafeMol unit as in previous works (20). Our rough mapping in timescale suggests that one MD step corresponds to ~1 ps (33). Thus, the duration of each trajectory is ~30 $\mu$s. To prevent the change of nucleosome density, the harmonic potential (the spring constant is set to 500 kcal/mol Å$^2$) restrained the center of mass (COM) of each nucleosome to its initial position, though this does nothing until the COM deviates $\geq$30 Å away from the initial position. Assumed here is that the added nuclear proteins would not change the overall topology or the nucleosome density of the chromatin.

## RESULTS

### Dynamics of nuclear proteins in chromatin

First, we address the diffusional dynamics of nuclear proteins in the model chromatin environments. We used three nuclear proteins: a signaling protein ERK, a small TF HMGB1, and a larger TF p53. As the chromatin environments, we prepared the two states with distinct densities; the high-density state (Fig. 2 *A*) and the low-density state (Fig. 2 *B*) as described above. We embedded each of the three nuclear proteins near the COM of the two model chromatins (*leftmost cartoons* in Fig. 3, *B–G*), resulting in the six setups. For each of the six setups, we conducted 20 independent simulations with different stochastic forces.

Fig. 3 *A* shows representative time courses of the COM displacements of nuclear proteins in chromatin from the initial position (see Fig. S2 for all the trajectories). Representative snapshots are depicted in Fig. 3, *B–G* (see the corresponding video files Video S1, S2, S3, S4, S5, and S6. Video S1 and S4 are for ERK, Video S2 and S5 are for HMGB1, and Video S3 and S6 are for p53 in high- and low-density chromatins, respectively).

In the case of high nucleosome density (Fig. 3 *A*, *top*), HMGB1 diffused most rapidly, which was followed by ERK. p53 showed the slowest diffusion. The higher the molecular weight is, the slower the diffusion is. We also note that the diffusion is relatively slow at the beginning and is accelerated in the middle of the time courses. The snapshots (Fig. 3, *B*, *D*, and *F*) suggest that the acceleration started when the nuclear protein reached the surface of the model chromatin. To clarify this event in the trajectories, we introduced a characteristic time $t_1$ when the nuclear protein touches the chromatin surface for the first time and another characteristic time $t_2$ when nearly the entire nuclear protein exits from the model chromatin environment (see the precise definition in Supporting Materials and Methods). We indicated these characteristic times in Fig. 3 *A* and depicted snapshots at these times in Fig. 3 *B*, *D*, and *F* for the high-nucleosome-density case.

In the case of the low-density chromatin environment (Fig. 3 *A*, *bottom*), we find qualitatively similar behavior. The diffusion was the fastest for HMGB1 and the slowest for p53 among the three proteins. In the case of p53, the molecule did not reach the surface of the model chromatin by the end of the simulation. Thus, we depicted two snapshots at arbitrarily chosen times indicated as $t_1'$ and $t_2'$ in Fig. 3 *G*.

We also note that irregular diffusion was observed for the time range $[t_1:t_2]$ at which the target protein resides on the surface of the model chromatin (Fig. S3).

### Quantifying the diffusion coefficient

From the trajectories, we estimated the diffusion coefficients of the three nuclear proteins, ERK, HMGB1, and p53, in chromatin with high and low nucleosome densities. As a control, we also estimated the diffusion coefficients without nucleosome. We calculated the mean-square displacement (MSD) of the COM of nuclear proteins as a function of duration, $\Delta$time steps (Fig. 4 *A*; Fig. S4). To reduce artifacts from the surface effect, we used the trajectory data up to $t_1$. (See Supporting Materials and Methods for the detailed explanation). We set the upper limit of $\Delta$time steps so that the MSD without nucleosome does not exceed 40 nm$^2$ (*yellow curves* in Fig. 4 *A*).

Fig. 4 *A* shows that the low nucleosome density did not markedly affect the MSD of any targets (Fig. 4 *A*, *blue curves*). The MSD was approximately proportional to $\Delta$time steps, which is a hallmark of the normal diffusion. In particular, the MSDs for ERK and HMGB1 in chromatin with low nucleosome density were nearly identical to those in the solution.

On the other hand, high nucleosome density significantly altered the MSD curves of all the three nuclear proteins (Fig. 4 *A*, *red curves*). The MSD plot as a function of $\Delta$time steps was convex-shaped, which is consistent with the anomalous diffusion. Given the limited size of the model chromatin, flexible nature of nuclear proteins, and limited timescale of simulations, quantifying the diffusion coefficient is not trivial.

Visual inspection suggests that some parts of target proteins are often bound to nucleosomes and linker DNAs at a short timescale. At a longer timescale, the proteins dissociate from DNA and bind to neighboring DNA. Thus, we hypothesized that the diffusion of the proteins contains two distinct modes. In the short duration, although parts of the protein are bound to DNA, other unbound parts can move with restraint, resulting in the diffusion of the COM. The diffusion in the unbound parts is similar to that in solution and has a large diffusion coefficient. In the long duration mode, the diffusion coefficient is smaller because the proteins repeatedly dissociate from and associate to DNA. This two-mode model results in the convex MSD curves (see Fig. S5; Supporting Materials and Methods) and thus is consistent with the MSD plots obtained with high nucleosome density (*red curves* in Fig. 4 *A*). We note that the

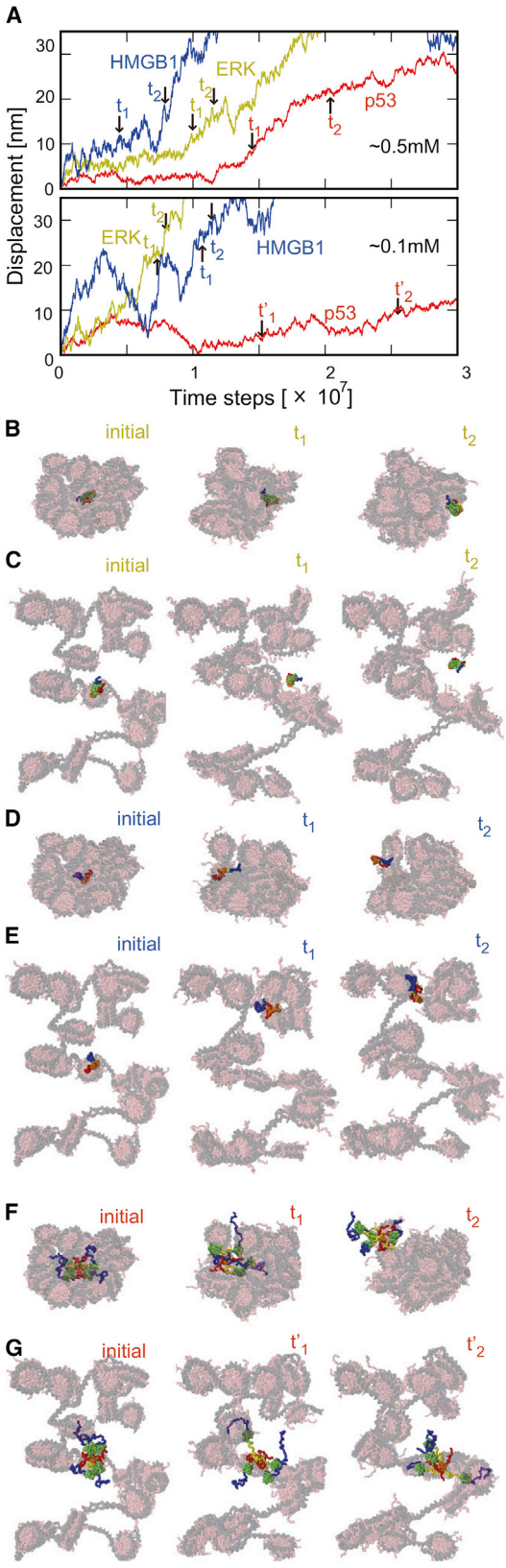


FIGURE 3 Movements of nuclear proteins in chromatin environments. (*A*) Representative time courses of displacements for COMs of nuclear proteins in the high-density (*upper panel*) and the low-density (*lower panel*) chromatins are shown: ERK (*yellow*), HMGB1 (*blue*), and p53 (*red*). (*B–G*) Snapshots at three times are given: $t = 0$ (*left*), $t_1$ (*middle*), and $t_2$ (*right*). The times $t_1$ and $t_2$ are labeled in (*A*) with black arrows. For p53 in high-density chromatin, $t_1'$ and $t_2'$ are arbitrarily chosen. The snapshots in (*B*)–(*G*) are for ERK, HMGB1, and p53, respectively. (*B*), (*D*), and (*F*) are for the high-density cases, whereas (*C*), (*E*), and (*G*) are for the low-density cases. The color usage for nuclear proteins is the same as that in Fig. 1, and chromatin is drawn transparent to highlight the proteins. To see this figure in color, go online.

contribution of the first mode is larger for the cases of HMGB1 and p53 than for ERK, perhaps because the formers have flexible linkers at their central regions. Thus, we estimated the diffusion coefficient using the long duration part of the MSD curves. Using the fourth quarter of the duration range in Fig. 4 *A*, we obtained the diffusion coefficients by the linear fitting.

From this analysis, the question we asked is how the protein diffusion is retarded by the high nucleosome density. In Fig. 4 *B*, we showed the relative diffusion coefficients ($D/D_0$) for the three nuclear proteins in three environments, where $D_0$ is the analytical diffusion coefficient in the solution (see Supporting Materials and Methods).

In the low-density chromatin environment, the diffusion coefficient for ERK seems to be slightly smaller than $D_0$. The diffusion coefficients for the two TFs, HMGB1 and p53, were moderately smaller than $D_0$. Thus, the motion of ERK was not sensitive to the presence of chromatin, whereas the diffusion dynamics of other TFs tend to be retarded by sequence-nonspecific interactions to chromatin, primarily DNA.

In chromatin with high nucleosome density, the diffusion coefficients of all the three proteins were markedly reduced. In this condition, nucleosomes served as steric obstacles to some extent and thus retarded diffusion of all the proteins. Particularly, the diffusion of p53 was most strikingly slowed. Given that p53 is larger than the HMGB1, the size could be a reason for this difference. Alternatively, the stronger affinity of p53 to DNA than HMGB1 could make the diffusion of p53 slower than HMGB1 (51) (Fig. S6). These two factors are not mutually exclusive.

## DNA sequence search by nuclear proteins in chromatin environment

From the functional point of view, TFs need to search their target sequences efficiently in genomic DNA. We estimate the rate of sequence search by the three nuclear proteins. We counted the cumulative numbers of DNA basepairs that were "searched" by the protein as a function of time. Here, the "searched" is defined by the detected interactions between DNA particles and amino acids in the known DBDs: the DBD of ERK, the A-box and B-box of HMGB1, and the core and CTD of p53. The cutoff distance of the interaction between a bead of DNA and an amino acid in DBDs of proteins was set to 8.0 Å.

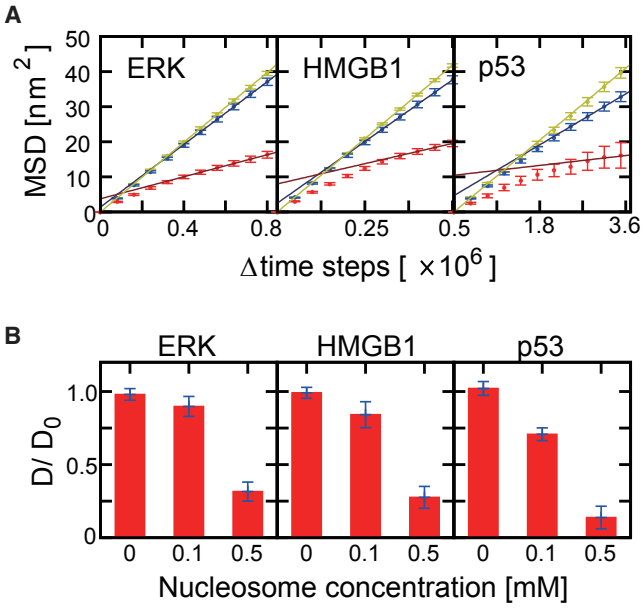


FIGURE 4 MSD and diffusion coefficients. (*A*) The mean-square displacement (MSD) is shown as a function of the time difference for three proteins. The yellow, blue, and red closed circles are for [nucleosome] ~0.0, 0.1, and 0.5 mM, respectively, with error bars that are defined as standard errors. The straight lines represent the linear fit using the fourth quarter of the time window. (*B*) Relative diffusion coefficients estimated for three proteins in three environments relative to the solution environment are shown: [nucleosome] ~0.0, 0.1, and 0.5 mM. The diffusion coefficient $D$ is normalized by the analytical diffusion coefficient $D_0$ for each protein in free solution. To see this figure in color, go online.

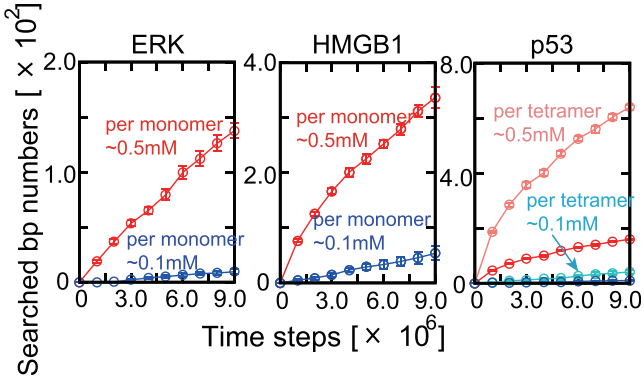


FIGURE 5 The cumulative number of DNA basepairs interacted with by nuclear proteins. Each panel shows the numbers of basepairs associated with DBDs of ERK, HMGB1, and p53 tetramer, respectively: the DBDs are the DBD of ERK, the A-box and B-box of HMGB1, and the core and CTD of p53. In each panel, red and blue symbols are for the number per monomer in high density and low density, respectively. For the number per tetrameric p53, we also plotted in light red and cyan symbols for the corresponding numbers in high- and low-density chromatins, respectively. Note that the scales in the vertical axes are different in the three proteins. The error bar for each symbol is defined as the standard error. To see this figure in color, go online.

Of the two DBDs of p53, only the core domain is suggested to "recognize" the sequence. Here, for comparison, we used both domains of p53 in the analysis. Also, p53 remains as a homotetramer during the simulation. Thus, there can be two ways to count the searched basepairs: either per monomer or tetramer.

Fig. 5 plots the cumulative numbers of searched DNA basepairs averaged over 20 trajectories. Remarkably, for each of the target proteins, the number of searched basepairs increased more rapidly in the case of the high nucleosome density than low. This tendency is the opposite to the diffusion coefficient described above. In chromatin with high nucleosome density, DNA is more crowdedly packaged. Thus, the protein can search the DNA sequence more rapidly even though it diffuses slowly. This tendency is especially apparent for ERK and HMGB1: in the long time regime, the numbers of the searched basepair grow linearly, suggesting that the simulation reaches the stationary phase. In this phase, the numbers of the newly searched basepairs per time (the slope of the curves) were significantly more abundant in chromatin with high nucleosome density than low. Unexpectedly, the ratio of the slopes in high and low density is larger than the ratio of the density, i.e., 5, in the case of ERK. This result suggests that the ERK tends to bind to chromatin more stably with the high nucleosome density. This could be attributed to the macromolecular crowding effect that increased the nonspecific binding between the target protein and DNA. In the case of HMGB1, this effect could not be detected. Because HMGB1, as a TF, has higher affinity to DNA and nucleosome compared to ERK, as a signaling protein (Fig. S6), the macromolecular crowding effect may be weak.

For p53, the cumulative searched basepair numbers are clearly curved as a function of time in the short time regime, whereas in the long time regime, the asymptotic behavior seems to become linear-like.

In general, in a transcriptionally inactive heterochromatin state, one tends to assume that TFs cannot search the target sites efficiently. Interestingly, our simulation results do not support this assumption and suggest that the TFs in the crowded nucleosomes (0.5 mM) can search the target sites more efficiently than in the diluted nucleosomes (0.1 mM). We note that we have constructed the high-density chromatin structure by the expansion-and-shrink protocol, and the 0.5 mM nucleosome concentration is close to the achievable upper limit in the current coarse-grained model (Fig. S1). Even in the limit, there remains unoccupied space large enough for the target proteins to move around. Thus, proteins can search the DNA sequence more efficiently because of the high concentration of DNA in crowded chromatin. In reality, however, there are many other proteins and nucleic acids, in addition to nucleosomes, that may prevent diffusion of TFs.

## Local DNA interaction profiles

We next addressed which parts of the DNA segment interact with proteins during the simulations. Using snapshots before $t_1$, we calculated the contact frequency of each DNA basepair with the DBDs of the three nuclear proteins (Fig. 6 *A*). We excluded rarely observed unwrapped nucleosomes from the analysis because it could be a computational artifact (see Supporting Materials and Methods). We note that some DNA partial unwrapping from the histone core occurred in the expansion-and-shrink simulation. In the figure, the DNA segments were aligned using the dyad position (set to 0) as a reference. Thus, the nucleosomal DNA corresponds to basepair number (−73:73) (*white* in Fig. 6 *A*), whereas the linker DNA corresponds to basepair numbers (−83:−74) and (74:83) (*gray* in Fig. 6 *A*).

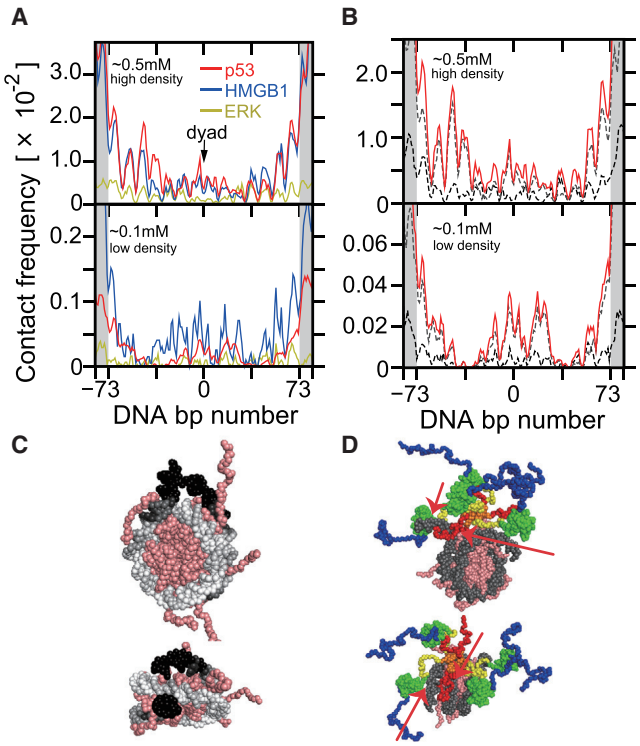


FIGURE 6 Contact frequency at each DNA site with nuclear proteins. The upper and lower panels correspond to results for the high- and low-density chromatin. The DNA fragments for single-nucleosome regions were aligned by the dyad position of the nucleosome. The left and right ends with the gray shade (ID (−83:−74) and ID (74:83)) correspond to the linker DNA regions, whereas the central part (ID [−73:73]) corresponds to the nucleosomal DNA (the bp ID = 0 corresponds to the dyad). (*A*) The contact frequency of entire DBDs of ERK (*yellow*), HMGB1 (*blue*), and p53 (*red*) is shown. (*B*) The contact frequencies of the core and CTD of p53 and the entire region of p53 are shown by black dashed, gray dashed, and red solid curves, respectively. (*C*) Contact frequency is mapped on the nucleosome structure (*side* and *top view*) by grayscale. (*D*) Representative complex snapshots of p53 bound on a nucleosome (*side* and *top view*) are shown. The red arrow indicates the major contacts between linker and/or dyad of DNA and core and/or CTD of p53. To see this figure in color, go online.

The most apparent feature in Fig. 6 *A* is that the contact frequencies in high nucleosome density are an order of magnitude larger than those in low nucleosome density. This difference is in harmony with the macromolecular crowding effect that enhances the nonspecific binding. Another prominent feature is that the contact frequencies in linker DNAs are much higher than those in nucleosomal DNAs; the ratios of the average frequency in linker DNAs to that in nucleosomal DNA were ∼2 for ERK and ∼8 for HMGB1 and p53 (quantified in Fig. S7). This preferential contact to linker DNA is reasonable because the linker DNA may be more accessible for the nuclear proteins compared with the nucleosomal DNA that wrapped histone octamers. We also note that even though the length of the linker DNA affects the ratio of contacts, the comparison among the three proteins would remain the same.

Within nucleosomal DNA, the contact frequency profiles were not uniform but exhibited ∼10 bp periodicity at least for HMGB1 and p53 regardless of the nucleosome density. In the high-nucleosome-density case in which more data samples are available than the low-nucleosome-density case, we see that peaks are higher near the entrance and exit regions ([−73:−53] and [53:73]) than the center. Termini of nucleosomal DNA tend to partially unwrap (20) where the nuclear proteins can access them. Also, we found a faint high peak near the dyad region.

By comparing the contact frequencies of the three proteins, we found that the contact frequencies of HMGB1 and p53 (per monomer) were significantly higher than that of ERK. For p53, we also separately examined the contact frequencies of the CTD and the core domain (Fig. 6 *B*). As a result, we found that the peaks around the dyad are mainly due to contacts by the CTD, not by the core domain. This tendency is consistent with the fact that the CTD of p53 has more prominent roles in nonspecific DNA binding than the core domain (46–48).

In Fig. 6 *C*, the contact frequencies between p53 and each DNA basepair are mapped as grayscale onto the simulation snapshot. From this figure, we can see that the linker DNA and entrance and exit part of nucleosomal DNA have higher frequencies (*almost black* in Fig. 6 *C*) than the dyad region (*dark gray*). The snapshot sampled in MD simulations at the low-density chromatins shows that p53 can make a complex with a nucleosome with the CTD and core domain (Fig. 6 *D*).

## Protein interaction profile

Next, we investigate which parts of the proteins interact with DNA. As expected, amino acids on the positively charged surface tend to have high probabilities to interact with DNA (Fig. 7 *A*), especially the highest peaks: residues 250–300 of ERK, residues 25–50 and around 150 of HMGB1, and residues near 125 and 350–393 of p53 agree with the experimentally suggested DBDs (41,42,46–48) (Fig. 7 *B*). For all the targets, the interaction surface patterns

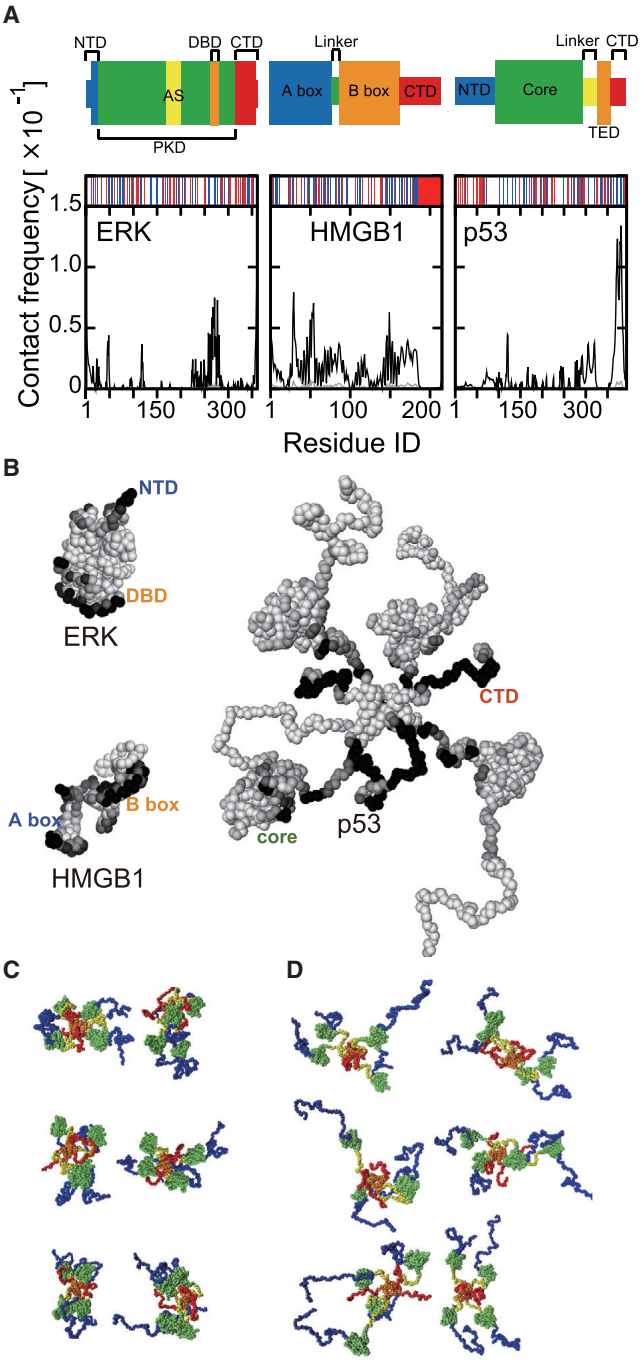


FIGURE 7 Contact frequencies of protein residues with DNA. (*A*) The black and gray curves correspond to the results in the high- and low-density chromatin environments, respectively. The upper barcodes show positively (*blue*) and negatively (*red*) charged residues for each protein. (*B*) Contact frequencies in the high-density chromatin mapped on snapshots are shown. (*C* and *D*) Three representative snapshots of p53 in the high-density chromatin with two views are given: $R_g$ of p53 in (*C*) and (*D*) corresponds to 60 ± 0.01 Å (near the minimal boundary of $R_g$ distribution in Fig. S9) and 78 ± 0.01 Å (around peak of $R_g$ distribution in Fig. S9), respectively. To see this figure in color, go online.

were indistinguishable between the low- and high-density chromatins (Fig. S8).

## Binding and dissociation kinetics

We also investigated the kinetics of the nuclear protein-DNA binding and unbinding. First, we confirmed that the simple Poisson process with the single exponential functions well described the binding and unbinding processes (Fig. S6 *A*) except for the binding process in low-density chromatin. The on- and off-rate constants that are determined as the inverse of averaged duration times are depicted in Fig. S6 *B*. The crowded chromatin environment facilitates the nuclear protein binding (the on rate increased) and decelerates the unbinding (the off rate decreased). Notably, the on rate was accelerated by a factor ∼13–20, which is larger than the ratio of the density, i.e., 5.

## Conformation of nuclear proteins in chromatin environment

In general, high-density crowding agents make flexible proteins more compact than in dilute solution because of the macromolecular crowding effects. Here, we investigated the conformational changes of the nuclear proteins in our model chromatin. We found that, although there was no marked change in the cases of ERK and HMGB1, p53 has a smaller radius of gyration $R_g^{TF}$ in chromatin with high nucleosome density than low nucleosome density (Fig. S9). The p53 snapshots with a smaller$R_g^{TF}$ (Fig. 7 *C*) differ from those with $R_g^{TF}$ (Fig. 7 *D*) at the peak of distribution in Fig. S9.

## Effects of acetylation on the CTD of p53

It has been implicated that p53 needs to be acetylated in its CTD for activation (52). Acetylation sites include K373, K382, and so on (52). The functional roles of the acetylation have been addressed experimentally (53,54) and computationally (22). In particular, using the same coarse-grained modeling approach as this work, we revealed that acetylation in the CTD of p53 remarkably reduces the sliding length so that the acetylated p53 tends to diffuse in three dimensions, whereas the unacetylated p53 can slide along DNA for a long duration (22,35). Here, we investigated how the acetylation alters the dynamics of p53 in chromatin. Namely, we focus on the influence of acetylation on the MSD profile, the diffusion coefficient, DNA-bp searching speed, and the contact frequency with DNA.

For this purpose, we conducted 20 independent molecular simulations in which p53 with the K373 and K382 residues acetylated was embedded in chromatin with high and low nucleosome density. To mimic the acetylation, we just neutralized the charges of the corresponding two lysine residues of all four monomers, whereas the rest of the simulation setups are the same as before. Whereas the unacetylated

p53 CTD contains +20 charges per a tetramer, the acetylated form has a reduced charge of +12, which reduces the affinity of the p53 CTD to DNA.

We plotted the MSD as a function of duration in Fig. S10 *A* and the diffusion coefficients in Fig. S10 *B*, comparing the acetylated p53 with the unacetylated p53. We found that the acetylation of p53 slightly weakened the crowding effect of chromatin on the diffusion, slowed the basepair searches (Fig. S10 *C*), and diminished the contact frequency between p53 and DNA, especially in the dyad (Fig. S10 *D*). The effect on the search speed is more pronounced for the CTD rather than the core domain.

## DISCUSSION

### Limitations in this study

In this study, we took a rather simple simulation setup that has many limitations.

First, we created structures of the chromatin with MD simulations without using experimental data. Recently, 3C-based experiments have provided structural restraints in the chromatin folding, which can be used to construct the chromatin structures (55).

Second, we employed simple Langevin dynamics without including hydrodynamic effects. Because the hydrodynamic effect is known to enhance the diffusion of large molecules, the diffusion coefficients evaluated by this study seem to give the lower limits (56). Notably, the hydrodynamic effect depends on the size of the proteins: the larger the protein, the more significant the impact of the hydrodynamics is. Thus, a quantitative comparison of the diffusion coefficients between p53 and HMGB1, for example, is inherently difficult in our work. However, qualitative comparison of the diffusion of the same proteins in different chromatin environments is likely to be less sensitive. Either way, including the hydrodynamic interaction is highly desired for the next step.

Third, real chromatin is much more complicated than just a polynucleosome array. Most notably, linker histones such as H1 bind on the nucleosome, which can affect binding and diffusion of nuclear proteins. Also, epigenetic marks such as histone acetylation would alter the chromatin structures and states greatly. These factors should be addressed in the future.

Fourth, our model does not approximate sequence-dependent protein-DNA interactions well. Nuclear proteins are attracted to DNA simply via the electrostatic interactions. Furthermore, the DNA model does not reflect the sequence-dependent bending rigidity, which could affect the model chromatin structure.

### Future challenges

Some challenges could be addressed in the future by extending our model. We will describe these challenges in the following three paragraphs.

It has been revealed in previous studies that HMGB1 facilitates deformation of nucleosome by repeated association to and dissociation from DNA that unwraps from histone core complexes (57). Then, the deformation of nucleosome helps TFs bind to their specific binding sites in nucleosomes (51,58). Also, the ability of HMGB1 to deform the nucleosome is regulated by the cysteine redox states (59). In experiments, the deformation requires a high concentration of HMGB1 (57). However, because only one HMGB1 molecule was included, the deformation could not be observed in our coarse-grained MD simulation. In the future, we will address the structural basis of nucleosome deformation activity of HMGB1 by simulating many HMGB1 molecules in chromatin.

As for the specific binding of p53 to its recognition site on DNA, rotational positioning of the site against histone core complexes regulates the binding (60,61). Also, the previous simulation studies indicated that the CTD of p53 interacts with the DBD and regulates the binding, too (62). The DNA model did not contain any specific binding sites in our study because we focused on the search of specific binding sites. Because of this, the simulation trajectories did not include p53 dynamics on specific binding sites. In the future, we will address the structural basis of specific DNA binding of p53 in nucleosomes.

In various simulation studies, the effect of crowder molecules on the search has been investigated (63–65). These studies indicated that the crowder molecules increased the fraction of the molecules which slides along DNA and promoted intersegmental transfer. In this study, we did not model the various kinds of proteins in the nucleus that could have molecular crowding effects. Thus, we could not discuss the impact of crowder molecules on the search in a chromatin environment, which can be addressed by naturally extending our model (65).

## CONCLUSIONS

We investigated diffusional dynamics, sequence search dynamics, and structures of two TFs, HMGB1 and p53, and a signaling protein, ERK, in chromatin with high and low nucleosome densities by coarse-grained MD simulations on the K computer, a massively parallelized supercomputer. Estimated diffusion coefficients in chromatin with a low nucleosome density were nearly the same as those in solution, whereas those in chromatin with a high nucleosome density were markedly reduced. Moreover, although the three nuclear proteins showed normal diffusion in chromatin with low nucleosome density, they exhibited anomalous subdiffusion in chromatin with high nucleosome density. Interestingly, we found that the rate of target search shows the opposite tendency: as the nucleosome density increases, the search rate increases. All the studied nuclear proteins preferred to contact with linker DNA rather than

nucleosomal DNA. We also studied the acetylation effect of the CTD of p53.

## SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.05.007.

## AUTHOR CONTRIBUTIONS

R.K.,T.T., and H.K. performed research and analyzed data. R.K., T.T., and S.T. conceived and designed the research. R.K. and S.T. wrote the manuscript. H.K. contributed simulation tools. All authors reviewed the final manuscript.

## ACKNOWLEDGMENTS

This work was supported in part by the Strategic Programs for Innovative Research "Supercomputational Life Science" of MEXT (Ministry of Education, Culture, Sports, Science, and Technology-Japan), in part by the MEXT KAKENHI 15H01351, 26104517, and 25251019, and in part by the Research Complex Promotion Program from the Japan Science and Technology Agency. Simulations were performed partly on the Supercomputer at Kyoto University and partly on K computer.

## REFERENCES

1. Alberts, B. 2015. Molecular Biology of the Cell, Sixth Edition. Garland Science, Taylor and Francis Group, New York.
2. Bell, O., V. K. Tiwari, …, D. Schübeler. 2011. Determinants and dynamics of genome accessibility. *Nat. Rev. Genet.* 12:554–564.
3. Tsompana, M., and M. J. Buck. 2014. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin.* 7:33.
4. Bigley, R. B., A. Y. Payumo, …, G. N. Huang. 2017. Insights into nuclear dynamics using live-cell imaging approaches. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 9:e1372.
5. Hihara, S., C. G. Pack, …, K. Maeshima. 2012. Local nucleosome dynamics facilitate chromatin accessibility in living mammalian cells. *Cell Rep.* 2:1645–1656.
6. Hibino, K., K. Kaizu, …, K. Maeshima. 2017. A combination approach based on live-cell imaging and computational modeling to further our understanding of chromatin, epigenetics, and the genome. Epigenetics and Systems Biology. Academic Press, pp. 225–238.
7. Yu, I., T. Mori, …, M. Feig. 2016. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife.* 5:e19274.
8. Perišić, O., and T. Schlick. 2017. Dependence of the linker histone and chromatin condensation on the nucleosome environment. *J. Phys. Chem. B.* 121:7823–7832.
9. Grigoryev, S. A., G. Arya, …, T. Schlick. 2009. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc. Natl. Acad. Sci. USA.* 106:13317–13322.
10. Bascom, G. D., K. Y. Sanbonmatsu, and T. Schlick. 2016. Mesoscale modeling reveals hierarchical looping of chromatin fibers near gene regulatory elements. *J. Phys. Chem. B.* 120:8642–8653.
11. Wedemann, G., and J. Langowski. 2002. Computer simulation of the 30-nanometer chromatin fiber. *Biophys. J.* 82:2847–2859.
12. Voltz, K., J. Trylska, …, J. Smith. 2008. Coarse-grained force field for the nucleosome from self-consistent multiscaling. *J. Comput. Chem.* 29:1429–1439.
13. Voltz, K., J. Trylska, …, J. Langowski. 2012. Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study. *Biophys. J.* 102:849–858.
14. Savelyev, A., and G. A. Papoian. 2009. Molecular renormalization group coarse-graining of polymer chains: application to double-stranded DNA. *Biophys. J.* 96:4044–4052.
15. Savelyev, A., and G. A. Papoian. 2010. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. USA.* 107:20340–20345.
16. Freeman, G. S., J. P. Lequieu, …, J. J. de Pablo. 2014. DNA shape dominates sequence affinity in nucleosome formation. *Phys. Rev. Lett.* 113:168101.
17. Freeman, G. S., D. M. Hinckley, …, J. J. de Pablo. 2014. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.* 141:165103.
18. Hinckley, D. M., G. S. Freeman, …, J. J. de Pablo. 2013. An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* 139:144903.
19. 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.
20. 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.
21. Tan, C., T. Terakawa, and S. Takada. 2016. Dynamic coupling among protein binding, sliding, and DNA bending revealed by molecular dynamics. *J. Am. Chem. Soc.* 138:8512–8522.
22. Terakawa, T., and S. Takada. 2015. p53 dynamics upon response element recognition explored by molecular simulations. *Sci. Rep.* 5:17107.
23. Chang, L., and S. Takada. 2016. Histone acetylation dependent energy landscapes in tri-nucleosome revealed by residue-resolved molecular simulations. *Sci. Rep.* 6:34441.
24. Saito, M., T. Terakawa, and S. Takada. 2017. How one-dimensional diffusion of transcription factors are affected by obstacles: coarse-grained molecular dynamics study. *Mol. Simul.* 43:1315–1321.
25. Shimizu, M., Y. Noguchi, …, S. Takada. 2016. Near-atomic structural model for bacterial DNA replication initiation complex and its functional insights. *Proc. Natl. Acad. Sci. USA.* 113:E8021–E8030.
26. Niina, T., G. B. Brandani, …, S. Takada. 2017. Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations. *PLoS Comput. Biol.* 13:e1005880.
27. Fan, Y., N. Korolev, …, L. Nordenskiöld. 2013. An advanced coarse-grained nucleosome core particle model for computer simulations of nucleosome-nucleosome interactions under varying ionic conditions. *PLoS One.* 8:e54228.
28. Korolev, N., L. Nordenskiöld, and A. P. Lyubartsev. 2016. Multiscale coarse-grained modelling of chromatin components: DNA and the nucleosome. *Adv. Colloid Interface Sci.* 232:36–48.
29. Khazanov, N., and Y. Levy. 2011. Sliding of p53 along DNA can be modulated by its oligomeric state and by cross-talks between its constituent domains. *J. Mol. Biol.* 408:335–355.
30. Krepel, D., D. Gomez, …, Y. Levy. 2016. Mechanism of facilitated diffusion during a DNA search in crowded environments. *J. Phys. Chem. B.* 120:11113–11122.
31. Daitchman, D., H. M. Greenblatt, and Y. Levy. 2018. Diffusion of ring-shaped proteins along DNA: case study of sliding clamps. *Nucleic Acids Res.* 46:5935–5949.
32. Kenzaki, H., N. Koga, …, S. Takada. 2011. CafeMol: a coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* 7:1979–1989.
33. Takada, S., R. Kanada, …, H. Kenzaki. 2015. Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc. Chem. Res.* 48:3026–3035.
34. Li, W., T. Terakawa, …, S. Takada. 2012. Energy landscape and multi-route folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc. Natl. Acad. Sci. USA.* 109:17789–17794.

35. 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.

36. Kurokawa, M. 2012. The K computer: 10 Peta-FLOPS supercomputer. In The 10th International Conference on Optical Internet (COIN2012). pp. 1.

37. Ou, H. D., S. Phan, T. J. Deerinck, …, C. C. O'Shea. 2017. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science.* 357:eaag0025.

38. Maeshima, K., R. Imai, …, Y. Joti. 2014. Chromatin structure revealed by X-ray scattering analysis and computational modeling. *Methods.* 70:154–161.

39. Maeshima, K., R. Imai, …, T. Nozaki. 2014. Chromatin as dynamic 10-nm fibers. *Chromosoma.* 123:225–237.

40. Roskoski, R., Jr. 2012. ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacol. Res.* 66:105–143.

41. Hu, S., Z. Xie, …, H. Zhu. 2009. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell.* 139:610–622.

42. Bianchi, M. E., and A. A. Manfredi. 2007. High-mobility group box 1 (HMGB1) protein at the crossroads between innate and adaptive immunity. *Immunol. Rev.* 220:35–46.

43. Gaglia, G., Y. Guan, …, G. Lahav. 2013. Activation and control of p53 tetramerization in individual living cells. *Proc. Natl. Acad. Sci. USA.* 110:15497–15501.

44. Joerger, A. C., and A. R. Fersht. 2008. Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.* 77:557–582.

45. Bell, S., C. Klein, …, J. Buchner. 2002. p53 contains large unstructured regions in its native state. *J. Mol. Biol.* 322:917–927.

46. Cho, Y., S. Gorina, …, N. P. Pavletich. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science.* 265:346–355.

47. Chen, Y., R. Dey, and L. Chen. 2010. Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure.* 18:246–256.

48. Weinberg, R. L., S. M. Freund, …, A. R. Fersht. 2004. Regulation of DNA binding of p53 by its C-terminal domain. *J. Mol. Biol.* 342:801–811.

49. Schalch, T., S. Duda, …, T. J. Richmond. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature.* 436:138–141.

50. Terakawa, T., and S. Takada. 2011. Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. *Biophys. J.* 101:1450–1458.

51. Rowell, J. P., K. L. Simpson, …, J. O. Thomas. 2012. HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail. *Structure.* 20:2014–2024.

52. Gu, W., and R. G. Roeder. 1997. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell.* 90:595–606.

53. Tang, Y., W. Zhao, …, W. Gu. 2008. Acetylation is indispensable for p53 activation. *Cell.* 133:612–626.

54. McKinney, K., M. Mattia, …, C. Prives. 2004. p53 linear diffusion along DNA requires its C terminus. *Mol. Cell.* 16:413–424.

55. Dekker, J., K. Rippe, …, N. Kleckner. 2002. Capturing chromosome conformation. *Science.* 295:1306–1311.

56. Takada, S. 2012. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* 22:130–137.

57. Scovell, W. M. 2016. High mobility group protein 1: a collaborator in nucleosome dynamics and estrogen-responsive gene expression. *World J. Biol. Chem.* 7:206–222.

58. Štros, M., E. Muselíková-Polanská, …, F. Strauss. 2004. High-affinity binding of tumor-suppressor protein p53 and HMGB1 to hemicatenated DNA loops. *Biochemistry.* 43:7215–7225.

59. Panneerselvam, S., P. Durai, …, S. Choi. 2016. Cysteine redox state plays a key role in the inter-domain movements of HMGB1: a molecular dynamics simulation study. *RSC Adv.* 6:100804–100819.

60. Sahu, G., D. Wang, …, A. K. Nagaich. 2010. p53 binding to nucleosomal DNA depends on the rotational positioning of DNA response element. *J. Biol. Chem.* 285:1321–1332.

61. Cui, F., and V. B. Zhurkin. 2014. Rotational positioning of nucleosomes facilitates selective binding of p53 to response elements associated with cell cycle arrest. *Nucleic Acids Res.* 42:836–847.

62. D'Abramo, M., N. Bešker, …, G. Chillemi. 2016. The p53 tetramer shows an induced-fit interaction of the C-terminal domain with the DNA-binding domain. *Oncogene.* 35:3272–3281.

63. Ma, Y., Y. Chen, …, K. Luo. 2016. How nonspecifically DNA-binding proteins search for the target in crowded environments. *J. Chem. Phys.* 144:125102.

64. Dey, P., and A. Bhattacherjee. 2018. Role of macromolecular crowding on the intracellular diffusion of DNA binding proteins. *Sci. Rep.* 8:844.

65. Krepel, D., and Y. Levy. 2017. Intersegmental transfer of proteins between DNA regions in the presence of crowding. *Phys. Chem. Chem. Phys.* 19:30562–30569.

**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 $C_\alpha$-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 $\phi_{ij}$ potentials, respectively. Each degree of freedom ($b_{ibd}$, $r_{ij}$, and $\phi_{ij}$) is biased to the native structure ($b_{ibd,0}$, $r_{ij0}$, and $\phi_{ij0}$), 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 $V_{\mathrm{ele}}$) 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, $\gamma$ is the friction coefficient, $\xi$ 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 $\Delta$ time steps of MSD time trajectories in Fig. 4A**

To compare the crowding effect by chromatin among three TFs easily, maximum $\Delta$-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 $\Delta$-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

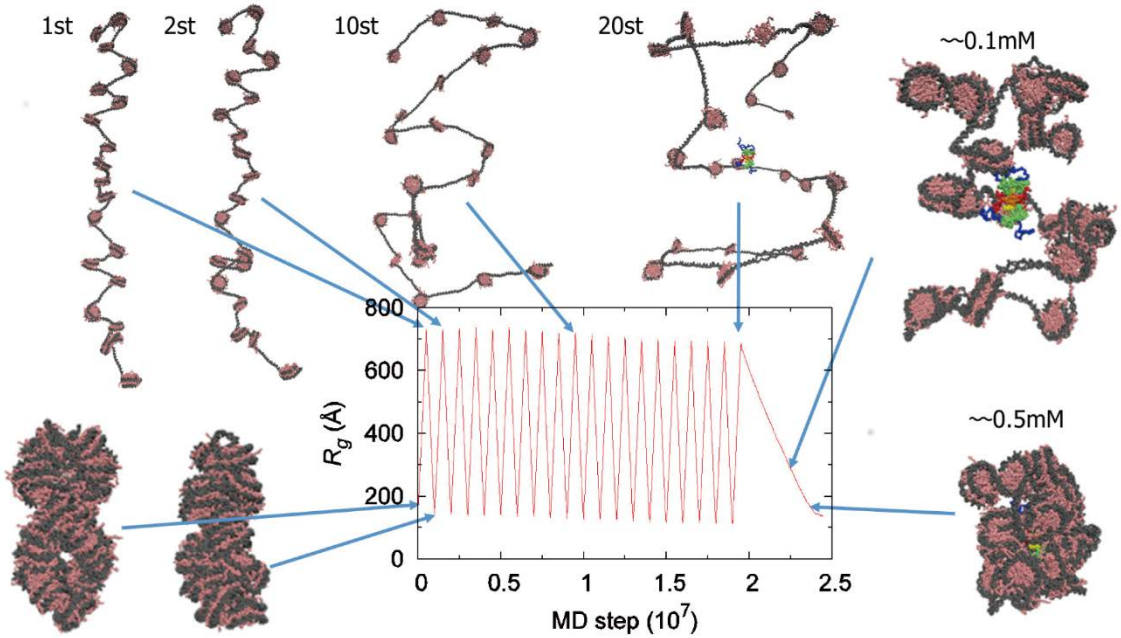


**Fig. S1: Construction of the model chromatin structures.** The time series of a radius of gyration, $R_g$, for the 20-nucleosome array, the model chromatin in the expand-and-shrink simulation are plotted together with representative snapshots. The 20-nucleosome array is enforced to expand-and-shrink 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.

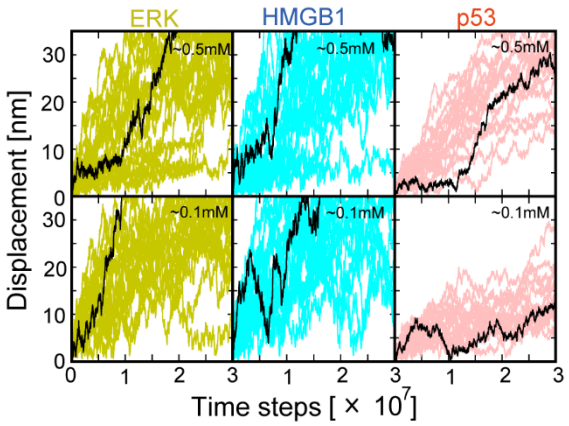


**Fig. S2: Movements of nuclear proteins in chromatin environments.** All time courses of displacements for centers of mass (COM) of nuclear proteins in the high-density (upper panel) and the low-density (lower panel) chromatins; ERK (yellow), HMGB1 (cyan), and p53 (pink). The black line in each panel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

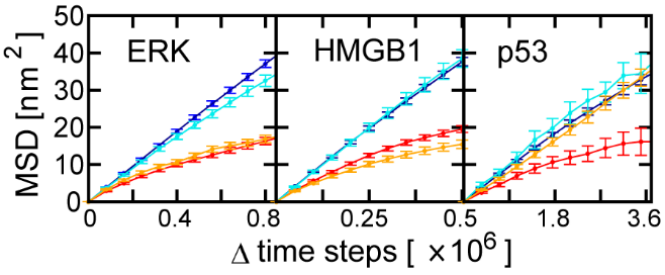

**Fig. S3: The comparison of the mean square displacements (MSDs) of nuclear proteins between those inside chromatin and those around the surface of chromatin.** The blue and red circles with error-bars are MSD plots within intra-chromatin, while the cyan and orange circles with error bars correspond 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.

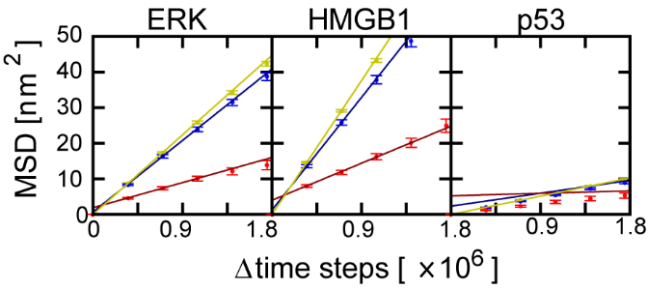


**Fig S4: The mean square displacements (MSDs) of nuclear proteins at the same $\Delta$time steps.** Each color for circles with error-bars is same as Fig. 4. The straight lines represent the linear fit line which is also same as Fig. 4. The error bars are defined as the standard errors.

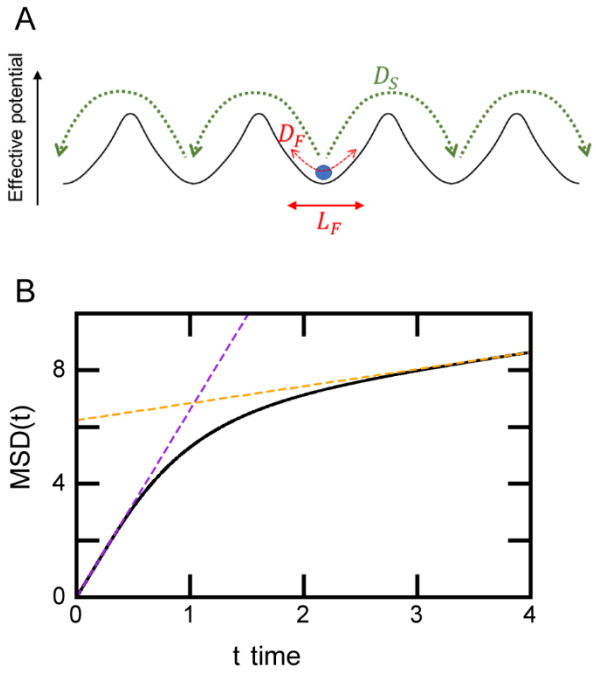


**Fig. S5: (A) The schematic view of the two-mode model of TF diffusion in crowded space.** The two-mode model may be expressed by the diffusional motion of a particle in effective potential. The blue particle represents the COM of the TF molecule. The width $L_F$ is the effective distance for the restricted 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.

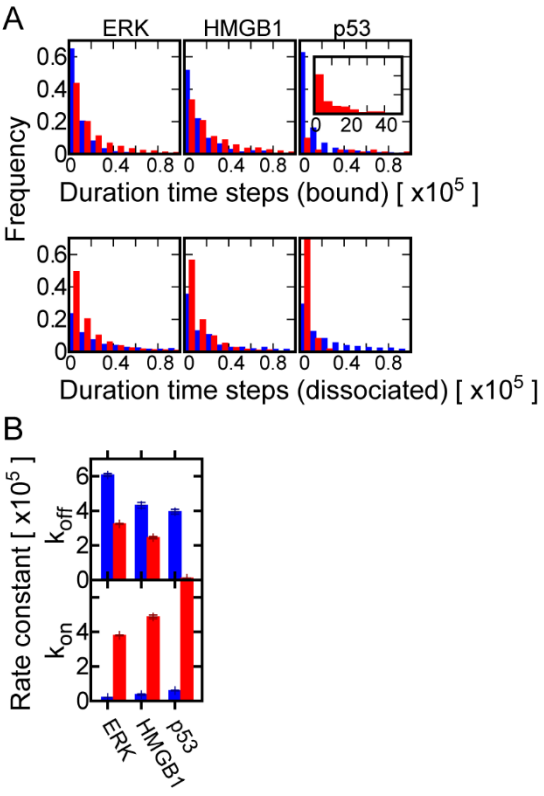


**Fig. S6: The kinetics of protein-DNA binding.** (A) The histogram of the duration time steps for the bound state (the upper panel) and the dissociated state (the lower panel) between proteins and DNA. The inset for p53 shows the distribution at the extended time steps. (B) The estimated rate constants $k_{off}$ 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.

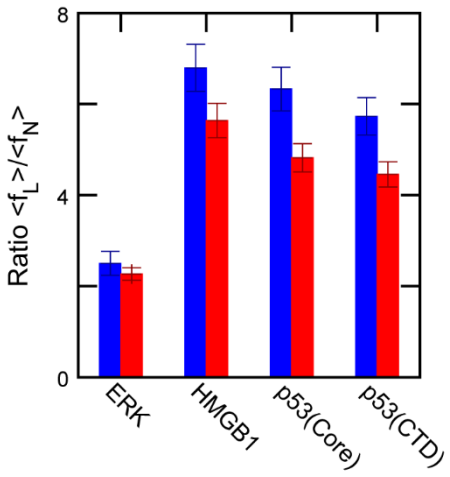


**Fig. S7: The ratio of the average contact frequencies in the linker-DNA: $\langle f_L \rangle$ relative to those in nucleosomal-DNA: $\langle f_N \rangle$.** The blue (red) bars correspond to the result for the low (high) density chromatin. The error bars are defined as the standard errors.

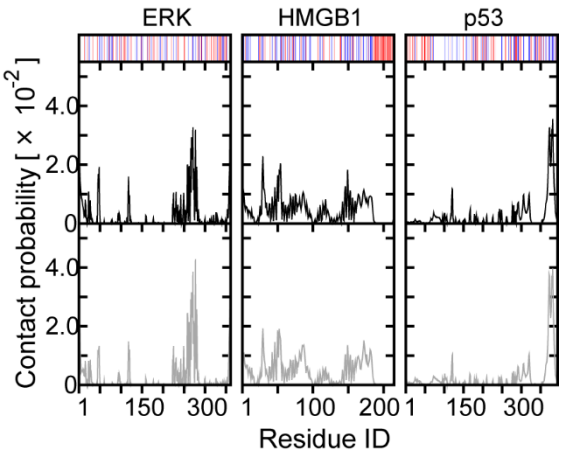


**Fig. S8: Contact probabilities (= normalized frequencies) of protein residues with DNA.** The black and gray curves correspond to the results for high-density and low-density chromatins, respectively. The upper panel with blue (+1) and red (-1) bars shows the charge distribution for each TF.

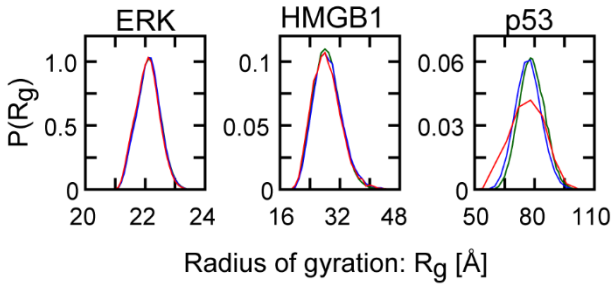


**Fig. S9: The probability distribution of radius of gyration $R_g$ for nuclear proteins in chromatins.** (A) In each panel, the green, blue, and red curves indicate the distributions at [nucleosome] =0.0 mM, 0.1mM, and 0.5mM, respectively.

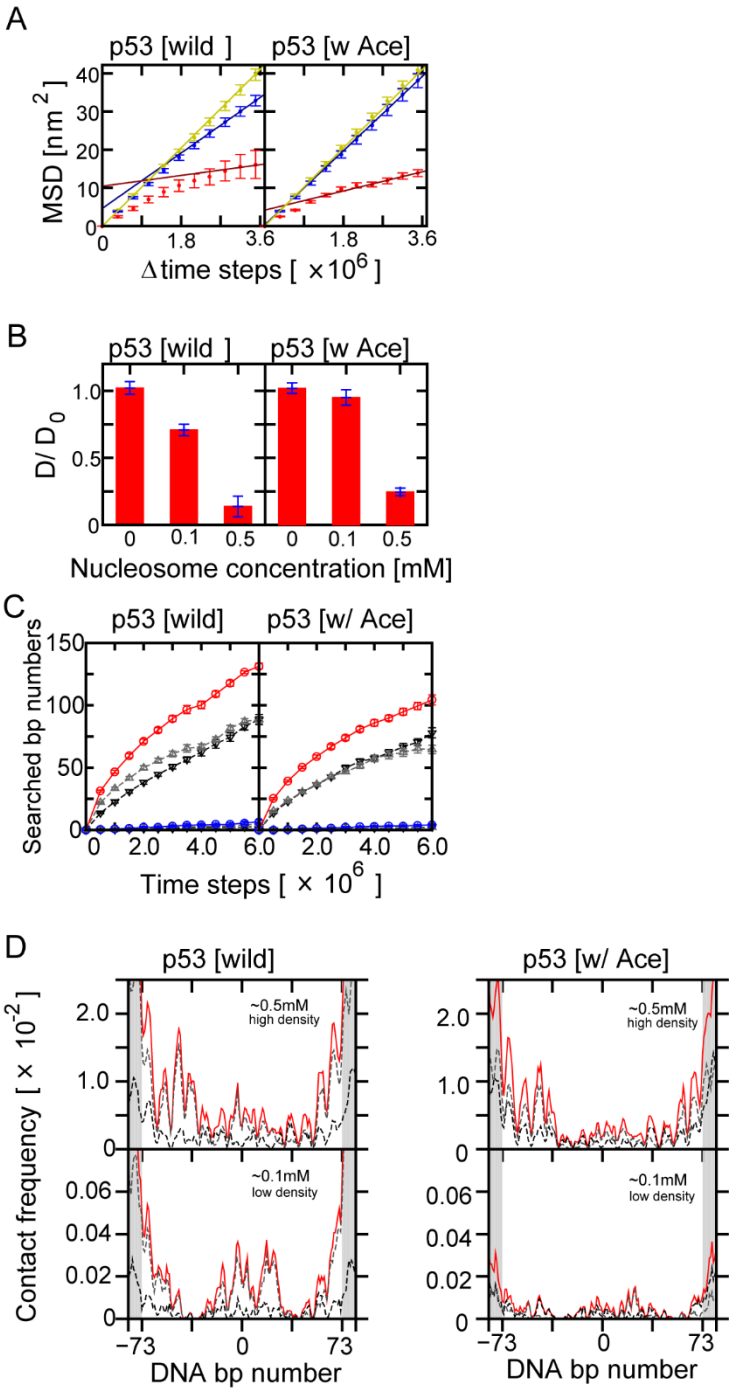


**Fig. S10: The comparison between wild-type p53 [wild] and the partially acetylated p53 [w Ace].** (A) MSDs, (B) the diffusion coefficients, (C) DNA-bp searching speeds, and (D) the contact-frequency profiles between p53 and DNA. The coloring rules and the meanings of panels (upper/lower) for panels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