

Doctoral Thesis

**Computational Analysis of
Molecular Dynamics in Biomolecular
Systems Including Nucleic Acids**

**Department of Mathematical and Life Sciences (MLS),
Graduate School of Science, Hiroshima University**

Takeru Kameda

Abstract

Biomolecular dynamics study based on computational science is an indispensable research field complement to experiments in structural biology, which primarily deals with proteins. In recent years, structures and dynamics of not only proteins but also nucleic acids (DNA and RNA) have been attracting attention in the fields of molecular and structural biology. Biological phenomena involving nucleic acids observed by rapidly developing experimental measurement technologies of genomics (e.g. RNA-seq, Chip-seq, etc.), are the background of that attention. These findings suggest that nucleic acids are no longer regarded merely as the media of genetic information. They should be considered as molecular machines that work with concrete structures and dynamics, in the same manner as proteins. Presently, structural biology involving nucleic acids is expected to be important in the research fields of life science and drug discovery.

Unfortunately, the dynamics of nucleic acids were not focused well in the biological research fields. Thus, these remain opportunities to extensively improve the dynamics analysis method of nucleic acids. For example, nucleic acid dynamics with sequence dependence, which is sensitive to alteration of nucleotide sequences, have not been much elucidated. Therefore, it is often difficult to evaluate the dynamics of molecular systems involving nucleic acids and their interaction with other molecules, by currently available analysis methods.

In this research, we studied several biomolecular systems involving nucleic acids using computational methods, to suggest effective dynamics analysis procedures. This proposal follows the historical developments of dynamics analysis methods of proteins in the field of structural biology. Through the analysis, we targeted following biomolecular systems: (i) simple double-stranded DNA; (ii) assembly of RNAs and proteins involving specific binding sites of two RNA segments; and (iii) complex of DNA and proteins bound by strong electrostatic interaction. I performed molecular dynamics analysis, and considered significance of simulated molecular dynamics. These topics and proposed methodologies are discussed, and are expected to apply to dynamics analysis of other biomolecular systems.

Index

1	Introduction	3
1.1	Evolution of Genomics	3
1.2	Development of Structural Biology	4
1.3	Analysis of Biomolecular Dynamics	5
1.4	Molecular Dynamics Analysis of Nucleic Acids	7
1.5	Outline of This Study	9
2	Methylation-Dependent DNA Dynamics	11
2.1	Introduction	11
2.2	Materials and Methods	12
2.3	Analyses and Results	14
2.3.1	Profiles of Double-Stranded DNA Stiffness	14
2.3.2	Local Geometry and Flexibility	16
2.3.3	Methyl Group Distribution	19
2.4	Discussion and Conclusion	21
3	Sequence Dependent Affinity of RNAs	23
3.1	Introduction	23
3.2	Materials and Methods	25
3.3	Analyses and Results	27
3.3.1	Binding Free Energy	27
3.3.2	Projected Free Energy	30
3.3.3	Binding Dynamics from the Structural Viewpoint	33
3.4	Discussion and Conclusion	35
4	Molecular Dynamics Analysis of Partially Disassembled Nucleosomes	37
4.1	Introduction	37
4.2	Materials and Methods	38
4.3	Analyses and Results	40
4.3.1	Structures of Histone Proteins	40
4.3.2	Structural Deformation of Nucleosomal DNA	42
4.3.3	Interactions Between Histone Tails and DNA	45
4.4	Discussion and Conclusion	47
5	Conclusion and Future Problems	49
6	Acknowledgments	51
7	Additional Information	52
7.1	Procedures of Molecular Modeling	52
7.2	Analysis of DNA Conformation Using X3DNA	53
7.3	Theory of Adaptive Biasing Force (ABF) Method	54
7.4	Supplemental Figures of Section 2	56
7.5	Supplemental Figures of Section 3	63
7.6	Supplemental Figures of Section 4	64

1 Introduction

1.1 Evolution of Genomics

Genetic information is well known to be encoded using DNA sequences [1]. This fact is common sense in life science [2,3]. Proteins are synthesized through the process of transcription and translation in the cell [4], and some of them work as molecular machines [5–7]. The projects to try to decode genetic information and to determine complete eukaryotic genome sequences started shortly after the discovery of genetic information [8]. Currently, genomes of many organisms (e.g. human and mouse) have been completely determined [8–11].

Complete determination of human genome sequence provided the fact that gene coding regions occupy approximately 1.5% of the whole genome [8], whereas gene coding regions occupy almost the whole genome in yeasts and bacteria [12, 13]. Therefore, non-gene coding (non-coding) regions should play other biological roles in the genomes of higher eukaryotes [8]. These non-coding regions have been studied since the discovery of them, with developing DNA sequencing technologies [14–18]. For example, enhancer regions activate the transcription of their spatially neighbor genes, by associations of binding proteins [19–22]. Insulator regions deactivate genes and interrupt genomic regions spatially [23–26], and their consensus sequences (in non-coding regions) have been revealed by recently developed Hi-C method [27–30]. Of course, large-scale genomic analysis projects have been proceeding [31–34].

The roles of DNA sequences including the non-coding regions have been studied focusing not only on biological but also physical properties [35–37]. For example, activities of transcription regulation regions showing higher-order structure formation require many modulation factors [38–40]. The factors show specific binding to genomic sites depending on DNA sequence, which has been observed by many biological experiments [41,42]. Because these factors are protein (i.e. that shows concrete structure), sequence-dependent specific affinities between DNA and protein should exist. Hence, it is reasonable that many biological functions are realized by their mechanical interactions, and this hypothesis is practically correct. Some significant mechanical interactions of them have been found [43–45].

Those previous findings suggest that DNA is not only genetic information container but also molecular machines showing distinct structure and dynamics, similar to protein. However, the physiological conditions involving nucleic acids and proteins are extremely complex environments including a huge number of molecules [46,47]. Precisely designed interactions of DNA and protein are at least required, to realize collective biological activity, though a comprehensive understanding of the mechanism is not enough due to the difficulty of observations. Currently, a lot of studies, which employ bottom-up structural biological approaches based on the structure and dynamics of biomolecules, have been conducted [48,49]. The development of structural biology is expected also in the field of genome biology.

1.2 Development of Structural Biology

There are two major ways in the analysis of biomolecular structure and dynamics. One of them is experimental structure analysis methods [50]. Biomolecular structures are determined by crystallography [51–53], nuclear magnetic resonance (NMR) method [54–56], and cryo electron microscopy (cryoEM) [57–59]. Then, their structural dynamics are inferred and discussed by employing dynamics measurement methods like NMR and FRET [60–62]. Recently, cryoEM structure determination has been significantly developed, and determined structures have been rapidly increasing [63–65]. Note that although most of these methods are experimental, there are also computational procedures (e.g. spectral attributions of excited atoms in NMR, structure construction of proteins through fitting electron density in cryoEM), as mentioned below [55, 57, 66].

Another approach in the field of structural biology is molecular simulation [67–69], in other words, numerical calculations of molecular properties using computational methods. More specifically, it is numerical calculation (of molecular dynamics, electron density distribution, etc.) methods based on theoretical physics and theoretical chemistry [70–72]. These methods provide microscopic molecular properties, difficult to capture by currently established experimental procedures [67–69]. Molecular simulation enables the complex operation of molecules (e.g. adding chemical modifications, exchanging monomer, etc.) and is indeed useful, while it requires experimentally determined information (e.g. reference molecular structures) before carrying out [73, 74]. Because of their advantages, molecular simulation has been developed and has played not separative, but cooperative roles in experimental approaches [75–77]. Practically, in studies focused on the structure and functions of proteins (e.g. enzyme activities), molecular simulation is indispensable and has contributed enormously [69, 77–79]. Additionally, it has been gradually recognized that computational approaches are useful for similar problems in the field of drug discoveries [80–84]. Further development of molecular simulation techniques and applications is hoped in a lot of research fields [73, 74, 85].

1.3 Analysis of Biomolecular Dynamics

Molecular simulation has been developed presently [85–87]. Molecular simulation methods are typically classified into two types. One of them is the quantum mechanics (QM) based method. This method can deal with chemical reactions involving electron transfer [88,89]. As it requires huge computational cost because of calculation of basis set [90,91], then it cannot focus on protein dynamics at domain scale in general. To overcome the problem, another method is employed in such cases. The method is based on classical mechanics (molecular mechanics; MM) and called molecular dynamics (MD). Although MD assumes that the molecular dynamics do not involve electron transfer [88,89], it can reach dynamics of large-scale molecules because of reduction of computational costs [92]. Additionally, the hybrid simulation method involving QM and MM (called QM/MM method) was proposed to gain both advantages and has been developed presently [92,93].

In this thesis, I only used the MD simulation method, to analyze the dynamics of biomolecular systems including nucleic acids (larger molecules than protein in general; explained in Sec. 1.4). The methodologies of MD are classified as follows: (i) obtaining the time series of molecular dynamics; (ii) evaluation of conformational ensembles and physical quantities (e.g. free energy), with employing extended theoretical methods and sacrificing time-series information [94]; (iii) an exhaustive evaluation of intermolecular affinities between their already determined structures (relatively similar to (ii)).

The method (i) is used to obtain real-time series of molecular dynamics instead of experimental observation. In this case, though the potential energy defined for intra- and inter-molecular interactions should be calculated by quantum mechanics in principle, it is defined by approximated classical mechanics assuming the molecular dynamics does not involve electron transfer [88,89]. In particular, first, atomic forces applied to molecules are obtained from the summation of potential energies. Then, the evolution of atomic position is calculated by Newton's equation of motion [95,96]. With assuming ergodicity, obtained time series of molecular dynamics can be regarded as ensembles [97]. However, the following major problem of the conformational sampling method is raised. Through the ordinary simulation (obtaining time series of molecular dynamics), their trajectories may be distributed only in local minimum regions of free energy profiles and this problem often induces biased sampling [98–101]. Following computational methods (ii) are approaches to overcome that problem.

The method (ii) provides more efficient sampling of molecular conformations and physical quantities than ordinary simulation and is called the extended simulation method [102,103]. On the other hand, ordinary simulation should be employed to obtain time series integrity because extended methods often sacrifice it. This extended simulation method includes replica exchange molecular dynamics simulation (REMD), free energy perturbation (FEP), etc. [94,102,104]. For example, the temperature REMD (T-REMD) method is a famous one [104–106]. T-REMD carries out molecular dynamics simulation of several independent simulation replicas, and each replica is set at different constant temperature. Then, exchanging trials of atomic coordinates between neighbor two replicas are probabilistically conducted at constant time intervals [104,105]. This exchange prevents

trapping at local minimum regions and allows accurate sampling of free energy profiles [98–100]. The advantages of REMD are confirmed in the studies of protein folding dynamics [107–109]. Currently, new sampling methods have been proposed along with the evolution of theoretical physics and computational science [110–114].

The method (iii) is called "Docking" simulation and employed to compare the affinity of many pairs of molecules (e.g. antigens and antibodies) [115–118]. For example, to evaluate structural affinity between target molecules (e.g. proteins), interaction scores (potentials) between molecules are calculated based on interaction functions (force fields), and candidates of the structures in interaction are suggested [119]. Its advantage is well known in the fields of designing antibodies and compounds [120–122]. Recently, other types of methods based on machine learning have been proposed [123–126].

To carry out the molecular simulation, although their execution programs can be made by own, several public (paid and unpaid) packages are available in the ranges of conventional simulation methods. AMBER [127, 128], CHARMM [129, 130], GROMACS [131, 132], GENESIS [133, 134], MODYLAS [135, 136], and NAMD [137, 138] are major molecular dynamics (MD) packages and have been well employed. On the other hand, in quantum mechanics, Gaussian (<http://gaussian.com>), etc. are widely used. In this research, NAMD (<https://www.ks.uiuc.edu/Research/namd/>) is employed (Sec. 7.1). NAMD is good at large scale parallel numerical calculation [137, 138].

Force field functions are also necessary for MD simulation. Force field involves fundamental parameters of functions and coefficients (e.g. elastic bonds and spring constant), to express physical constraints of atomic (molecular) interaction such as chemical bonds and van der Waals (VDW) force [139]. Major ones are AMBER [140, 141], CHARMM [142, 143], GROMOS [144], and OPLS [145]. These force field parameters have been updated periodically.

Coarse-grained modeling, in which multiple atoms are substituted with one particle has been developed [146–149]. This method provides fast numerical simulations because of the representation, and contributed to research focusing on relatively large molecular systems [150–156]. Although not used in this study, our group also has tried coarse-grained modeling of proteins and DNA [157–160].

1.4 Molecular Dynamics Analysis of Nucleic Acids

As mentioned above, molecular simulation methods have been greatly developed [85–87]. Practically, the recent acceleration of numerical calculation by the development of computer science enables a long-time, large-scale molecular simulation [161,162]. This acceleration expanded the targets of numerical simulation. In the past, almost all targets were small molecules (e.g. single proteins). Currently, molecular complexes and nucleic acids, which involve larger atom numbers than conventional cases, have been gradually dealt with by computational methods [163,164]. As referred in Section 1.1, mechanical properties of biomolecular systems involving nucleic acids have been gradually focused. Therefore, molecular simulation of biomolecular systems involving nucleic acids is expected to occupy important research areas in life sciences and drug discoveries [165].

Unfortunately, biomolecular complex systems involving nucleic acids are less studied [166] as they are difficult to deal with by mechanics [167,168]. Let us consider DNA. First, the gyration radius of nucleotides is larger than that of amino acids. In addition, their polymerized segments show much extended and linear conformation [169,170] while peptide segments (proteins) often show folded and compacted structure [171,172]. These different shapes result in different solvated system volumes and numbers of water molecules (of course size of DNA is large). It makes computational costs of solvated systems of DNA much larger than that of protein. Next, the mechanics of DNA is extremely anisotropic unlike proteins, because their structures are composed of two flexible segments coupled by weak hydrogen bonds [173–176]. Furthermore, although it has been already known that mechanical properties of DNA depend on DNA sequences [43,177,178], unified solution of sequence-dependent DNA dynamics was hardly worked out because of the various combinations of sequence (nucleotides). Besides, it is extremely hard to distinguish the interaction among DNA and proteins, electrostatic force and sequence-dependent structural affinity [179,180]. These complexities make it difficult to elucidate specific nucleotide recognition mechanisms used by binding proteins. Two DNA segments (= double strand) are connected by only hydrogen bonds, then the cleavage of them easily occurs at typically 350~360 K, i.e. not high temperature [181]. This less thermal stability is sometimes useful in experiment [182]. However, that makes it impossible to apply T-REMD, which is a powerful computational approach to obtain structural ensembles of proteins [104–106], because DNA cannot endure the high temperature (e.g. 600 K, which is often used in T-REMD of proteins). Double-strand separates rapidly, and cannot be expected to revert to the original structure in simulation time scales. Therefore, the acceleration of structural sampling by tempering methods is not applicable. Instead, methods such as reaction coordinate-based methods may be appropriate [110–113].

Furthermore, several difficulties exist in the case of RNA. Due to the irregularity of single-strand RNA dynamics, physical determination of their folded (secondary) structure is much difficult [166,183], while mathematical estimation is not easy [184,185]. Currently, though non-coding RNAs have been attracted attention along with the rapid development of sequencing technology, their biological roles are hardly elucidated [186,187]. Studies of their roles from a physical

viewpoint are somehow conducted [188,189]. Appropriate computational approaches may enable us to elucidate their roles through the evaluation of their dynamics and molecular interactions [165].

Presently, though above-mentioned unfortunate problems are raised, several computational approaches have been performed and their outcomes have been presented. For example, the dynamics of a nucleosome, which is well known biomolecular complex of DNA and proteins, have been studied by several research groups [190–195]. In particular, Kono et al. (QST, Japan) analyzed dynamics of histone core and tails [163,164,196,197], tail dynamics dependent on histone modification [198,199], and DNA unwrapping dynamics [194,200,201], using molecular dynamics simulation. These studies commonly employed free energy evaluation using long-time and extended methods, to bridge biological roles and intractable nucleic acid dynamics. To analyze biomolecular systems involving nucleic acids, which is hard to deal with, it is necessary and appropriate to capture structural dynamics and their sequence dependence with eliminating unnecessary dynamics, in the extent of currently available computational methods.

1.5 Outline of This Study

To survey the various roles of genomic DNA (Sec. 1.1), I have considered effective methods to evaluate the mechanics and dynamics of nucleic acids, regarding them not as primitive media of genetic information but as molecular machines, which involve structures and information. In this thesis, I discussed several research topics currently I propose [202–205]. These topics focused on several target biomolecular systems involving nucleic acids and employed effective analysis methods to evaluate the molecular systems (Fig. 1). My suggested methods are expected to be employed in similar studies in the future.

(i) First, in Section 2, we considered a statistical analysis method to detect structural dynamics changes of double-stranded DNA at atomic scale induced by chemical modification [202]. This evaluation method provides effective insight into sequence-dependent double-stranded DNA physics and is expected to be applied in the studies of sequence-dependent nucleotide recognition [206–208].

(ii) Second, in Section 3, we considered sequence-dependent RNA binding dynamics, which is thought to be more complicated than double-stranded DNA, based on free energy profiles obtained with respect to multiple reaction coordinates [203]. This is a cooperative study with chemical biology researchers and expected to complement the experimental research, to analyze microscopic dynamics of RNA binding. This free energy-based evaluation should be adopted in further studies of analysis and prediction of RNA structural dynamics.

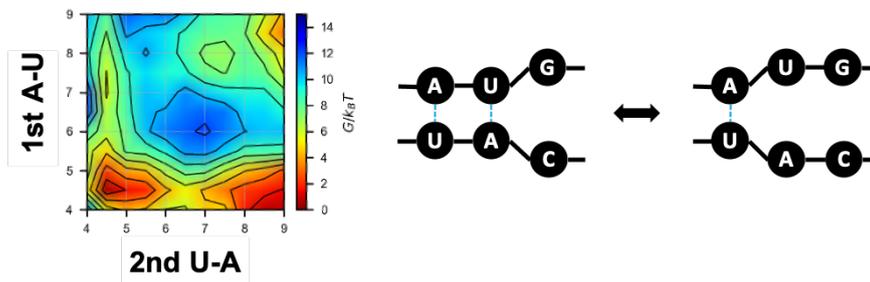
(iii) Finally, in Section 4, I evaluated nucleosome dynamics. In particular, I proposed a statistical analysis method to evaluate effective dynamics of DNA, which is restrained by interaction with histone proteins [204, 205], from long time simulation trajectories. This research also employed multiple probability profiles (equivalent to free energy profiles) and associated them with experimentally observed biological phenomena involving nucleosome dynamics.

These methodologies, results, and conjectures will be discussed as follows (Fig. 1).

(i) DNA Dynamics and Modification Dependence



(ii) Free Energy Profile and Corresponding Dynamics of RNAs



(iii) Effective Deformation of Nucleosomal DNA

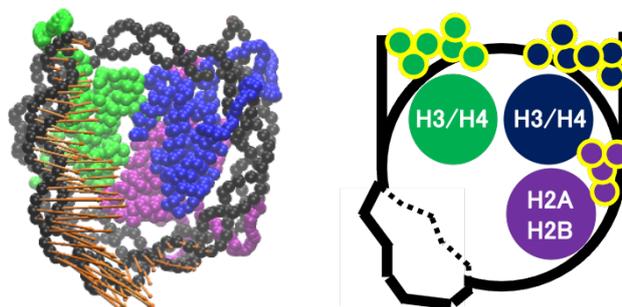


Figure 1. Performed Researches in This Study. (i) Methylation Dependent DNA Dynamics. (ii) Sequence Dependent Binding Dynamics of RNAs. (iii) Molecular Dynamics Analysis of Partially Disassembled Nucleosomes.

2 Methylation-Dependent DNA Dynamics

This section is based on ref. [202].

2.1 Introduction

DNA methylation is one of well known biochemical modifications of genomic DNA and has been associated with many biological phenomena by a huge number of biological researches [209–211]. A methyl group (CH_3 -) binds to the cytosine (C) base in a C \rightarrow G (CpG) dinucleotide [209]. CpG with a methylated cytosine is conventionally called mCpG. mCpGs are widely and frequently distributed in higher eukaryotic genomes and affect transcription regulation [210,212]. In particular, whole genomic regions are globally methylated in mammalian genomes (e.g. human and mouse) [213]. Though mCpGs are distributed all over the genome, their biological roles show the dependency on the genome regions [214]. For example, DNA methylation at transcription start sites (TSSs) induces condensation of TSS with binding or recruitment of several mCpG recognition enzymes [215]. Then, the expression of (TSS methylated) genes is silenced. For the regulation of selective gene expressions, only parts of TSSs including CpG (called CpG Island (CGI)) are unmethylated while almost all genes are silenced [216].

DNA methylation has been studied from the aspects of not only biological and experimental but also physical properties of nucleic acids. Through the previous researches from the viewpoint of physics, the following mechanical effects and responses of DNA induced by methylation were observed: (i) methylation decreases flexibility of DNA as chainlike polymer [217]; (ii) methylation might affect (facilitates or prevents) DNA strand separation [218]; (iii) methylation enhances attractions of homologous DNA [219]; (iv) methylation modulates dynamics of nucleosomes involving mCpG [220]. Therefore, DNA methylation should be an important factor of sequence-dependent mechanics of DNA, which is the fundamental basis of molecular recognition and complex formation with other biomolecules (e.g. proteins). In particular, extreme methylation levels of DNA (e.g. TSSs in mammalian genomes) are considered to determine their mechanical properties, however their details are mostly unclear. Thus, a further systematic analysis which can distinguish the origin of the effects on structural dynamics of DNA, the dependence of sequence and methylation patterns, should be more appropriate than previous researches [221–225].

In this study, to systematically evaluate the changes of DNA dynamics by methylation pattern in more detail, we performed a fully atomic molecular dynamics simulation of double-stranded DNA with several typical methylation patterns, and analyzed differences of mechanics among the patterns [202]. The findings in this study may serve as a basis for a discussion on methylation-dependent DNA dynamics.

2.2 Materials and Methods

Atomic coordinates of double-stranded DNA are necessary to carry out fully atomic molecular dynamics (MD) simulation, in the same way as proteins. Typically, atomic coordinates of proteins were determined by experiments [51,55,58], and can be obtained from Protein Data Bank (PDB) (<https://www.rcsb.org>). However, it is impossible to obtain DNA structures consisting of assigned sequences in the same way (though some short DNA structures can be downloaded from PDB). Therefore, we constructed DNA coordinates of designed sequences using X3DNA (the method referred to in Sec. 7.2). In this research, to focus on structural dynamics of DNA that depend only on methylation patterns (i.e. eliminating (sacrificing) sequence pattern dependency), we restrict the targets to 50 base-pairs, repetitive CpG dinucleotide sequences (i.e. 5'-(CG)₂₅-3'). Then, we modified some CpG to mCpG (Fig. 2 (a)) to obtain sequences corresponding to certain mCpG ratios, as listed in Tab. 1. To infer the coordinates of missing atoms, we employed CHARMM36 force-field (2017 update) and VMD [226]. For the detailed procedures, please see Sec. 7.1.

Table 1. Target Sequences. For each model, methylated CpG (mCpG) content ratio, methylated cytosine (mC) sites, and repeat units are listed. All sequences are composed of only CpG dinucleotides. mC sites are indexed as shown in Fig. 2 (c). Note that the models, except SEQ 0.00 and SEQ 1.00, include partial (fractional) sequences at the ends (Tab. 2).

Model	mCpG Content	mC Site (Nucleotide Indices)	Repeat Unit
SEQ 0.00	0% (0)	-	CG
SEQ 0.25	25% (1/4)	1 9 17 25 33 41 49 51 59 67 75 83 91 99	CGCGmCGCG
SEQ 0.33	33% (1/3)	1 7 13 19 25 31 37 43 49 51 57 63 69 75 81 87 93 99	CGmCGCG
SEQ 0.50	50% (1/2)	1 5 9 13 17 21 25 29 33 37 41 45 49 51 55 59 63 67 71 75 79 83 87 91 95 99	CGmCG
SEQ 0.67	67% (2/3)	3 5 9 11 15 17 21 23 27 29 33 35 39 41 45 47 53 55 59 61 65 67 71 73 77 79 83 85 89 91 95 97	CGmCGmCG
SEQ 0.75	75% (3/4)	3 5 7 11 13 15 19 21 23 27 29 31 35 37 39 43 45 47 53 55 57 61 63 65 69 71 73 77 79 81 85 87 89 93 95 97	CGmCGmCGmCG
SEQ 1.00	100% (1)	1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99	mCG

Constructed DNA molecules were solvated in a $60 \text{ \AA} \times 60 \text{ \AA} \times 200 \text{ \AA}$ water box of TIP3P water model, and neutralized by K^+ , and added 150 mM KCl (Fig. 2 (b)). Periodic boundary condition with Particle-Mesh Ewald (PME) electrostatics [227,228] was employed and a cutoff of 12 \AA (with switching from 10 \AA) was used for nonbonded interactions. Langevin thermostat (damping coefficient: 5/ps) and Langevin-piston barostat were adopted in MD simulation [229,230]. Temperature and pressure were set at 300 K and 1 atm, respectively. Energy minimization (10,000 steps) and 10.0 ns of annealing were carried out before the 50.0 ns of production run in equilibrium. During annealing and production run (time-step: 2fs), weak harmonic restraints were applied to C1' atoms of nucleotides at

both ends of the DNA segment (Fig. 2 (b); spring constant: $1.0 \text{ pN}/\text{\AA}$, centered at the position after the energy minimization). This restraint was applied to prevent translation and rotation of DNA molecule, to reduce computational costs. Individual ten simulation trials were carried out for each DNA model (Tab. 1). Numerical simulation was performed using NAMD (version 2.13 multicore with CUDA) [137, 138].

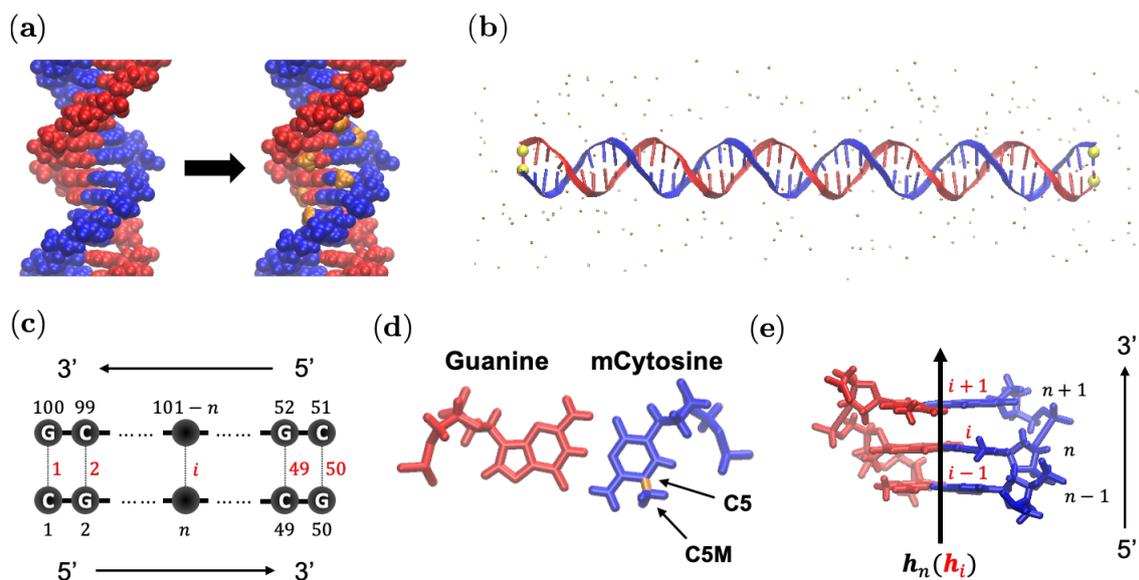


Figure 2. Overview of the Model. (a) Structure of DNA with poly-CpG (left) and poly-mCpG (right). The methyl groups are shown in orange. (b) Solvated DNA model. Small particles represent ions. C1' atoms located at both ends (shown in yellow) are restrained in the simulation. (c) Nucleotide indices n (shown in black) and base-pair indices i (shown in red) of the double-stranded DNA. (d) Locations of C5 and C5M atoms in 5-methylated cytosine. (e) Schematic representation of helical axis $h_n(h_i)$.

2.3 Analyses and Results

2.3.1 Profiles of Double-Stranded DNA Stiffness

First, we confirmed that DNA is stiffened by methylation, observed in experimental research [217], through evaluation of (i) overall structural variation of DNA [157] and (ii) persistence length as polymer physics [231].

On the overall structure variation (i), the following statistical method was applied. For atomic coordinates (x_k, y_k, z_k) for all heavy atom ks at individual time points were transformed as $(\Delta x_k, \Delta y_k, \Delta z_k) = (x_k - \langle x_m \rangle_m, y_k - \langle y_m \rangle_m, z_k - \langle z_m \rangle_m)$ ($\langle \rangle_m$ means the average for all atom ms). Then, covariance matrix I of $(\Delta x_k, \Delta y_k, \Delta z_k)$ for individual time points were calculated as Eq. 2.1. The descending ordered eigenvalues $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, and $\sqrt{\lambda_3}$ ($\sqrt{\lambda_1} > \sqrt{\lambda_2} \geq \sqrt{\lambda_3} > 0$) were obtained from eigendecomposition of matrix I . λ_1 and λ_2 corresponded to the extension and distortion of double-stranded DNA, respectively [157]. Scatter plots of $(\sqrt{\lambda_1}, \sqrt{\lambda_2})$ were shown in Figs. 3 and S2. Note that $(\sqrt{\lambda_1}, \sqrt{\lambda_2})$ can be obtained from each time point τ . Thus, those scatter points correspond to individual time point τs .

$$I := \begin{pmatrix} \langle (\Delta x_k)^2 \rangle_k & \langle \Delta x_k \Delta y_k \rangle_k & \langle \Delta x_k \Delta z_k \rangle_k \\ \langle \Delta y_k \Delta x_k \rangle_k & \langle (\Delta y_k)^2 \rangle_k & \langle \Delta y_k \Delta z_k \rangle_k \\ \langle \Delta z_k \Delta x_k \rangle_k & \langle \Delta z_k \Delta y_k \rangle_k & \langle (\Delta z_k)^2 \rangle_k \end{pmatrix} \quad (2.1)$$

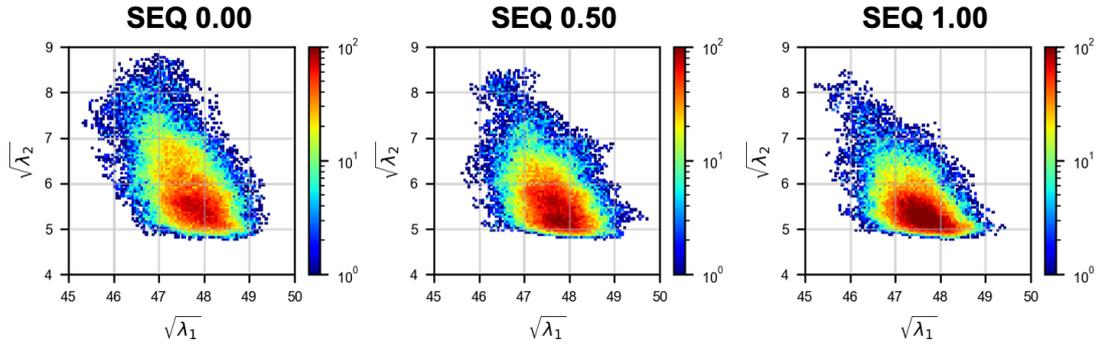


Figure 3. Profiles of Overall Geometry. The distribution of $(\sqrt{\lambda_1}, \sqrt{\lambda_2})$ summed over ten trials is shown.

Methylation content ratios of DNA (Tab. 1) correlate with increasing of larger $\sqrt{\lambda_1}$ and lower $\sqrt{\lambda_2}$ (Fig. 3). More details, mostly, $(\sqrt{\lambda_1}, \sqrt{\lambda_2})$ was in the range of $47 \leq \sqrt{\lambda_1} \leq 49$ and $5 \leq \sqrt{\lambda_2} \leq 6$, corresponding to a conformation similar to the shape of stretched DNA. The distribution outside of this range was different for these models. The probability within this range was higher for SEQ 1.00 than that for SEQ 0.00. $\sqrt{\lambda_2}$ of SEQ 1.00 tended to be smaller than that of SEQ 0.00. This implied that high mCpG content resulted in a stretched, or stiffer DNA shape. The other models (Tab. 1) showed intermediate stiffness between SEQ 0.00 and SEQ 1.00 (Figs. 3 and S2). The variation of $\sqrt{\lambda_2}$ was suppressed as the mCpG content increased. These results showed that higher methylation content stiffened

DNA in equilibrium conditions. The typical extension level of the DNA structure is shown in Fig. 4.

Next, persistence length L_p of DNA was evaluated by following procedures. The set A_d , inner-product of helical axes \mathbf{h}_i (Fig. 2 (e)) separated by d nm ($d \in [0.3, 10.0]$ at intervals $\Delta d (= 0.01)$), was defined as

$$A_d := \left\{ \mathbf{h}_i \cdot \mathbf{h}_j \mid d - \frac{1}{2}\Delta d \leq C_{ij} < d + \frac{1}{2}\Delta d \right\}$$

$$C_{ij} := \sum_{l=i}^{j-1} |c_{l+1} - c_l| \quad (2.2)$$

$$(11 \leq i < j \leq 40)$$

where c_i is the center of base-pair i . c_i and \mathbf{h}_i were obtained using X3DNA [232, 233]. Then, pair sets $\{(d, \langle H \rangle_{H \in A_d})\}$ ($\langle \dots \rangle_H$, which indicates the average for all H s) were fitted by function $f(d) := \exp(-d/L_p)$. Profiles of persistence length L_p for each methylation level were shown in Fig. 4. Each point corresponds to L_p obtained from one simulation trajectory (5,000 snapshots). Overall, a clear correlation was observed between mCpG content and L_p , the significant difference of L_p was confirmed. Note that evaluated L_p scores well agreed with experimentally obtained persistence length of DNA (~ 50 Å [234–236]).

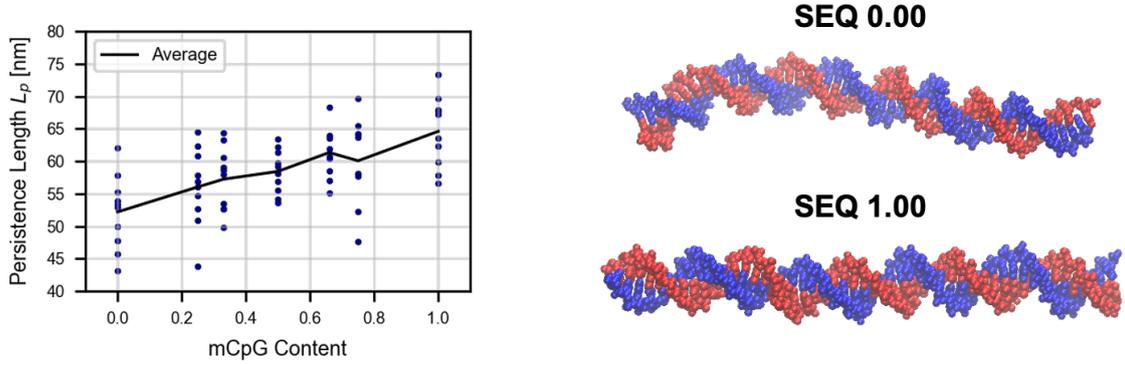


Figure 4. Profiles of the Persistence Length of DNA. (Left) Persistence length L_p . The average within each trial and over ten trials are shown. The horizontal axis corresponds to the mCpG content (Tab. 1). p -values (by Welch’s t -test) between mCpG content 0.0 and 0.5, 0.5 and 1.0, and 0.0 and 1.0 are 9×10^{-3} , 8×10^{-3} , and 1×10^{-4} , respectively. (Right) Snapshots of the structure of SEQ 0.00 and SEQ 1.00 models at 40 ns.

2.3.2 Local Geometry and Flexibility

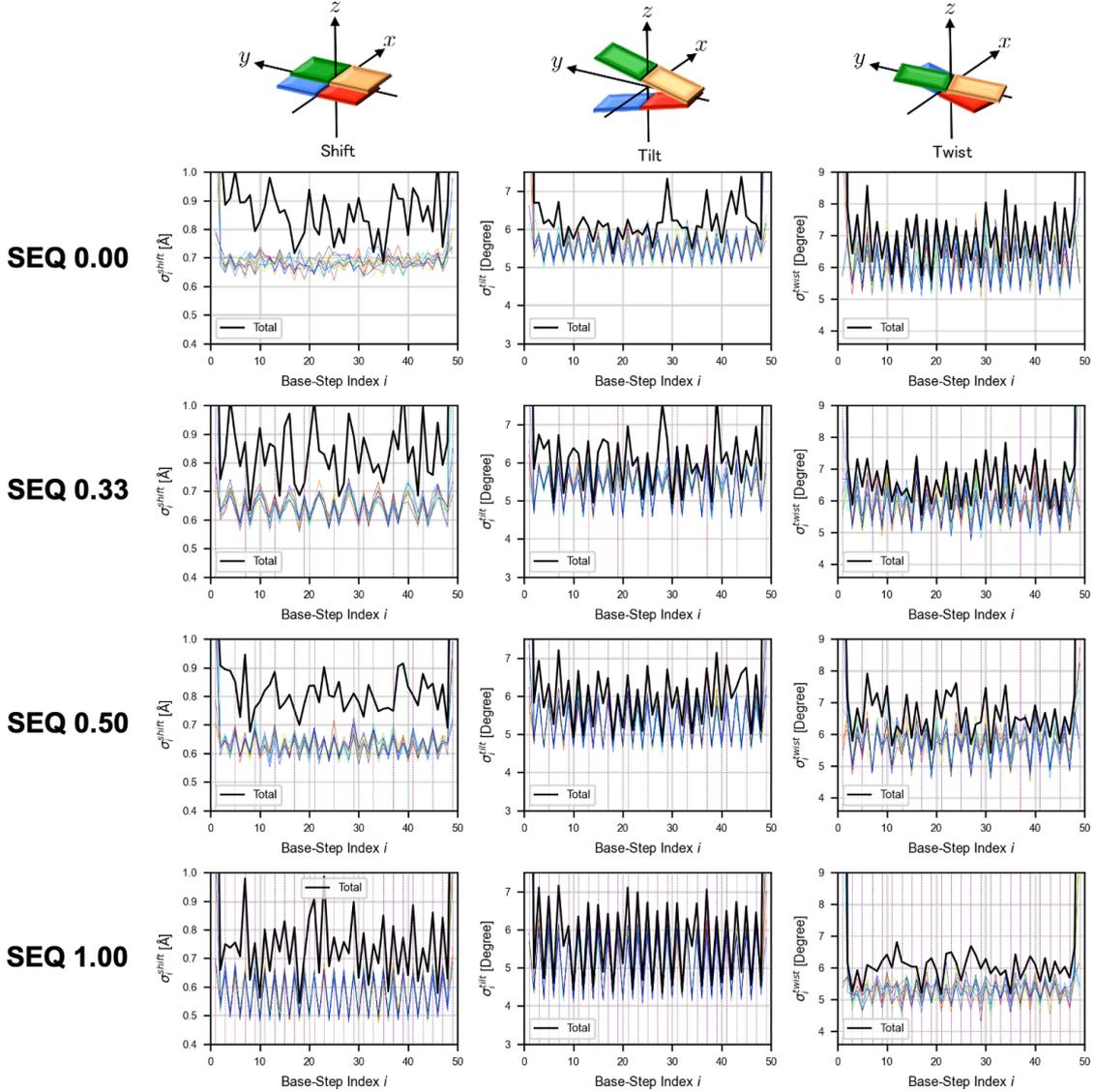


Figure 5. Profiles of σ_i^{shift} , σ_i^{tilt} , and σ_i^{twist} . Colors show different simulation trials and the black line shows the overall flexibility (structural variation) in ten trials. In the horizontal axis, odd and even numbers correspond to base-steps of C→G and G→C, respectively (see Fig. 2 (c)). Purple lines show mC→G (C→G in the case of mC) base-steps.

We confirmed DNA methylation stiffened double-stranded DNA (Figs. 3 and 4). Then, we analyzed the origin of DNA stiffening by evaluating structural geometries of base-steps using X3DNA [232, 233] as follows. (i) The scores $X_i^\bullet(\tau)$ for base-step parameter (BSP, see Fig. S1) \bullet of base-step i (base-pair indices from i to $i+1$ (Fig. 2 (c))) at respective simulation time point τ , were evaluated. (ii) $X_i^\bullet(\tau)$ was transformed as $\hat{X}_i^\bullet(\tau) := X_i^\bullet(\tau) - X_i^\bullet(0)$, where $X_i^\bullet(0)$ means the score just after energy minimization (see Sec. 2.2). Thus, $\hat{X}_i^\bullet(\tau)$ corresponds to difference of BSP

• of base-step i from (meta-)stable (= energy minimized) conformation. (iii) The flexibility score σ_i^\bullet of BSP • at base-step i was defined as the standard deviation (S.D.) of $\hat{X}_i^\bullet(\tau)$. Note that σ_i^\bullet was calculated either in each simulation trial or over ten trials for the same model. The results of BSPs (Fig. S3) that show especially remarkable changes of σ_i^\bullet among different mCpG contents were shown in Fig. 5.

By considering mCpG sites, the following trends could be observed. σ_i^{shift} and σ_i^{tilt} decreased at the base-steps G→C adjacent to the mCpG, while not in the middle C→G base-step (Fig. 5). In particular, similar σ_i^{tilt} profiles over individual simulation trials were observed (while not observed in σ_i^{shift} case). This result suggests that mCpG restricts neighbor \hat{X}_i^{tilt} dynamics and their deformation in determined ranges. On the other hand, a fundamentally different trend was observed in terms of σ_i^{twist} (Fig. 5). Overall decrease of σ_i^{twist} appeared with increasing of mCpG content, then it is shown that methylation globally restricts twisting between base-pairs. However, the decrease of twisting was not observed at mCpG base-step.

Table 2. Sequence Constitution of the Models. \star^m represents m -times of the iteration of the repeat unit.

Model	Repeat Unit	Sequence Constitution
SEQ 0.00	CG	5'- \star^{25} -3'
SEQ 0.25	CGCGmCGCG	5'-mCGCG- \star^5 -CGCGmCG-3'
SEQ 0.33	CGmCGCG	5'-mCGCG- \star^7 -CGmCG-3'
SEQ 0.50	CGmCG	5'-mCG- \star^{12} -3'
SEQ 0.67	CGmCGmCG	5'- \star^8 -CG-3'
SEQ 0.75	CGmCGmCGmCG	5'- \star^6 -CG-3'
SEQ 1.00	mCG	5'- \star^{25} -3'

To clarify the above-mentioned trends, we aligned and averaged σ_i^\bullet in terms of repeat units with eliminating boundary ten base-steps, i.e. only base-steps i in $11 \leq i \leq 40$ were employed. The repeat units were listed in Tab. 2. We calculated σ_i^\bullet for each simulation trajectory, then averaging σ_i^\bullet for equivalent base-steps l in repeat units noted as

$$\overline{\sigma_l^\bullet} := \langle \sigma_i^\bullet \rangle_{i \sim l} \quad (2.3)$$

where $i \sim l$ means that i corresponds to equivalent location of l (Tab. 2). Average \pm S.D. over simulation trajectories were evaluated (Fig. S4), and only remarkable results were shown in Fig. 6. Suggested trends of σ_i^{tilt} and σ_i^{twist} could be observed.

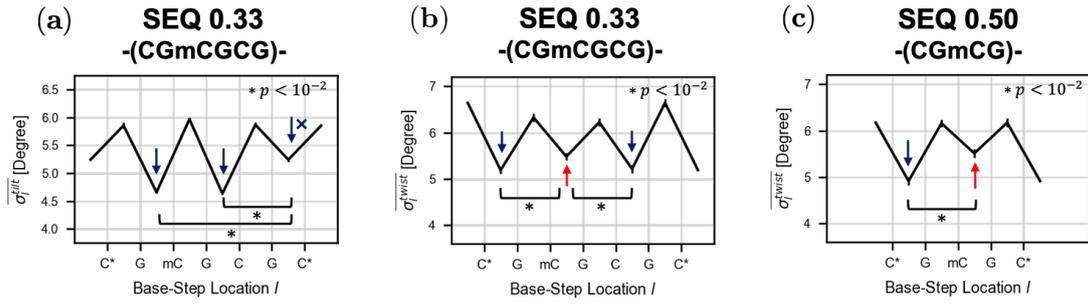


Figure 6. Profiles of $\overline{\sigma_l^\bullet}$ for the Repeat Unit. σ_i^{tilt} and σ_i^{twist} (shown in Fig. 5) are aligned and averaged in terms of the repeat unit (Tab. 1). Error bars show mean \pm S.D. over ten trials; see Sec. 7.2 for the method. As these profiles are for base-step parameters, data points are shown between two consecutive bases. Nucleotides at both ends (shown with * in the horizontal axis) are identical. (a) $\overline{\sigma_l^{\text{tilt}}}$ of SEQ 0.33 (CGmCGCG, averaged over five iterations); (b) $\overline{\sigma_l^{\text{twist}}}$ of SEQ 0.33 (CGmCGCG, averaged over five iterations); (c) $\overline{\sigma_l^{\text{twist}}}$ of SEQ 0.50 (CGmCG, averaged over seven iterations). In (a), reduction of S.D. at G \rightarrow C is significantly larger at both neighbors of mCpG (* $p = 3 \times 10^{-16}$ each). In (b) and (c), S.D. at mC \rightarrow G is significantly higher than unmethylated C \rightarrow G (* $p = 1 \times 10^{-7}$ (b; left), 1×10^{-6} (b; right), and 6×10^{-11} (c)).

2.3.3 Methyl Group Distribution

Based on the previous results (Figs. 5 and 6), changes of base-step dynamics σ_i^\bullet can be considered to originate from physical interaction between DNA and methyl groups. Then, we analyzed methyl group dynamics from simulation trajectories. At first, we calculated θ_n , the angle between vectors of the direction of the methyl group and the plane formed by the base-pair (Fig. 7 (a)). In particular, employing normalized direction vectors \mathbf{h}_n of base-step axis at base n (Fig. 2 (e)), \mathbf{d}_n^p of base-pair at base n , which connects atoms of C1' of base n to opposite strand, and \mathbf{d}_n^m of methyl group at base n (Fig. 2 (d)), θ_n was obtained as follows.

$$\begin{aligned}\theta_n &:= \theta_n^m - \theta_n^p \\ \theta_n^m &:= \cos^{-1}(\mathbf{h}_n \cdot \mathbf{d}_n^m) \\ \theta_n^p &:= \cos^{-1}(\mathbf{h}_n \cdot \mathbf{d}_n^p)\end{aligned}\tag{2.4}$$

Note that length of \mathbf{h}_n , \mathbf{d}_n^m , and \mathbf{d}_n^p were equal to 1. Statistics of θ_n at base n over ten simulation trial for each model were evaluated (Figs. 7 (a) and S5). Overall, profiles of θ_n were negative but could fluctuate. These negatively directed methyl group dynamics suggested that any physical interactions between methyl groups and neighbor (5'- side) nucleotide. Then, we evaluated contact frequency between them averaged over ten simulation trial (Fig. 7 (b)). A pair of a methyl group and nucleotide was regarded in contact if their shortest atomic distance was less than 3.0 Å. In all cases, methyl groups contacted to neighbor (5'- side) nucleotide (Figs. 7 (b) and S6). We considered the electrostatic interaction between methyl groups and others. However, no interaction pairs were identified (data not shown). Therefore, it was shown that changes of σ_i^\bullet by positional effects of methyl groups (Fig. 6) were induced by weak van der Waals (VDW) interactions and simple excluded volume effects.

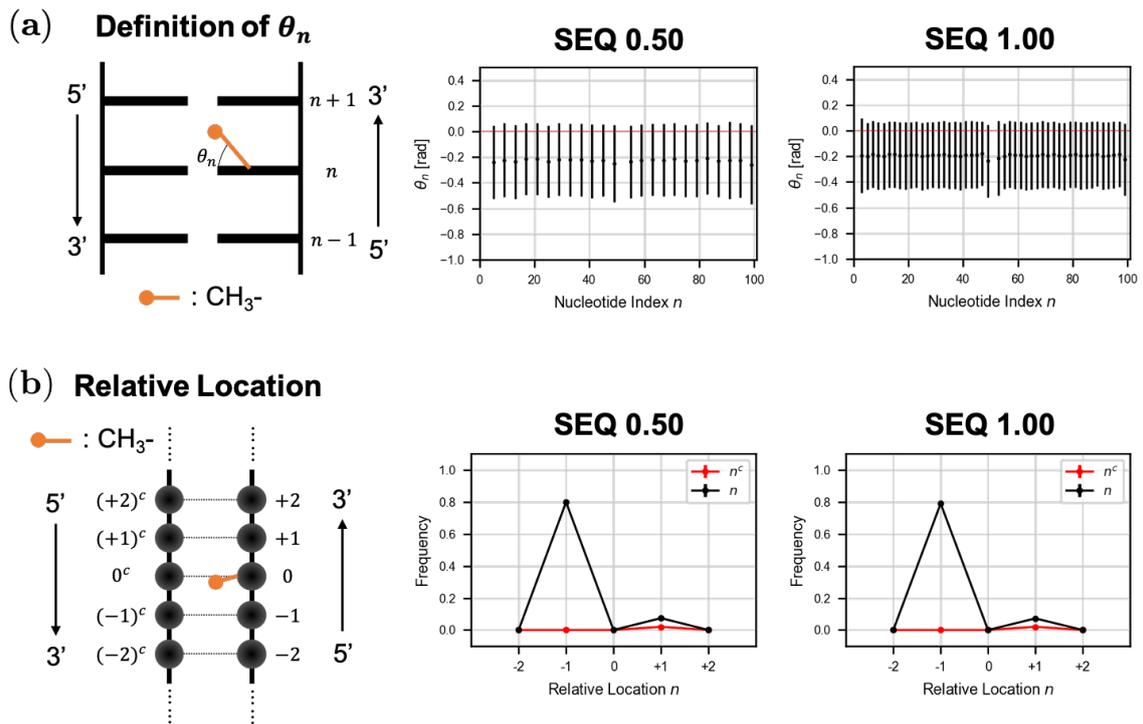


Figure 7. Orientation of Methyl Groups and Their Interactions with DNA. (a) The definition and observed distribution of θ_n . The average \pm the standard deviation (S.D.) of θ_n for each methyl group is shown. (b) The definition of nucleotide indices relative to the methyl group, and estimated contact frequencies. The average \pm S.D. of contact frequency for all methyl groups is shown.

2.4 Discussion and Conclusion

The schematics (Fig. 8) show the result of this research [202]. Due to site-specific methylation, BSP (Fig. S1) dynamics corresponding to local bending of DNA shape (shift and tilt) in the neighbor of mCpG were prevented (Fig. 5). In addition, winding and unwinding dynamics of base-steps (twist) were restricted and decreased in the long-range, whereas the mCpG base-step itself shows an inverse trend. These restrictions of BSP were induced by physical interaction between methyl groups and their neighbor nucleotides (Fig. 7). Further analysis of molecules involving DNA methylation (e.g. exploring dynamics of methylation recognition enzyme [206,207]) might provide additional interesting findings.

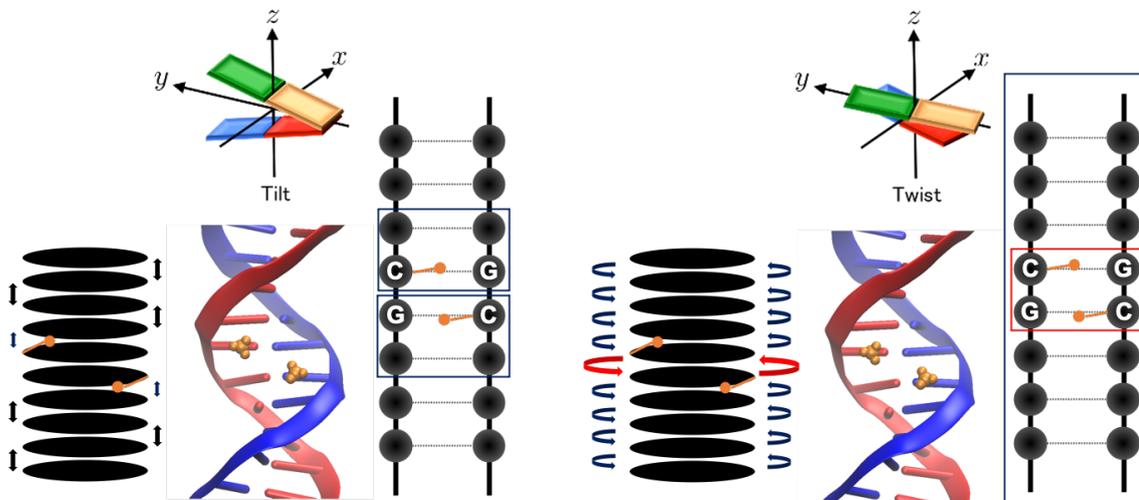


Figure 8. Schematics of Directed Effects Induced by DNA Methylation. Suggested effects on tilt and twist dynamics. Navy (or red) squares represent ranges where site-specific DNA methylation (i.e. existence of mCpG) makes the tilt or twist mode stiffer (or relatively flexible).

Macroscopic mechanics in terms of CGI (CpG Island), i.e. typical hypomethylated DNA, was investigated previously [237]. This research evaluated the mechanical response of hypermethylated DNA to stretching by nanometry using optical tweezers. Stiffening of DNA was observed, though disappeared when DNA was overstretched. Our simulation study suggested that the base-step dynamics (Fig. 5) was suppressed by dynamic interactions between DNA and methyl groups (Fig. 8). These interactions are expected to occur only in B-DNA like conformation. Thus, our proposed mechanism (Fig. 8) agrees well with this experimental study [237].

Another study suggested that hydration around methyl groups affects base-step dynamics [225]. These chemical properties might be involved in the DNA stiffening mechanism.

Previous research suggests that mCpG stabilizes nucleosome positioning [238]. The nucleosome structure was determined by X-ray crystal analysis using typical sequences of DNA showing high affinity to core histone proteins. In the structures, DNA conformation was not uniformly curved around histones but bent at several

sites [239]. The result of our study can be associated with this fact, because (above-mentioned) experimental observation supports the hypothesis that site-specific DNA methylation may change local base-step dynamics (Fig. 5). This change might induce a similar situation and enhance nucleosome positioning.

In this study, we employed fully atomic molecular dynamics simulation, to focus on detailed DNA dynamics. However, coarse-grained modeling of DNA with reduced atom numbers and degrees of freedom has proceeded to investigate dynamics of long DNA with reducing computational costs [156–158, 240]. Methylation-dependent DNA dynamics showed in this study will be incorporated into those extended modeling in the future.

3 Sequence Dependent Affinity of RNAs

This section is based on ref. [203].

3.1 Introduction

A biological process to synthesize protein in gene expression is called translation [241,242]. Protein is synthesized based on the nucleotide sequence of (already spliced) mRNA in the molecular complex "ribosomes" through the conversion of nucleotide sequence to amino acid sequence [243]. Each three-nucleotide unit from the start codon (AUG) toward the 3'-terminal corresponds to one amino acids [244]. AUG is employed as typical start codon over all organisms [245].

The eukaryotic translation process is established by following procedures [246]. First, ribosomal subunit 40S changes to 43S preinitiation complex (PIC) with the help of several modulation factors (eukaryotic Initiation Factors; eIFs) [247]. Next, the 43S PIC forms 48S PIC, which can start to read mRNA sequence, and seeks a start (AUG) codon with several eIFs. Finally, after scanning of mRNA, recognition of start codon, and binding of Met-tRNA_i^{Met} at P-site, the 48S PIC structure forms 80S with combining 60S ribosomal subunit and additional factors [248,249]. Elongation of peptide starts after the formation of 80S ribosomal complex. Note that the ribosomal molecular complex consists of rRNA (nucleic acids) and initiation factors (proteins), and its structural determination is much interesting topic in structural biology [250–252].

In the start codon recognition in the procedure above, typical start codon (AUG) and its complementary nucleotides (CAU) of Met-tRNA_i^{Met} are known to bind strongly [253,254]. Therefore, the binding affinity of AUG-CAU bases (codon and anticodon) determines the possibility of translation initiation of 48S PIC. Actually, in PIC structure, scanning of mRNA starts at the 5'-terminal and reads three-nucleotide sites sliding at one nucleotide interval [241,242].

However, other three nucleotides (i.e. not AUG) could play a role of the start codon in previous experimental studies [255,256]. For example, in archaea and eubacteria, not only AUG but also GUG and UUG could work as start codons though initiation frequency is much lower than AUG [255]. On the other hand, in eukaryotes, CUG can initiate translation though less frequently than AUG [256]. The accuracy of start codon selectivity has been studied, and relationships to eIFs have been revealed [257,258]. In eukaryotes, several eIFs (eIF1 and eIF1A) show specific interaction during start codon recognition in 48S PIC. They are responsible for providing strict AUG recognition in mRNA sequence [259,260]. This strictness (AUG recognition) and exception (CUG start codon) are conflicting, which is an interesting biological problem. Additionally, although several mismatch codons (e.g. GUG) hardly work as start codon, their chemically modified derivatives do it at a high frequency. Therefore, nucleotide-dependent free energy profiles of codon-anticodon binding should be an important information to determine the possibility to start translation initiation. In other words, binding dynamics of three nucleotides of mRNA and CAU of Met-tRNA_i^{Met} in 48S PIC determines the frequency of translation initiation by the three nucleotides.

Previously, a computational study on the start codon recognition was performed using molecular dynamics simulation [261]. This study focused on the binding affinity of CAU of Met-tRNA^{Met}_i, and typical (AUG) and near-cognate (e.g. GUG) codons of mRNA in the partial model of 48S PIC [261]. The free energy difference between AUG and other codons in (artificial) bound states was calculated using free energy perturbation (FEP [94]). However, the codon-anticodon bound state defined in this FEP-based method conjectured merely through substituting a base with the corresponding base in the AUG. The result of this method cannot solve the problem: what molecular dynamics and interaction cause the changes of binding affinity between various types of codons. For example, the transition process from unbound to bound states cannot be analyzed, then differences of the transition between different codons cannot be discussed. Generally, the structure of a single RNA segment is much flexible and disordered than that of folded proteins. This uncertainty makes it difficult to determine the apparent bound state (also unbound state). Therefore, to discuss the changes of nucleotide, other simulation methods are required.

In this study, another free energy evaluation method, adaptive biasing force (ABF [113, 262, 263]), was employed to discuss the binding dynamics of 48S PIC model (similar to previous study [261]) involving several types of codons. The established procedure of free energy evaluation enables us to discuss binding dynamics of different codons, using ABF at individual base-pair resolution. In particular, free energy profile in terms of three reaction coordinates, which correspond to the distance of individual base-pairs of codon-anticodon, was calculated. Then, changes of binding dynamics can be discussed by the changes of the free energy profile. These changes provide a structural insight into binding dynamics depending on nucleotide sequence [203]. This method is expected not only to be employed in binding analysis of further various (including chemically modified) codons but also to provide further solutions for free energy-based RNA dynamics analysis in other biomolecular systems.

3.2 Materials and Methods

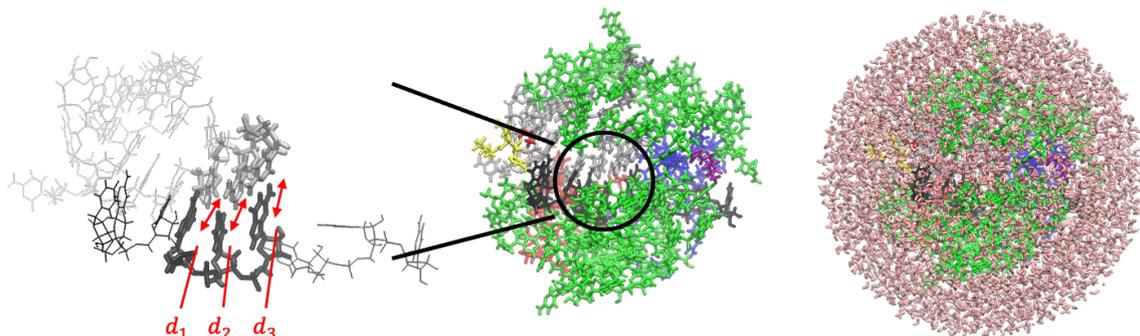


Figure 9. PIC Structure Models and Reaction Coordinates in This Study. (Left) mRNA and tRNA around the codon-anticodon base-pairs (Tab. 3) are shown in black and silver, respectively. Distances d_1 , d_2 and d_3 used as reaction coordinates are indicated. (Center) rRNA segments (split) are shown in green, and protein components are shown in other colors. (Right) Water molecules are shown in pink, which enclose the system and form a sphere.

Previous research of start codon recognition in eukaryotic translation initiation using all-atom molecular dynamics (MD) simulations [261] was referred in this study. To reconstruct codon-anticodon interaction in solution, open pre-initiation complex (PIC) structure (PDB ID: 3J81) determined by CryoEM [264] was employed. To reduce computational cost for MD simulation, atoms within 25 Å from N1 atom in the middle base of anticodon in tRNA molecule were extracted [261]. Then, nucleotides were edited to reconstruct PIC models involving target codons in our study (Tab. 3). When editing the nucleotide (e.g. AUG → GUG), first, atoms except N1 and N9 in base group, sugar group, and phosphate group were deleted. Then, coordinates of missing atoms were inferred. All histidine residues were configured as ϵ -protonated. These molecules were soaked into a 36 Å radius water sphere (Fig. 9), neutralized by K^+ , and added 150 mM KCl. TIP3P water model was employed. VMD [226] was used to infer missing atom coordinates, solvate the model, and visualize the structure throughout the study. For the detailed procedures, please see Sec. 7.1.

Table 3. Codons Modeled in Our Simulation. Nucleotide sequence and reference atoms to define the codon-anticodon distance.

Codon	1st Nucleotide		2nd Nucleotide		3rd Nucleotide	
AUG	A	N1 and N6	U	N3 and O4	G	N1, N2, and O6
GUG	G	N1, N2, and O6	U	N3 and O4	G	N1, N2, and O6
CUG	C	N3, N4, and O2	U	N3 and O4	G	N1, N2, and O6

All simulations were carried out using NAMD (version 2.13 multi-core) [137]. CHARMM36 force-field (July 2019 update) was used [142,143]. Multilevel summation method (MSM) electrostatics [265] was employed. Cutoff at 12 Å (with

switching from 10 Å) was applied to non-bonded interactions. Temperature and pressure were set at 310 K and 1 atm, respectively; Langevin thermostat (damping coefficient: 5/ps) and Langevin-piston barostat were adopted [229,230]. All C1' (nucleotide) and C α (amino acid) atoms farther than 22 Å from the center of the system (i.e. water sphere) were restrained at their initial positions, and water molecules crossing the boundary of water sphere (radius 36 Å) were restrained. Harmonic potential functions with spring constant 10 pN/Å were adopted as the restraint of molecules. After energy minimization (10,000 steps), the system was equilibrated for 10 ns, and then simulated for 1 μ s (time-step: 2 fs); the biasing force was applied only after collecting 200 samples in the bin. Each model (Tab. 3) was simulated five times.

Adaptive biasing force (ABF) molecular dynamics method [113,262,263] was performed to evaluate multi-dimensional free energy profiles in terms of d_1 , d_2 , and d_3 (Fig. 9), which were defined as the distances of the 1st, 2nd, and 3rd base-pairs in Å, respectively (Tab. 3). Specifically, these d_i were evaluated as distance between the centers of hydrogen donor and acceptor atoms of codon and anticodon (Table 3 shows the atoms in a mRNA segment). Each d_i was sampled over $4.0 \leq d_i \leq 9.0$ with bin width $\Delta d = 0.5$ Å, and was restrained by harmonic bonds with spring constant 10 pN/Å if d_i crosses lower boundary (3.0 Å) or upper boundary (10.0 Å). The details of ABF is explained in Sec. 7.3.

3.3 Analyses and Results

3.3.1 Binding Free Energy

The free energy profile $G(d_1, d_2, d_3)$ with respect to three variables d_1 to d_3 was obtained through the analysis of ABF results. The probability of Gibbs free energy for each state $P(d_1, d_2, d_3)$ obeys Eq. 3.1:

$$P(d_1, d_2, d_3) := \frac{\exp\left(-\frac{G(d_1, d_2, d_3)}{k_B T}\right)}{\sum_{4.0 \leq d_1, d_2, d_3 \leq 9.0} \exp\left(-\frac{G(d_1, d_2, d_3)}{k_B T}\right)}. \quad (3.1)$$

Furthermore, $P(d_1, d_2, d_3)$ was averaged over the simulation trials for each model (see Sec. 3.2), and then $G(d_1, d_2, d_3)$ was evaluated as Eq. 3.2:

$$G(d_1, d_2, d_3) = -k_B T \ln P(d_1, d_2, d_3) + \text{const.}, \quad (3.2)$$

assuming $\min\{G(d_1, d_2, d_3) | 4.0 \leq d_1, d_2, d_3 \leq 9.0\} = 0$.

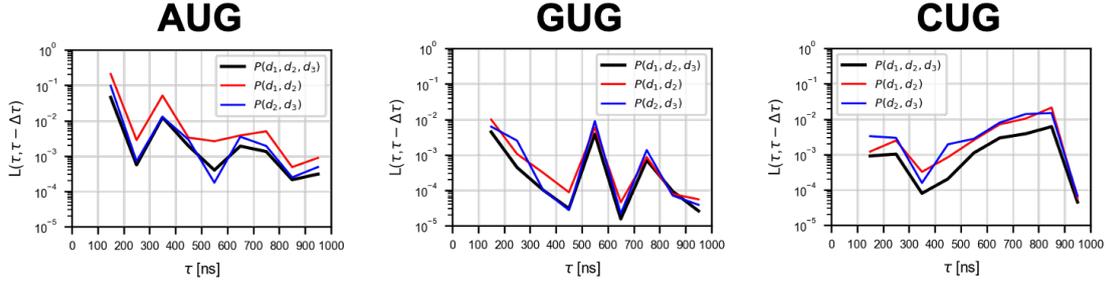


Figure 10. Convergence of Free Energy Profiles with Time Evolution. The $L(\tau, \tau - \Delta\tau)$ of $P(d_1, d_2, d_3)$, $P(d_1, d_2)$, and $P(d_2, d_3)$ are shown (see Eq. 3.3; $\Delta\tau = 100$ ns). The data points are plotted between the two consecutive time points $(\tau - \Delta\tau)$ and τ .

Probability $P(d_1, d_2, d_3)$ was obtained at equal intervals $\Delta\tau$ ($= 100$ ns) through the ABF simulation. The convergence of $P(d_1, d_2, d_3)$ with increasing simulation time steps was evaluated as follows. To compare profiles of $P(d_1, d_2, d_3)$ at two different time points (τ_1, τ_2) , squared-error function of two probability distributions was employed. These profiles of different time points τ_1 and τ_2 were represented by $P(d_1, d_2, d_3; \tau_1)$ and $P(d_1, d_2, d_3; \tau_2)$, respectively. Then, the squared-error function $L(\tau_1, \tau_2)$ was calculated as Eq. 3.3:

$$L(\tau_1, \tau_2) := \sum_{4.0 \leq d_1, d_2, d_3 \leq 9.0} (P(d_1, d_2, d_3; \tau_1) - P(d_1, d_2, d_3; \tau_2))^2 \quad (3.3)$$

The convergence of $P(d_1, d_2, d_3)$ at τ [ns] was tested by $L(\tau, \tau - \Delta\tau)$ (Fig. 10). The convergence of $P(d_1, d_2)$, $P(d_2, d_3)$, and $P(d_1, d_3)$ was evaluated in the same

way. Almost small difference of L and fast convergence could be observed in AUG and GUG, while could not observed in CUG. This suggest that to confirm convergency of the result of CUG, further long time, many trial of simulation is expected.

Then, to evaluate the free energy difference between the codon-anticodon bound and unbound states, the free energy scores G_{bound} and G_{unbound} , and their gap $\Delta G_{\text{binding}}$ were defined as Eq. 3.5:

$$\begin{aligned}
 G_{\text{bound}} &:= \frac{\sum_{4.0 \leq d_1, d_2, d_3 \leq 6.0} G(d_1, d_2, d_3) P(d_1, d_2, d_3)}{\sum_{4.0 \leq d_1, d_2, d_3 \leq 6.0} P(d_1, d_2, d_3)} \\
 G_{\text{unbound}} &:= \frac{\sum_{7.0 \leq d_1, d_2, d_3 \leq 9.0} G(d_1, d_2, d_3) P(d_1, d_2, d_3)}{\sum_{7.0 \leq d_1, d_2, d_3 \leq 9.0} P(d_1, d_2, d_3)} \\
 \Delta G_{\text{binding}} &:= G_{\text{bound}} - G_{\text{unbound}}
 \end{aligned} \tag{3.4}$$

Here the bound and unbound states are defined as $\forall i : 4.0 \leq d_i \leq 6.0$ and $\forall i : 7.0 \leq d_i \leq 9.0$, respectively (Fig. 11). The distance range for the bound state corresponds to the codon-anticodon (AUG-CAU) structure [264], and that for the unbound state is based on a previous research that described unbound conformation of the complex [266]. G_{bound} and G_{unbound} were hence weighted average of $G(d_1, d_2, d_3)$ in the ranges of bound and unbound states, respectively (Eq. 3.5).

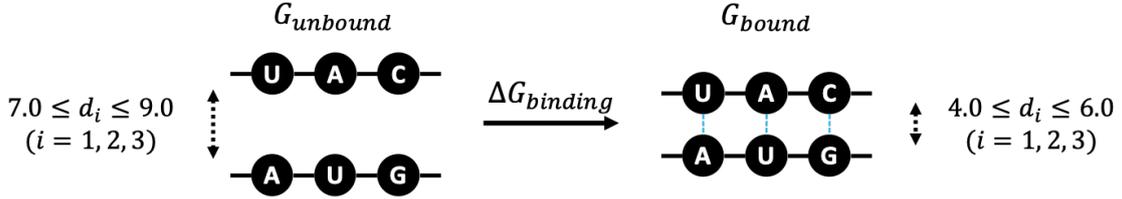


Figure 11. Binding Free Energy. Schematic representation of G_{bound} , G_{unbound} , and $\Delta G_{\text{binding}}$ (Eq. 3.5).

$$\begin{aligned}
 G_{\text{bound}} &:= \frac{\sum_{4.0 \leq d_1, d_2, d_3 \leq 6.0} G(d_1, d_2, d_3) P(d_1, d_2, d_3)}{\sum_{4.0 \leq d_1, d_2, d_3 \leq 6.0} P(d_1, d_2, d_3)} \\
 G_{\text{unbound}} &:= \frac{\sum_{7.0 \leq d_1, d_2, d_3 \leq 9.0} G(d_1, d_2, d_3) P(d_1, d_2, d_3)}{\sum_{7.0 \leq d_1, d_2, d_3 \leq 9.0} P(d_1, d_2, d_3)} \\
 \Delta G_{\text{binding}} &:= G_{\text{bound}} - G_{\text{unbound}}
 \end{aligned} \tag{3.5}$$

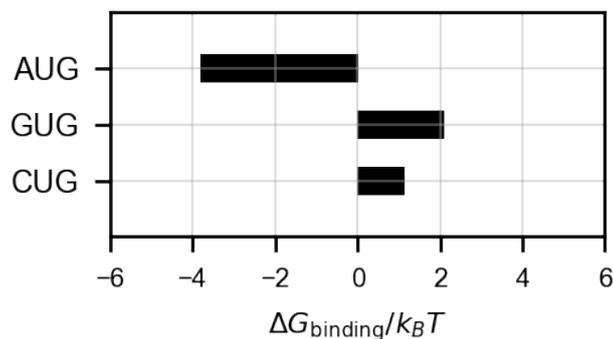


Figure 12. Estimated Binding Free Energy. $\Delta G_{\text{binding}}$ scores (Eq. 3.5; see also Fig. 11) are shown. The scores were obtained from $P(d_1, d_2, d_3)$ averaged over five simulation trials for each model.

The free energy scores of bound and unbound states (G_{bound} and G_{unbound} ; see Eq. 3.5) averaged over five simulation trials (Fig. S7), and presented their difference $\Delta G_{\text{binding}}$ (Fig. 12) were evaluated. In the case of the cognate start codon AUG, $\Delta G_{\text{binding}}$ must be negative to stabilize the initiation of translation, and was indeed $\sim -4 k_B T$. In contrast, the GUG codon, less frequently used as a start codon [258,267], showed a positive $\Delta G_{\text{binding}} \sim 2 k_B T$. For the CUG codon, which is considered as a stronger start codon than GUG [258,267], $\Delta G_{\text{binding}}$ showed an intermediate value $\sim 1 k_B T$. Thus, $\Delta G_{\text{binding}}$ accounts for observed initiation rates from the respective start codons.

Furthermore, the difference of free energy score of bound state for two different ribonucleotides (Tab. 3) was calculated to confirm consistency to different computational approach [261]. Assuming that G_{unbound} is common for all the models, i.e. the free energy of the unbound state is independent of the codon, G_{bound} and $\Delta G_{\text{binding}}$ are equivalent. The difference, or penalty, of binding free energy ($\Delta\Delta G$) induced by AUG \rightarrow GUG and AUG \rightarrow CUG substitution was $\simeq 6 k_B T$ (3.6 kcal/mol) and $\simeq 5 k_B T$ (3.0 kcal/mol), respectively. This result is largely consistent with another computational approach using the free energy perturbation (FEP) [261].

3.3.2 Projected Free Energy

To visualize the profiles, $G(d_1, d_2, d_3)$ was projected (marginalized) onto 2-dimension, as $G(d_1, d_2)$ in Eq. 3.6. Projected profile $P(d_1, d_2)$ was obtained from $G(d_1, d_2)$.

$$G(d_1, d_2) := \frac{\sum_{4.0 \leq d_3 \leq 9.0} G(d_1, d_2, d_3) P(d_1, d_2, d_3)}{\sum_{4.0 \leq d_3 \leq 9.0} P(d_1, d_2, d_3)} \quad (3.6)$$

$$P(d_1, d_2) := \frac{G(d_1, d_2)}{\sum_{4.0 \leq d_1, d_2 \leq 9.0} G(d_1, d_2)}$$

$G(d_1, d_3)$, $G(d_2, d_3)$ and $P(d_1, d_3)$, $P(d_2, d_3)$ were defined in the same way.

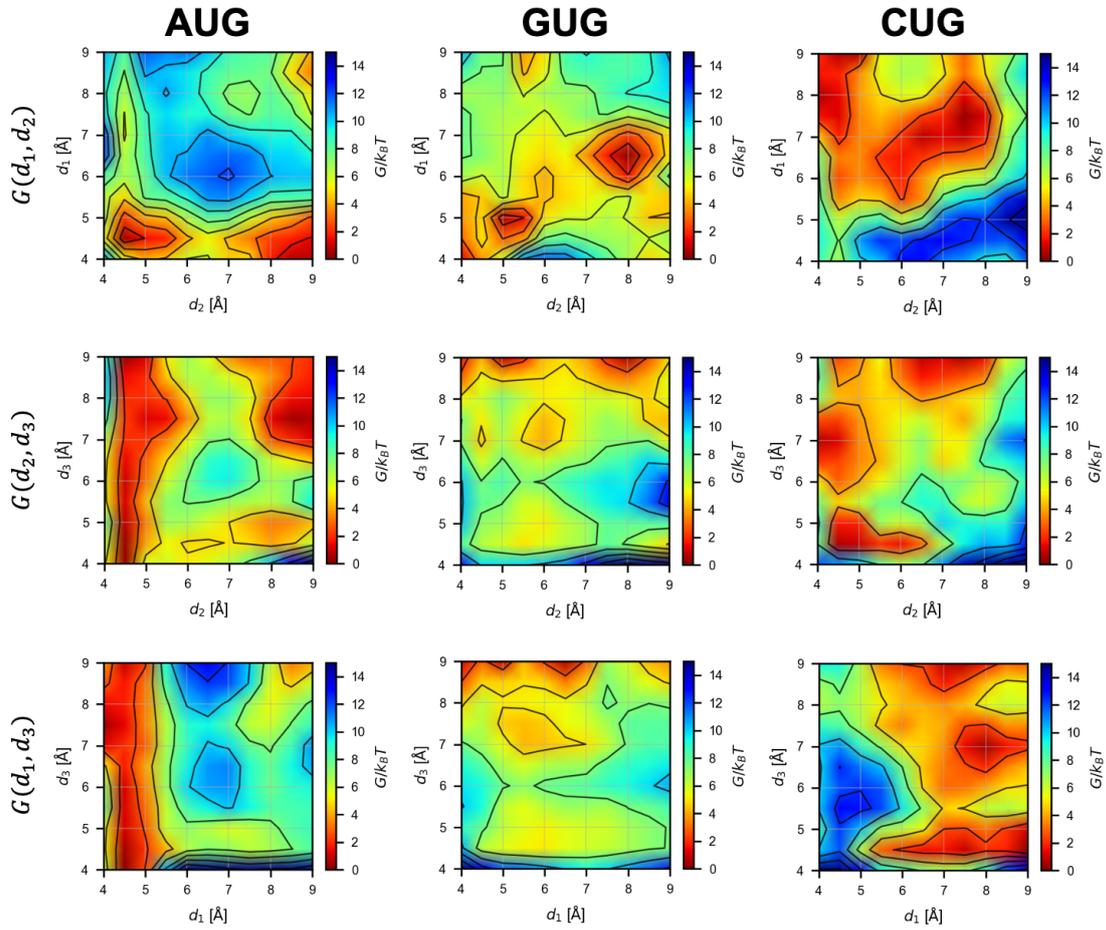


Figure 13. Projected Free Energy Profiles. Profiles of $G(d_1, d_2)$, $G(d_2, d_3)$, and $G(d_1, d_3)$ obtained from Eq. 3.6 for each model is shown by contour plots.

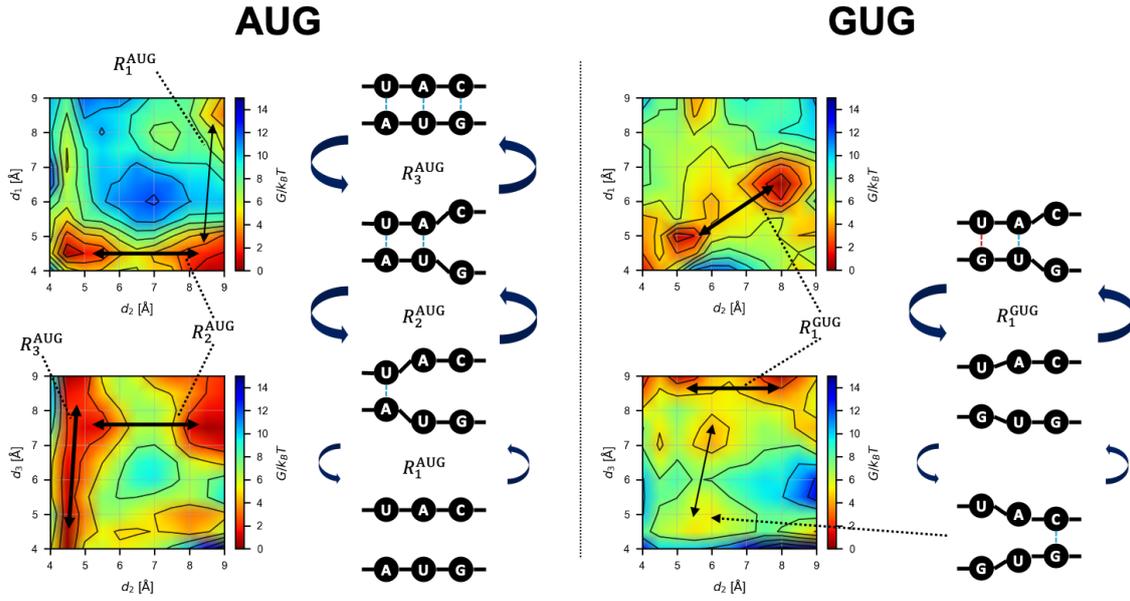


Figure 14. Schematics of the Base-Pair Binding Dynamics. Conformational changes inferred from the free energy landscape (Fig. 13). The transition path R_n^\bullet (\bullet is AUG or GUG) is shown by black arrows.

The projected free energy profiles are shown in Fig. 13, and the transition dynamics were inferred. The suggested paths and their schematics are shown in Fig. 14. In the case of the AUG start codon, the following transitions were expected in the AUG-CAU dynamics in equilibrium (Fig. 14). Starting from the bound state, d_3 shows large fluctuations while d_1 and d_2 show small ones (R_3^{AUG}). d_1 and d_2 are bistable (bound and unbound), and once the 3rd G:C base-pair is broken (large d_3), the 2nd U:A base-pair may become unbound (R_2^{AUG} to the large d_2 state). Only after that, the 1st A:U base-pair dissociates (R_1^{AUG} to the large d_1 state); this is expected to occur less frequently due to the higher barrier than those for R_2^{AUG} and R_3^{AUG} . Starting from the unbound state and reversing the process above, the AUG codon should bind to the CAU anticodon from the side of the 1st A:U base-pair, followed by the 2nd and then 3rd base-pairs (Fig. 14). This result suggests that the recognition of the 1st A:U base-pair is very important for the accurate start codon recognition, in agreement with the role and location of eIF1 in the P-site [264, 268] (see Fig. 16 below).

In the case of the GUG codon, d_1 and d_2 show the transition (R_1^{GUG}) between two distinct (metastable) states (both bound and both almost unbound) as shown by $G(d_1, d_2)$ in Fig. 14, while d_3 is mostly high. Binding of the 3rd base-pair (transition to lower d_3) is possible but less frequent, and simultaneous binding of the 1st and 3rd base-pairs is rare (Fig. 13). As expected, the affinity of the 1st base-pair (wobble G:U) is lower than the case of AUG. This result is consistent with infrequent GUG initiation observed in the previous works [258, 267].

In the case of the CUG codon, many meta-stable states were observed as shown in Fig. 13. Transition paths seem to be more complicated than the AUG and GUG cases. Although concurrent binding of the 2nd and 3rd base-pairs (lower d_2 and

d_3) is possible, the 1st base-pair cannot form simultaneously with these other base-pairs (Fig. 13), which makes the CUG pairing unstable compared to AUG base-pairing. Overall, however, the binding free energy $\Delta G_{\text{binding}}$ is lower for CUG than for GUG (Fig. 12) (see below). Note that, technically, the rugged free energy landscape (Fig. 13) demanded more computational cost for the ABF sampling, as suggested by the slow convergence shown in Fig. 10.

3.3.3 Binding Dynamics from the Structural Viewpoint

To observe binding dynamics from the structural viewpoint, typical structures, or atomic coordinates corresponding to a specific reaction coordinate (d_1, d_2, d_3) , were obtained by averaging the sampled atomic coordinates as follows. Here, we assumed that the reaction coordinate (d_1, d_2, d_3) is represented by $(\tilde{d}_1, \tilde{d}_2, \tilde{d}_3)$ if Eq. 3.7 is satisfied. Then, each atomic coordinate was averaged over all the snapshots (sampled at 10 ps intervals) corresponding to the representative reaction coordinate $(\tilde{d}_1, \tilde{d}_2, \tilde{d}_3)$.

$$\tilde{d}_i - \frac{1}{2}\Delta d \leq d_i < \tilde{d}_i + \frac{1}{2}\Delta d \quad (i = 1, 2, 3) \quad (3.7)$$

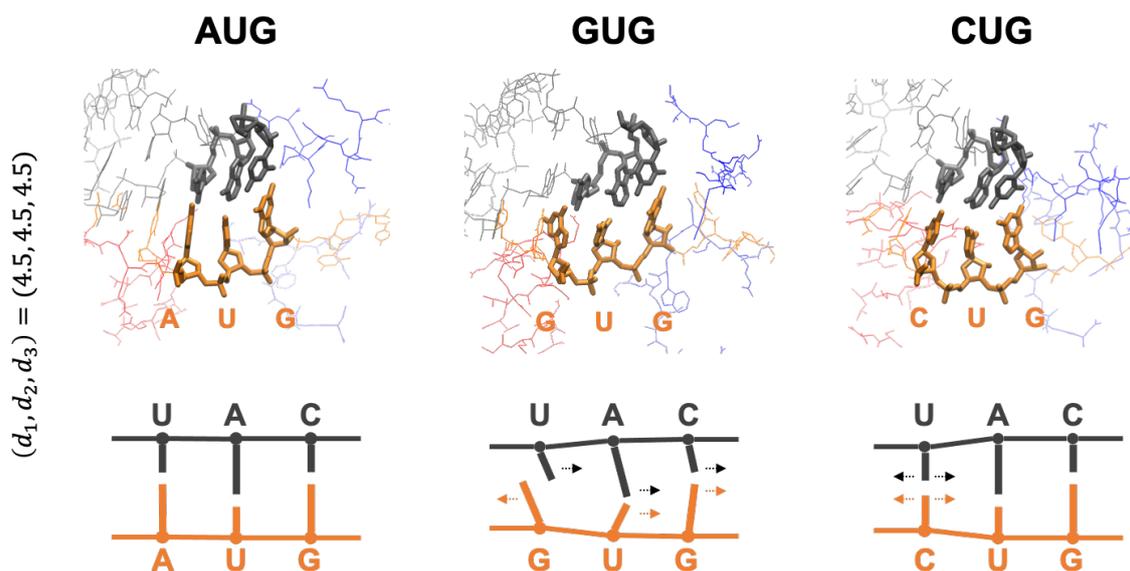


Figure 15. Typical Bound Structures. Averaged structures corresponding to reaction coordinate $(\tilde{d}_1, \tilde{d}_2, \tilde{d}_3) = (4.5, 4.5, 4.5)$ (see Eq. 3.7). In the structures (top), nucleotides of the codon and anticodon are drawn by thick lines. In the schematics (bottom), lines show the direction of the bases. mRNA and tRNA are drawn by orange and gray, respectively. Red and blue lines are parts of eIF1 and eIF1A, respectively.

The averaged structures and schematics of the codon and anticodon are shown in Fig. 15. In the case of AUG, the averaged bound-state structure is ordered and tightly bound. It is reasonable, as it is the correct start codon, and the binding free energy is negative (Fig. 12). Note that eIF1 and eIF1A molecules (shown in red and blue in Fig. 15, respectively) are present near the AUG-CAU base-pair. It was experimentally suggested that these proteins contribute to the accurate start codon recognition [247, 259, 260, 264, 269]. (see Fig. 16).

By contrast, the structure of the GUG-CAU base-pair is disordered (Fig. 15). The mismatched bases (the 1st G:U) avoid each other (rather than forming a

wobble base-pair) and the uracil in tRNA tilts toward the 2nd U:A base-pair. The directions of the 2nd and 3rd base-pairs were consequently affected, resulting in the unstable bound state (Fig. 12). Although the projected free energy profile $G(d_1, d_2)$ (Fig. 13) suggests cooperative binding of the 1st and 2nd base-pairs (Fig. 15), the 3rd base-pair is mostly separate, which may prevent the recognition of the GUG start codon.

In the case of CUG, the structure is relatively ordered (Fig. 15). Although the 1st C:U base-pair is mismatched, cytosine is smaller than guanine and adenine (purine bases), which may mitigate steric hindrance at the 1st position. As shown in Fig. 13, many meta-stable conformations are possible, which we propose to be attributed to combinations of bound and unbound conformations of the base-pairs. It is therefore reasoned that some near-bound states can occasionally allow translation initiation at this codon.

3.4 Discussion and Conclusion

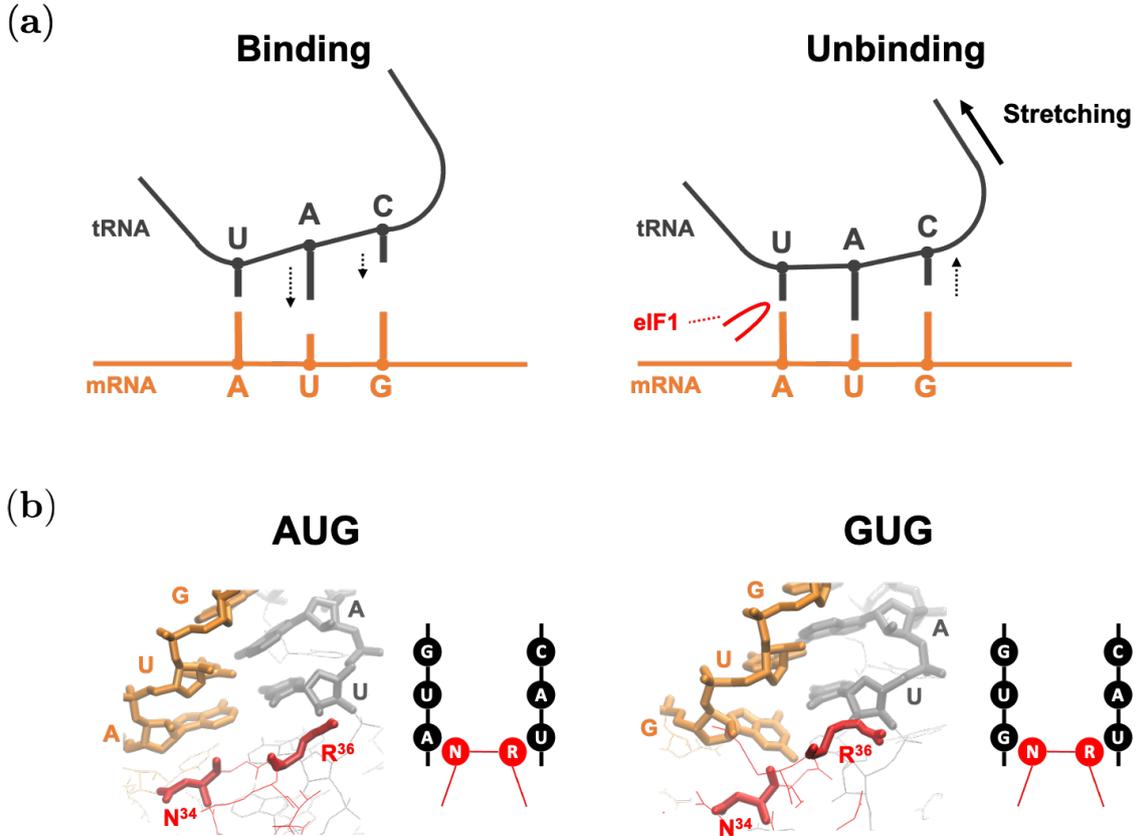


Figure 16. Binding Mechanism Conjectured by This Study. mRNA and tRNA are drawn by orange and gray, respectively. Red lines show parts of eIF1. (a) Schematics of binding and unbinding dynamics. (b) Averaged structures corresponding to reaction coordinate $(\tilde{d}_1, \tilde{d}_2, \tilde{d}_3) = (4.5, 4.5, 4.5)$ (see Eq. 3.7), and their schematic representations. Nucleotides of the codon and anticodon, and N^{34} (Asn-34) and R^{36} (Arg-36) are drawn by thick lines.

Our ABF-based approach provided not merely the binding free energy but information on the nucleic acid binding dynamics represented by the free energy landscape (Figs. 13 and 14), in contrast to the previous work. The free energy profiles shown in Fig. 13 suggested an unexpected stability of the 1st A:U base-pair, compared to the 3rd G:C base-pair. According to the free energy profile of AUG binding, dissociation of the triplet base-pairs starts at the 3rd G:C (Fig. 14, right column). In the open PIC model that is suggested to occur during the scanning process prior to start codon recognition [264], the tRNA is not perpendicularly attached to the mRNA, in contrast to the P-site tRNA positioning during the elongation phase. This conformation appears to allow the 5'-side (i.e. cytosine side) of the anticodon to curve away from the start codon, suggesting a stretching force towards the tRNA side (Fig. 16(a)). We propose that this stretching decreases the affinity of the 3rd G:C base-pair during the scanning process (Fig. 16(a)). In contrast, the affinity of the 1st A:U base-pair is likely to be increased

by interaction with eIF1, so is that of the 2nd U:A base-pair by eIF1A, as proposed previously [247, 259, 260, 264, 268, 269] (Fig. 16(a)). In strong agreement with the role of eIF1 in stabilizing the 1st A:U base-pair, our averaged simulation structure indeed positions Asn-34 and Arg-36 in its proximity (Fig. 16(b)). In fact, the residues Asn-34:Gly-35:Arg-36, termed β -hairpin loop 1, is absolutely conserved from yeast to human. Mutations altering Asn-34 and Arg-36 display significant increase in UUG initiation [270], in agreement with their crucial role in maintaining open scanning-competent PIC conformation.

The free energy landscape of GUG-anticodon base-pairs (Fig. 13) and its averaged simulation structure in the P-site (Figs. 15 and 16(b)) also suggest that the same structural restriction in turn prevents G:U pairing at the 1st position, that otherwise occurs frequently in its free form. The disordered 3rd G:C base-pair seen with the GUG structure appears to be consistent with this idea (Fig. 15). Since we did not observe a strong disorder in CUG-anticodon structure (Fig. 15), we propose that the near-cognate start codon usage characteristic of eukaryotic initiation is mostly explained by a strong perturbation on GUG accommodation in the P-site due to steric restriction imposed by eIF1 β -hairpin loop. In agreement with this thesis, the level of CUG initiation is just equivalent to that of GUG initiation in yeast *S.cerevisiae* [271] (and personal observations of Dr. Asano), although the former is significantly stronger than the latter in various distinct contexts in human cells [272].

4 Molecular Dynamics Analysis of Partially Disassembled Nucleosomes

This section is based on refs. [204, 205].

4.1 Introduction

Nucleosome is a fundamental unit of eukaryotic chromosome [273–275]. The nucleosome consists of around 150 base-pairs of DNA and 4 types of 8 histone proteins [239]. DNA wrapping structure around histone proteins makes long eukaryotic genome compact [276, 277]. This fact is convincing with the genomic length of the human genome, 3.4 \AA (base-step) \times 30 billion (base-pair counts) \times 2 (diploid). In addition, histone proteins are well known to play biological roles other than chromatin compaction [278, 279]. Histone proteins have two regions, stable "core" and disordered "tail" [280, 281]. Chemically modified tails serve as various biochemical signals [282–284]. For example, methylation of lysine at residue index 9 of histone H3 (H3K9me3) induces condensed nucleosomes by several factors, then makes the locus heterochromatin [285]. Huge numbers of biological researches were published about histone modifications [280, 281, 283, 286, 287].

As mentioned above, nucleosomes have the potential for biological roles. However, the nucleosome is a structure in which DNA wraps around histone proteins [239]. How the nucleosomes (histone proteins) behave through the interaction of DNA binding proteins along with transcription? An answer is that, nucleosomes may work as physical obstacles [288, 289]. Actually, histone proteins are known to be evicted completely or partially in such situations, by several modulation factors [290–293]. Nucleosomes that lack some histone proteins (intermediate nucleosomes; or "partially assembled nucleosome" [294]) have been studied in recent years, and the following facts have been shown. (i) DNA in partially disassembled nucleosomes shows higher accessibility than that in canonical nucleosomes, i.e. partially disassembled nucleosomes facilitate binding of proteins to DNA [295]. (ii) the dissociation frequency of histone H2A/H2B is significantly higher than that of H3/H4, and nucleosomes lacking only H3/H4 have not been experimentally observed [296]. However, analysis of mechanical properties of the nucleosome as a complex of biomolecules, which underlie previous observation [295], have not proceeded.

In this research, we analyzed dynamics of these partially disassembled nucleosomes by computational approaches, to evaluate their mechanical features [204, 205]. In particular, nucleosome models of canonical one [239], and partially disassembled one lacking a histone dimer H2A/H2B (from the canonical structure), were constructed. Then, their dynamics were compared by fully atomic molecular dynamics simulation. Additionally, we also constructed an H3/H4-lacking model for comparison.

4.2 Materials and Methods

The nucleosome structure which consists of the histone octamer and human centromere palindromic sequences (α -satellite), published in PDB (PDB ID : 1KX5 [239]) was employed as canonical nucleosome structure in this study. Then, one of the (four) histone dimers was deleted, to obtain partially disassembled nucleosome structures lacking one histone dimer. Note that components of each pair of proteins H3/H4 and H2A/H2B are known to show strong binding affinity at the atomic scale [297, 298]. Thus, the units of lacking histone proteins were H3/H4 and H2A/H2B. We named nucleosomes lacking H3/H4, H2A/H2B, H3'/H4', and H2A'/H2B' (Tab. 4) as Δ H3/H4, Δ H2A/H2B, Δ H3'/H4', and Δ H2A'/H2B', respectively (Fig. 17).

Table 4. Residue IDs of Core Region and Tail Region for each Histones. Each chain ID and residue ID refer to PDB ID: 1KX5 [239]. Only histone H2A has a C-terminal tail.

Histone (Chain ID)	Core Region	Tail Region
H3 (A) and H3' (E)	45 – 135	1 – 44
H4 (B) and H4' (F)	25 – 102	1 – 24
H2A (C) and H2A' (G)	18 – 98	1 – 17 (N-terminal) 99 – 128 (C-terminal)
H2B (D) and H2B' (H)	35 – 122	1 – 34

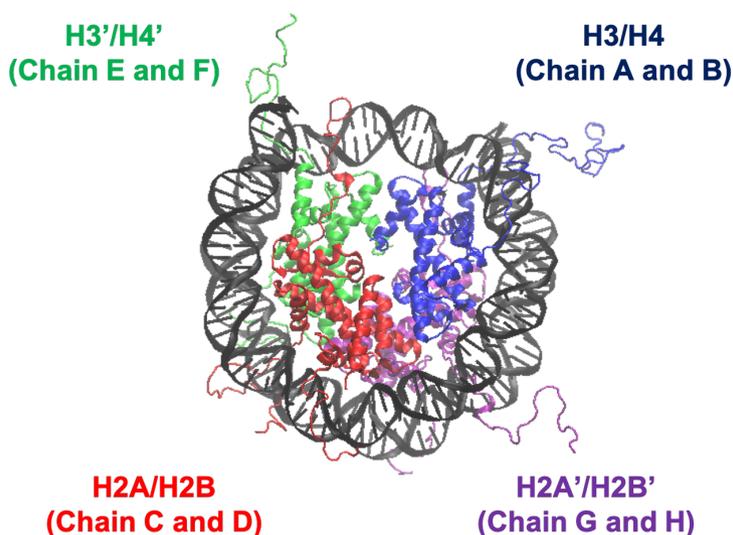


Figure 17. Nucleosome Structure and Histone Indices. PDB ID: 1KX5. Histones H3/H4 (Chain A and B), H2A/H2B (Chain C and D), H3'/H4' (Chain E and F), and H2A'/H2B' (Chain G and H) are shown in blue, red, green, and purple, respectively. Following these indices, we named each nucleosome lacking one of the histone-dimers as Δ H3/H4, Δ H2A/H2B, Δ H3'/H4', and Δ H2A'/H2B'.

To solvate these nucleosome models, TIP3P water molecules and KCl were

added to construct the water box. The water box was neutralized by K^+ and added 150 mM KCl (see Sec. 7.1). Periodic boundary condition with Particle-Mesh Ewald (PME) [227,228], non-bonded interactions with 12 Å cutoff (with switching from 10 Å), Langevin thermostat (damping coefficient: 5/ps) and Langevin-piston barostat were adopted in this simulation [229,230]. Temperature and pressure were set at 310 K and 1 atm, respectively. In the numerical (molecular dynamics) simulation, first, 10,000 steps of energy minimization was carried out with restricting the positions of $C\alpha$ atoms in amino acids and $C1'$ atoms in nucleotides, at initial coordinates (i.e. crystal structure) by harmonic restraint (spring coefficient: 100pN/Å). Then, 100 ns production run was carried out. Each model was simulated ten times. Numerical simulation was performed using NAMD (version 2.12 multicore with CUDA) [137,138].

4.3 Analyses and Results

4.3.1 Structures of Histone Proteins

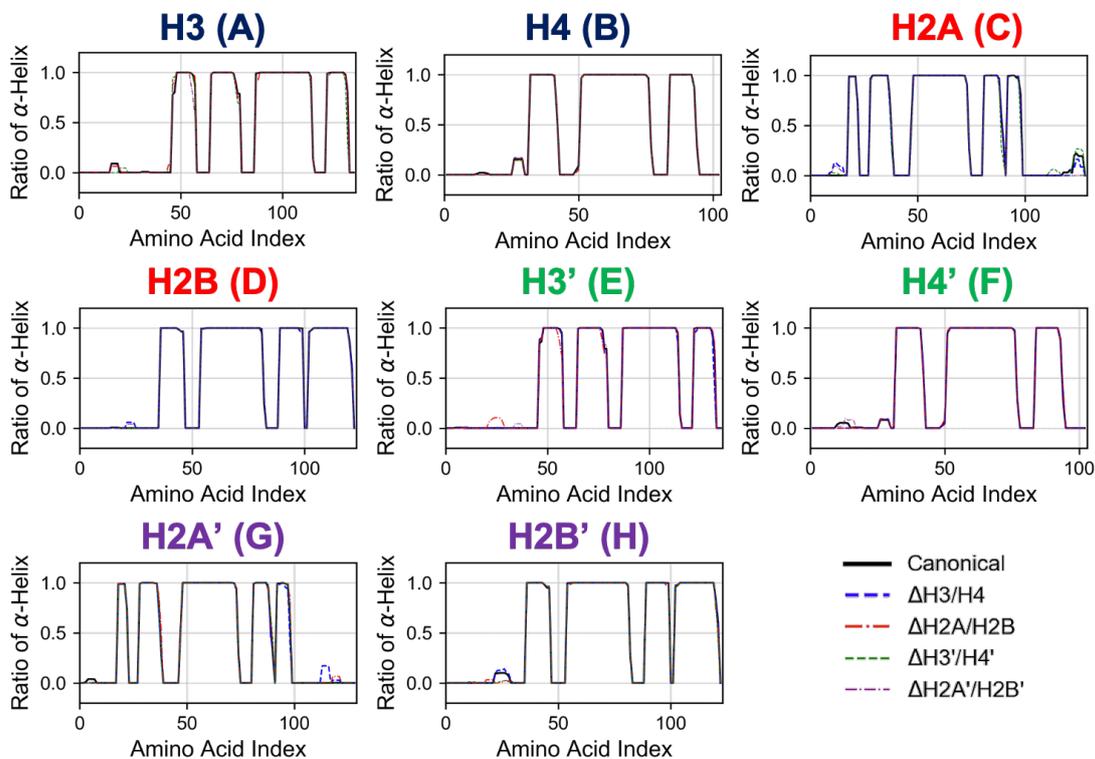


Figure 18. Secondary Structure Stability of Histones in Each Nucleosome. Averaged over ten trials (each trial is shown in Fig. S8). Horizontal and vertical axes of each graph show amino acid index and ratio of α -helix formation in respective histones as indicated in figure.

At first, we evaluated ratios of secondary structure formation and collapse of histone proteins in the remaining nucleosome structure, by partial dissociation of histones. We employed DSSP [299], which is often used to determine the secondary structure of proteins. It is already known that histone proteins have only several α -helices in core regions [239]. Then, we only calculate α -helix formation ratio in all histones over ten simulation trials in each nucleosome model (Fig. 18). As shown in the results, profiles of formation ratio are hardly different, and the ratios of each α -helix forming residue show ~ 1.0 (100 %). This result suggests that partial histone dissociation does not affect nucleosome structures, considering the fact: canonical nucleosome shows much stable structure i.e. conformation of histone proteins hardly change.

The α -helix formation was slightly observed in tail regions (Fig. 18). These agreed with previous research that tail regions transiently show α -helix formation when interacting with DNA [300,301].

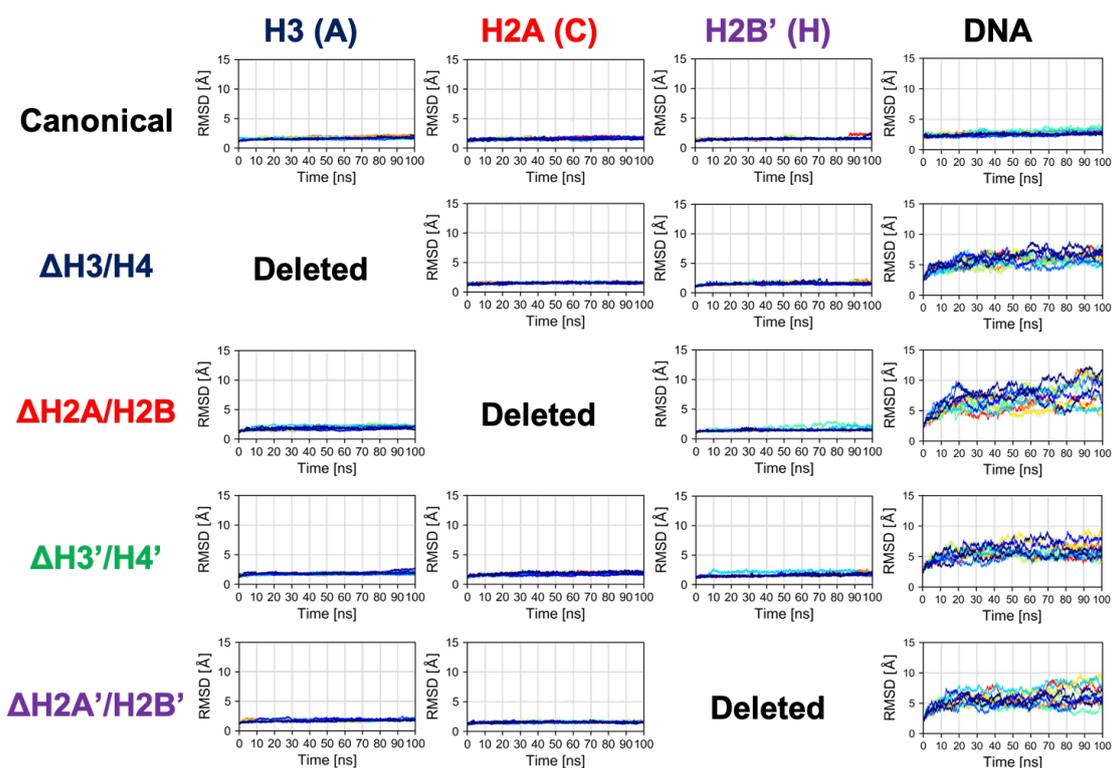


Figure 19. Time Series of RMSD of Histone and DNA in Each Nucleosome. Colors show different simulation trajectories (total 10).

Next, we evaluated RMSD (Root Mean Square Displacement) of histones and DNA for each nucleosome model, to focus on the tertiary structure of these molecules. At the time of RMSD computations, we eliminated histone tails (Tab. 4) in removing translation and rotation of nucleosome molecules, because histone tails (intrinsically disordered regions) show much unstable conformations [302,303]. Atomic coordinates just after energy minimization were referred to align the simulation trajectories, and we note those transformed coordinates of atom i as \hat{x}_i . The time series of RMSD of each molecule for each nucleosome model was shown in Figs. 19 and S9. Regarding all histone proteins, there are negligible differences between canonical nucleosomes and partially disassembled nucleosomes. This result well agreed with experimental observations [304,305]. On the other hand, only DNA in partially disassembled nucleosomes showed an increase of RMSD, while DNA in the canonical nucleosome did not show. This result (and including DSSP result) suggests that partial dissociation of histone proteins affect only wrapping DNA.

4.3.2 Structural Deformation of Nucleosomal DNA

According to the analysis above, partial dissociation of histone proteins only affected DNA dynamics (Fig. 19). To evaluate the changes in DNA dynamics, we evaluate RMSF (Root Mean Square Fluctuation) of DNA. At the time of RMSF computations, we aligned simulation trajectories based on only core regions of histone proteins (not the same as RMSD calculation part), and note the coordinate of atom i as \hat{x}_i . Then, with focusing on only C1' atoms of nucleotides, RMSF is defined as $\langle |\hat{x}_i - \langle \hat{x}_i \rangle_{\text{time}}|^2 \rangle_{\text{time}}$, where $\langle \cdot \rangle_{\text{time}}$ means the averaging of all simulation trajectories. RMSFs were calculated over ten simulation trials for each nucleosome model (Fig. 20). Note that $\langle \cdot \rangle_{\text{time}}$ means averaging over the latter half of simulation trajectory (50.01 – 100 ns) and ten trials for the same nucleosome model. As overall trends, several nucleotides in partially disassembled nucleosomes, originally interacted to dissociated histones, show higher RMSD scores than others.

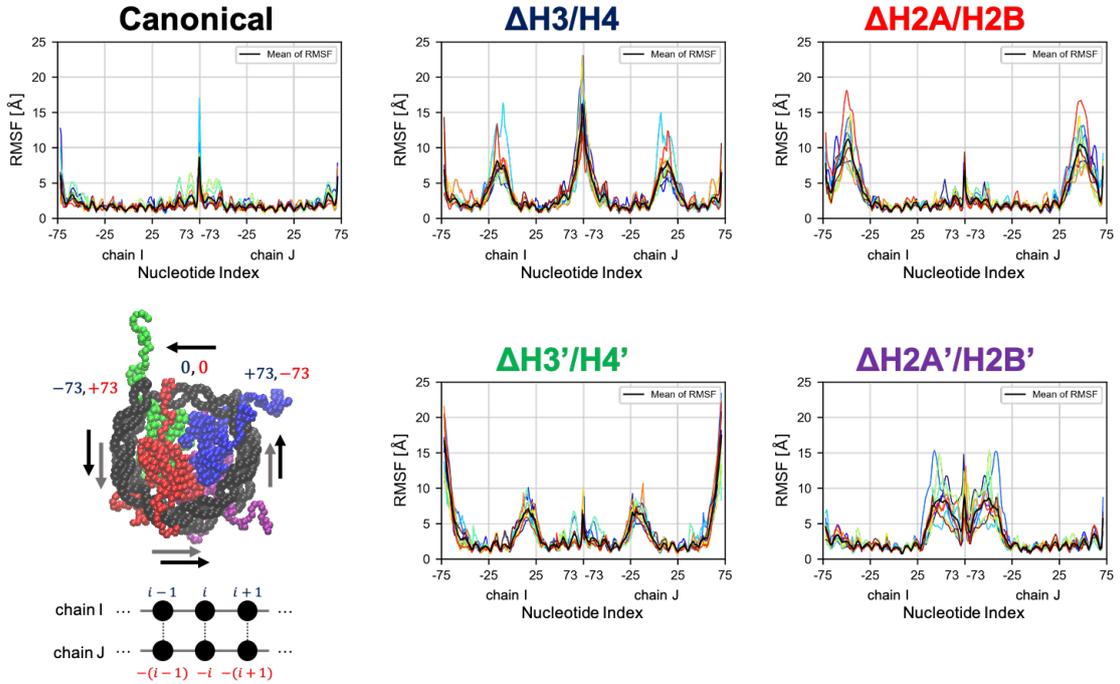


Figure 20. Root Mean Square Fluctuation (RMSF) of DNA. Each graph shows RMSF profiles of respective nucleosomal DNA. Horizontal axis shows nucleotide indices i . Colors show different simulation trajectories (total ten). The nucleotide indices are indicated in the bottom-left panel (each particle represents a nucleotide or an amino acid); the indices in chains I and J (-73 to 73 each in the PDB file) are shown in blue and red, respectively.

To evaluate not only magnitudes of non-directive fluctuation (RMSF; Fig. 20), but also effective structural deformation of DNA, we employed PCA (Principal Component Analysis). PCA has been used in a similar case of protein dynamics analysis [306]. In particular, the coordinate vector of time τ : $X(\tau) = (\hat{x}_1^T(\tau), \hat{x}_2^T(\tau), \dots, \hat{x}_N^T(\tau))^T$, which consists of nucleotides positions \hat{x}_i (eliminated translation and rotation), was defined. Then, eigenvalue decomposition of the co-

variance matrix of $X(\tau)$ was carried out. Note that in obtaining covariance matrix, τ s in only the latter half of simulation trajectory over ten simulation trials were used, to eliminate the dependency of the initial structure. The eigenvectors \mathbf{v}_j were sorted by descending order of eigenvalues, and the largest \mathbf{v}_j was expected to correspond to the most drastic deformation. In fact, the two largest motions mostly corresponded to the DNA breathing outward from the nucleosome (Figs. 21 and S10).

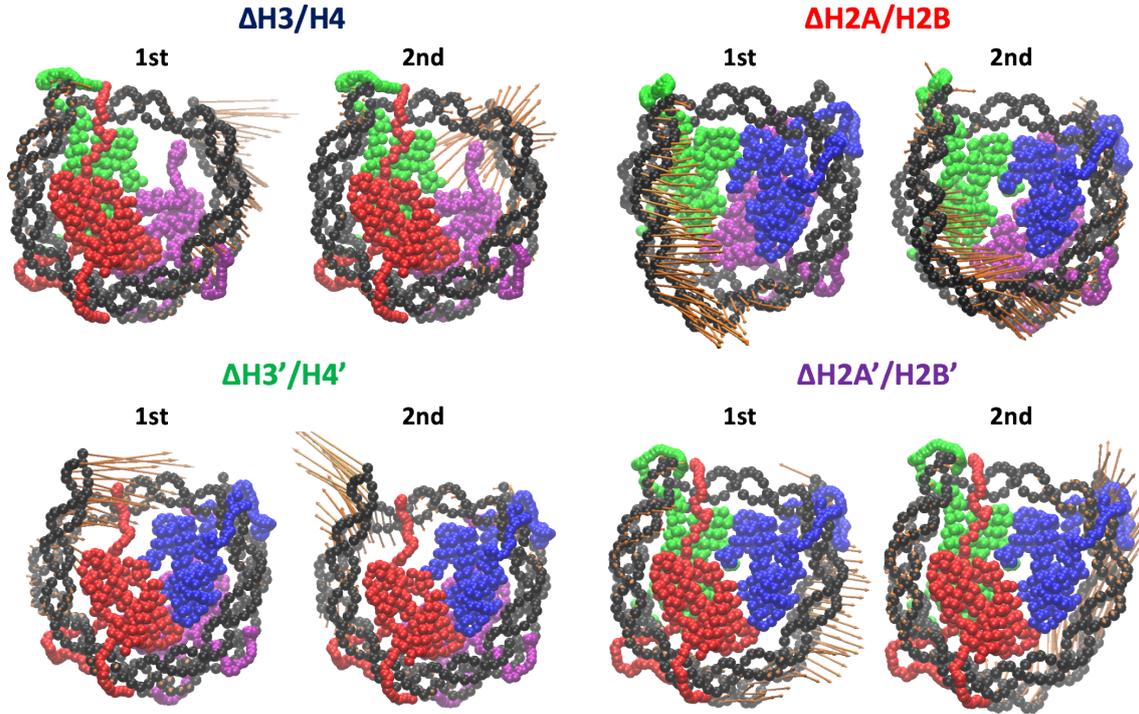


Figure 21. Effective Directions of DNA Deformation (Front View). Visualized from the side of chains A to D (H3, H4, H2A, and H2B). Nucleotide (C1' of each nucleotide) and amino acid (C $^{\alpha}$ of each amino acid) coordinates are shown by spheres. Orange arrows show the vectors of the 1st and 2nd PC modes. Arrow lengths indicate the norm of the vector.

Deformation level of $X(\tau)$ for the j -th largest deformation direction $S(\tau, j)$ was defined as $S(\tau, j) := (X(\tau) - \langle X \rangle_{\text{time}}) \cdot \mathbf{v}_j$. Profiles $S(\tau, j)$ of two largest j over ten simulation trajectories for each model were shown in Fig. 22. The plots were smoothed using Kernel Density Estimation (KDE). $S(\tau, j)$ of initial conformation ($S(0, 1)$, $S(0, 2)$) was shown in magenta. Individual time series of simulation trajectories were drawn simultaneously (Fig. 22).

Following different profiles depending on dissociated histone dimers were observed from Fig. 22. In the case of H2A/H2B-lacking nucleosome models (Δ H2A/H2B and Δ H2A'/H2B'), initial coordinate ($S(0, 1)$, $S(0, 2)$) locates far from high probability regions, i.e. DNA behavior in equilibrium is much different from the initial histone-wrapping conformation. Individual simulation trajectories show drift and irreversible transition. Thus, H2A/H2B dissociation induces high mobility and irreversible transition of DNA structure. In contrast, in the

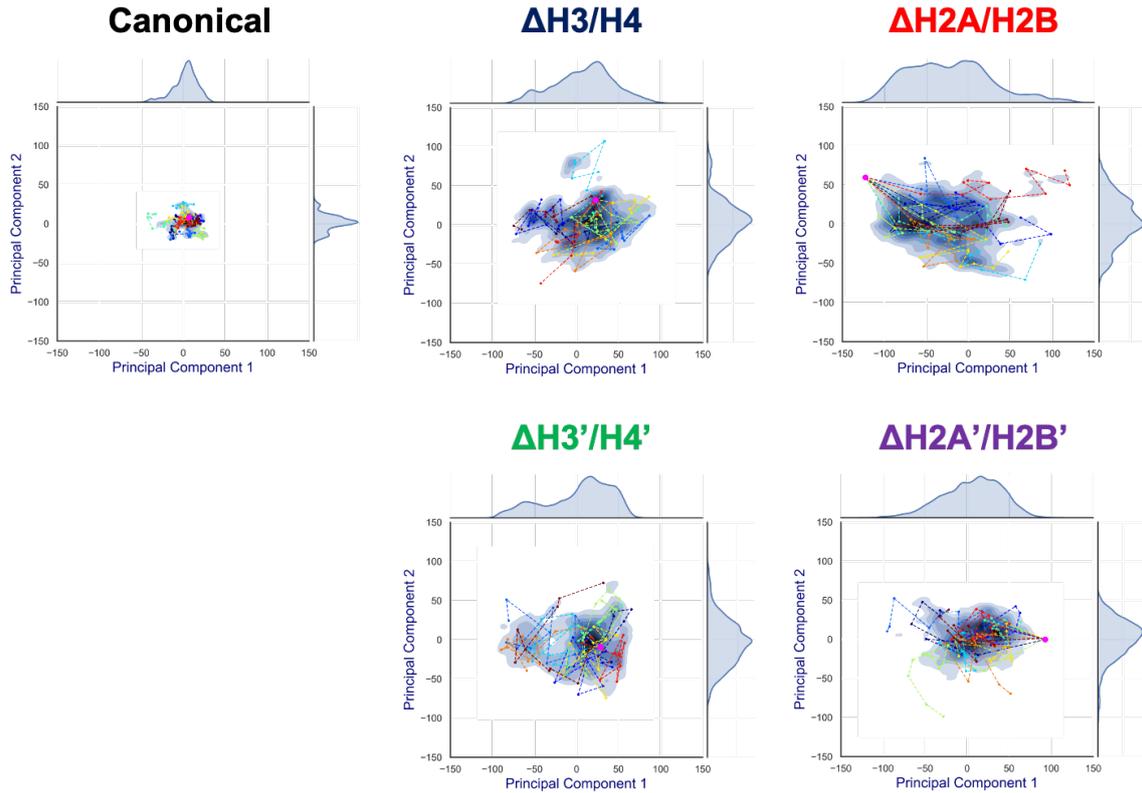


Figure 22. Free Energy Landscape of DNA Dynamics to 1st and 2nd PC Modes. The axes correspond to deviations ($S(t, \text{mode})$) to the 1st and 2nd PC vector directions from the averaged structure ($= S(t, 1)$ or $S(t, 2)$). The color (blue) shows the probability density of S . Each profile was fitted by kernel density estimation (KDE). Magenta dot corresponds to the initial conformation ($S(0, 1)$, $S(0, 2)$), and each line (in different colors) shows a simulation trajectory (total ten; plotted at 10 ns intervals).

case of H3/H4-lacking nucleosome models ($\Delta\text{H3}/\text{H4}$ and $\Delta\text{H3}'/\text{H4}'$), initial coordinate ($S(0, 1)$, $S(0, 2)$) locates near high probability regions. In other words, DNA behavior in equilibrium is similar to initial histone wrapping conformation. Many of simulation trajectories distribute around ($S(0, 1)$, $S(0, 2)$) location. Contrary to expectations, this result shows that dissociation of H3/H4 hardly induces irreversible DNA deformation.

Overall, the following fact was shown: DNA deforming dynamics in partially disassembled nucleosomes depends on dissociated histone proteins (Fig. 22).

4.3.3 Interactions Between Histone Tails and DNA

Considering the result of PCA (Fig. 22), the following question was raised: What causes the difference of DNA dynamics dependent on lacking histone proteins? In the protein (secondary and tertiary) structure analysis (Figs. 18 and Fig. 19), no apparent difference was observed, with excluding histone tails from targets. Hence, we evaluated contact frequency between the DNA and histone tails in canonical and partially disassembled nucleosomes (Fig. 23). It is to be noted that the histone tails were not counted in aforementioned analysis.

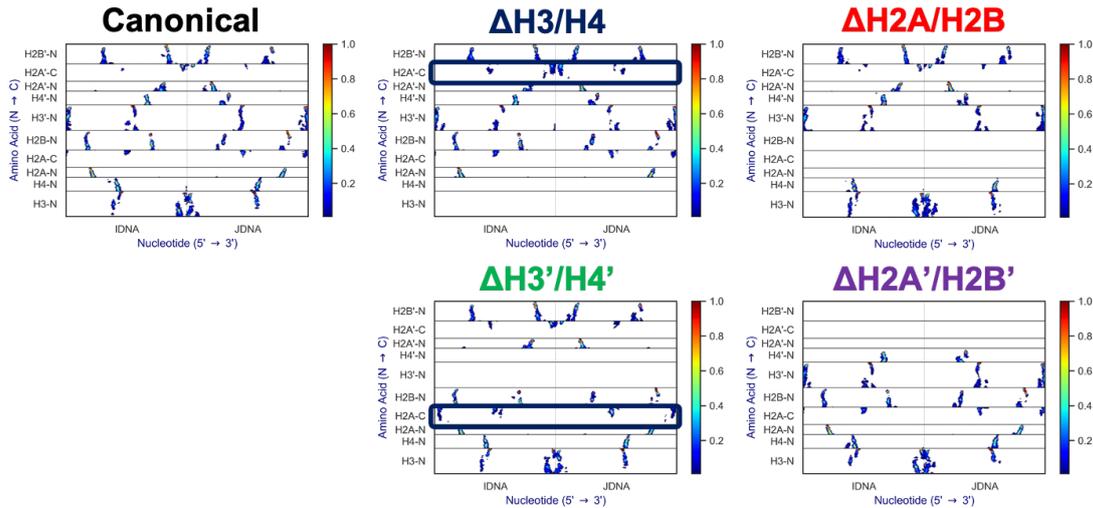


Figure 23. Contact Matrix of DNA and Histone Tails. Contact frequency of nucleotide of DNA and amino acid of histone tails.

In comparison among all types of nucleosomes, no apparent differences were observed in the majority of components, which is consistent with a previous MD simulation study showing that tails were trapped by the adjacent DNA [196]. Only in H3/H4-lacking nucleosomes ($\Delta H3/H4$ and $\Delta H3'/H4'$), most obvious differences are interaction frequency between DNA and C-terminal tail of H2A (H2A-C). Considering this result, we found the following trend by observations of simulation trajectories: H2A-C invades the space induced by H3/H4 dissociation. This invasion induces additional interaction of H2A-C and DNA, which originally contacted with lacking H3/H4. Then, the mobility of that DNA was restricted though fluctuation levels of that DNA increased. This restriction makes DNA conformation similar to the initial histone wrapping DNA structure (Fig. 24).

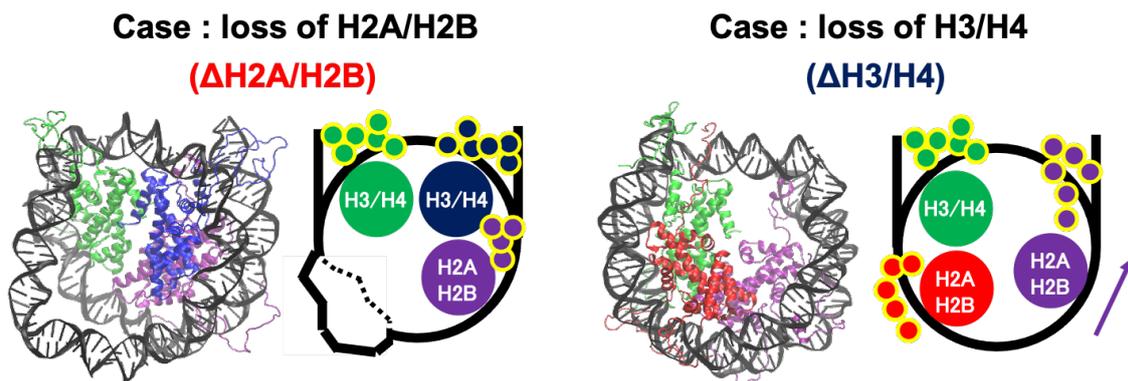


Figure 24. Histone Tail Dynamics in Δ H3/H4 and Δ H2A/H2B. Snapshots from simulation trajectories and schematics of Δ H3/H4 and Δ H2A/H2B are shown. Histone tails are represented by small circles in the schematics.

4.4 Discussion and Conclusion

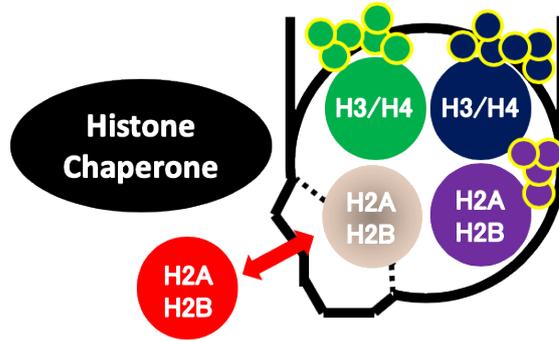
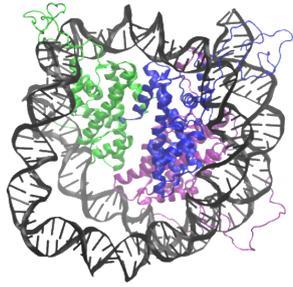
In this study, histone tail dynamics in partially disassembled nucleosomes, and their biological roles were investigated [204, 205]. In the structural analysis of histone proteins remaining in the nucleosomes, no significant deformation was observed (Figs. 18 and 19). Specifically, the nucleosome models lacking an H2A/H2B (Δ H2A/H2B and Δ H2A'/H2B'; observed in experiments) show similar RMSD to canonical one. We also confirmed that no significant differences were found between the H2A/H2B-lacking structure and the hexasome part in the overlapping di-nucleosome [307], (RMSD: 2.554 Å). Thus, it is suggested that drastic structural deformation occurs only in the DNA in partially disassembled nucleosomes. Structures of histones (remaining in nucleosomes) seem almost the same as the canonical structure.

On the other hand, dynamics of DNA were apparently different between the partially disassembled nucleosomes lacking H2A/H2B and those lacking H3/H4 (Fig. 22). In the H2A/H2B-lacking models (Δ H2A/H2B and Δ H2A'/H2B'), long H3 tails were trapped at the dyad parts and the ends of nucleosomal DNA (Fig. 24). Because of the trapping of H3 tails, free space induced by the loss of H2A/H2B was not filled. Then, DNA that interacted with the dissociated H2A/H2B shows breathing spontaneously. In the H3/H4-lacking models (Δ H3/H4 and Δ H3'/H4'), H2A-C region invades the space induced by the loss of lacking H3/H4 (Fig. 24). Then, additional interaction of DNA and H2A-C occurs. This interaction restricts DNA breathing.

The obtained result can be associated with biological experiments as follows. H2A/H2B detachment seems effective modulation based on the dynamics of H2A/H2B-lacking nucleosome. The H2A/H2B dissociation enhances mobility of DNA which originally contacted with the H2A/H2B (Fig. 24). Actually, this modulation process is employed in intra-cellular function [291]. On the other hand, detachment of H3/H4 is considered inefficient, because H3/H4 dissociation induces additional interaction of DNA and H2A-C, then increase of DNA mobility cannot be expected (Fig. 24).

Contrary, let us suppose that how to restore partially disassembled nucleosomes to canonical (histone octamer) nucleosome. Insertion of H2A/H2B is thought to be easy because the space enhanced by high DNA mobility remains in the H2A/H2B-lacking nucleosomes (Fig. 25). On the other hand, the loss of H3/H4 induces H2A-C invasion. Then, the space originated from H3/H4 dissociation will be quickly filled (Fig. 25). It is inconvenient in the case of insertion of H3/H4. Thus, H2A/H2B detachment and attachment are preferred by dynamic features of molecular complexes [204, 205].

Loss of H2A/H2B



Loss of H3/H4

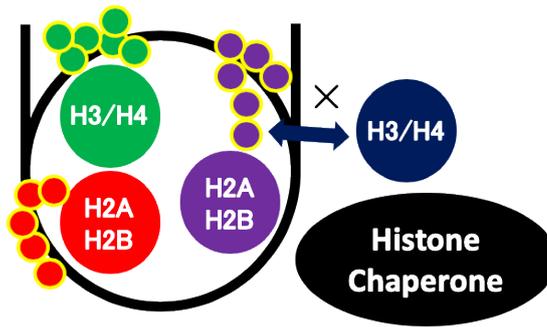
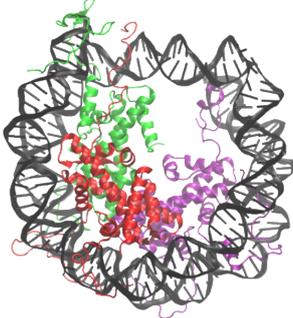


Figure 25. Suggested Schematics of Nucleosome Remodeling based on the Results. Schematics of histone dissociation and nucleosome remodeling by histone chaperone (like FACT). The relationship between partially disassembled nucleosome dynamics and the process of nucleosome remodeling is suggested.

5 Conclusion and Future Problems

In this thesis, the problem mentioned in the introduction part (Sec. 1) — physical properties of molecular systems involving nucleic acids — was considered and discussed.

In Section 2, dynamics of double-stranded DNA involving several methylation patterns was simulated, and structural deformation was evaluated at base-pair resolution (Fig. 5). The analysis was performed from physical viewpoint, i.e. first-principles approach [202]. This strategy is employed in the field of soft matter physics [308,309]. Particularly, I could obtain DNA dynamics depending on chemically modified sites. This result is expected to be a basis of site-specific nucleotide recognition by DNA binding proteins [206–208]. However, these sequence-dependent dynamics of nucleic acids are considered as cooperative dynamics of constituent atoms [202]. Then, the application of the analysis method to further longer DNA should be an extremely difficult problem. For example, coarse-grained models proposed by our group [157,158] cannot incorporate atomic-scale structures. Of course, chemical modification cannot be considered. Hierarchical molecular dynamics modeling and analysis of nucleic acids are expected to be discussed and developed in the future (it is also my major research theme).

In Section 3, dynamics of two RNA segments (mRNA and tRNA) involved in the molecular complex of other RNAs (rRNAs) and proteins (eIFs) were simulated, and multi-dimensional free energy profiles in terms of their base-pairs were evaluated [203]. Distances of base-pairs (codon and anticodon) were employed as reaction coordinates to discuss their cooperativeness. Practically, obtained free energy profiles reproduced the difference of binding strength of base-pairs (A:U and G:C). The difference of these profiles provided the sensitive changes of binding affinity and transition path among AUG, GUG, and CUG. There are few studies that focused on precise binding dynamics of nucleic acids, like this study. Therefore, the following extension can be considered. First, the same molecular dynamics analysis of additional codons and association with biological experiments, are wanted. These additional works will be published in the future with experimental researches. The mechanisms of precise translation initiation by various factors (eIFs) also should be evaluated in the same way [261,310]. Next, the proposed method (multi-dimensional ABF) of RNA binding is to be applied to other environments. As mentioned in Section 1.4, bound and folded structures of RNA are historically studied by not the dynamics (physical) but mathematical estimation based on match/mismatch/gap score of base-pair [311–313]. Studies of secondary structure prediction of RNA are typical examples [184]. Of course, structural predictions have been currently conducted from physical viewpoints [314–316], analysis method of RNA dynamics based on reaction coordinate may provide more appropriate physiological RNA conformations and their transient pathway. In the analysis of the biological function of non-coding RNA from the viewpoint of interaction, the dynamics analysis method should be important [165].

In Section 4, the statistical analysis method for the dynamics of DNA, which

shows strong interaction with proteins, is considered by focusing on nucleosomes [204, 205]. PCA (Principal Component Analysis), which has been used to obtain effective structural deformation of protein [306], was employed in this study. Structural transition of DNA was detected through the analysis of long-time DNA dynamics and their probability distributions using PCA. However, though this PCA-based method revealed transitions, it is not assured that this method is appropriate for other biomolecular systems. As a further development, statistical dynamics analysis methods to detect effective structural deformations should be discussed. To evaluate effective structural changes from simulation data, classification method by time-scale of deformations based on ICA (Independent Component Analysis) [317–319] is developed in the recent years [111, 320–322], although PCA (covariance-based method) is often used. In the equilibrium state, slower vibration has a larger amplitude assuming the law of equipartition of energy [323]. With assuming that low frequency and slower structural motion is expected to be larger. Therefore, long time simulation trajectories should provide further effective dynamics by the ICA method, along with the development of computer science.

It is much important to distinguish sequence-dependent structural affinity and sequence-independent electrostatic affinity from the interaction of DNA and proteins in the molecular complex [179, 180]. This would be a breakthrough in the field of structural biology. For example, that may reveal mechanisms of the determination of nucleosome positioning and its stability [288, 324–326], which is a major, completely unsolved problem in chromatin biology. It may assist in designing nucleotide recognizing proteins (e.g. TALEN [327, 328] and CRISPR [329–331]) as their recognition is based on molecular dynamics of proteins and nucleic acids. Its application is not limited to those molecules but also to studies on biological information processing and network analysis based on molecular interactions [332–334]. Efficient structural sampling method may be appropriate and solve that problem. For example, REST2 considers only systematic regulation of electrostatic interaction [114, 335].

Currently, in the field of life science, research methods except experiment are largely classified into two types; numerical calculation-based molecular simulation and informatics (and sequencing) based bioinformatics [184, 336–338]. In the bioinformatics field, though dynamics and energy profiles are hardly discussed, an exhaustive analysis of molecular systems involving nucleic acids is enabled based on mathematical science [31, 32]. Although these computational and informational sciences cannot be considered closely related, I consider that both advantages should be utilized simultaneously. Particularly, in macroscopic chromosomal scale, polymer physics and sequencing analysis have cooperated with each other, and discovered additional knowledge [339–342]. In microscopic few-molecular systems, many experimentally measured biological phenomena that might be associated with molecular dynamics are suggested [343–345]. The importance of model (dynamics) and data-driven computational biology is expected to be recognized in the future.

6 Acknowledgments

These works were supported by RIKEN Junior Research Associate Program (<https://www.riken.jp/en/careers/programs/jra/index.html>), "TSUBAME Encouragement Program for Young/Female Users" of Global Scientific Information and the Computing Center at Tokyo Institute of Technology (https://www.gsic.titech.ac.jp/en/encouragement_program), and RIKEN supercomputer HOKUSAI GreatWave (<https://i.riken.jp/en/>).

I am grateful to supervisors Dr. Togashi (Hiroshima Univ.) and Dr. Awazu (Hiroshima Univ.) for their polite guidance and many help in research and writing papers. Dr. Togashi provided me the basic advice on computational biology and theoretical biophysics since master's course. I could not complete this thesis if I had not received his helpful guidance. Dr. Awazu gave me rigorous and appropriate advice to obtain the basic knowledge of theoretical and mathematical physics since undergraduate. I also thank Dr. Nishimori (Meiji Univ.).

In the DNA methylation study (Sec. 2), I am grateful to Dr. Suzuki (Nagoya Univ.) for fruitful discussions from planning to publish this research [202]. Not referred in this thesis, much bioinformatic analysis was performed in this research. Through the analysis, I had received many helpful supports and advice from Dr. Nikaido (Tokyo Medical and Dental Univ.), Dr. Ozaki (Tsukuba Univ.), and Dr. Matsumoto (Nagasaki Univ.). They also belong to Laboratory for Bioinformatics Research, RIKEN Center for Biosystems Dynamics Research (BDR; I also belong; <https://bit.riken.jp>).

In the RNA binding study (Sec. 3), Dr. Asano (Kansas Univ.) provides me the fundamental knowledge of ribosomes and much appropriate advice for the collaborative research [203] since the doctor course. This research required a long time, then I worry him, but his help was conscientious. I hope the proposed computational method (Sec. 3) will be helpful for his future biological researches.

I had several difficulties in the nucleosome dynamics study (Sec. 4) because this was the research of structural biology for the first time [204, 205]. During that study, Dr. Kawasaki (Hiroshima Univ.) and Dr. Saikusa (AIST) gave me opportunities for fruitful discussions. This experience was a nice opportunity to decide the concept of future research activities (computational biology, chemistry, and drug discovery).

Additionally, in the researches not referred in this thesis, I am grateful to Dr. Shinkai (RIKEN BDR), Dr. Sugawara (Tokyo Univ.), Dr. Tate (Hiroshima Univ.), Dr. Tochio (Teikyo Univ.), Dr. Sakamoto (Hiroshima Univ.), Mr. Ikegaya (Hiroshima Univ.), and Mr. Matsushita (Hiroshima Univ.) for fruitful discussions and helpful advice. Many help of everyone made me be able to complete researches and this thesis. I would like to take this opportunity to thank you.

7 Additional Information

7.1 Procedures of Molecular Modeling

In this thesis, molecular dynamics (MD) simulation was employed (Secs. 2 ~ 4). Following execution programs and datasets were employed.

- **VMD: Visual Molecular Dynamics** [226]. VMD is developed by Theoretical and Computational Biophysics group in Illinois University (<https://www.ks.uiuc.edu>). VMD provides molecular structure modeling, molecular dynamics analysis, and molecular structure visualization. VMD is used before and after numerical simulation using NAMD. Solvation and ionization also can be done by VMD.
- **NAMD: Scalable Molecular Dynamics** [137,138]. Numerical molecular simulation program package developed in the same place as VMD. NAMD is famous for providing the fast numerical simulation of large-scale systems by efficient parallel computing. Accommodating tutorials can be obtained from the web site (<https://www.ks.uiuc.edu/Training/Tutorials/>).
- **CHARMM Force Fields: Chemistry at HARvard Macromolecular Mechanics** [129,130]. Parameters of molecular structures and interaction force fields developed by chemistry laboratory in Harvard University (<https://www.charmm.org>). CHARMM force-field includes not only typical atomic interaction functions but also an additional potential function of restraining peptide bond surface of the protein. Though mainly employed in studies of proteins and lipids in the past, CHARMM has been employed in studies of nucleic acids and complex with proteins after a major update to c36 version [142,143].

Therefore, target molecules were modeled by VMD and simulated by NAMD, employing CHARMM36 force-field in the study in this thesis. NAMD also provides Adaptive Biasing Force (ABF) method [113,262] in Collective Variables (CV) modules (<https://colvars.github.io/colvars-refman-namd/colvars-refman-namd.html>). Furthermore, elimination of translate and rotate motions of molecules, and RMSD calculation can be performed by loading NAMD outputs trajectories using VMD. VMD also extracts and outputs the data of coordinates, velocities, and potential energies of molecules, then these data can be analyzed by our own scripts.

7.2 Analysis of DNA Conformation Using X3DNA

Structure analysis using fully atomic scale methods is included in VMD. However, these are typically appropriate to proteins, and not very suitable for DNA. Then, to analyze DNA at atomic scale, we employed X3DNA (<http://x3dna.org>). X3DNA can construct atomic coordinates of DNA with selected base-pair and base-step parameters, and analyze structural degrees of DNA (<http://home.x3dna.org/highlights/schematic-diagrams-of-base-pair-parameters>). In this research [158,202], atomic DNA coordinates were constructed using the helical parameter [43,346,347]. Typical structural degrees of DNA are shown in Fig. S1.

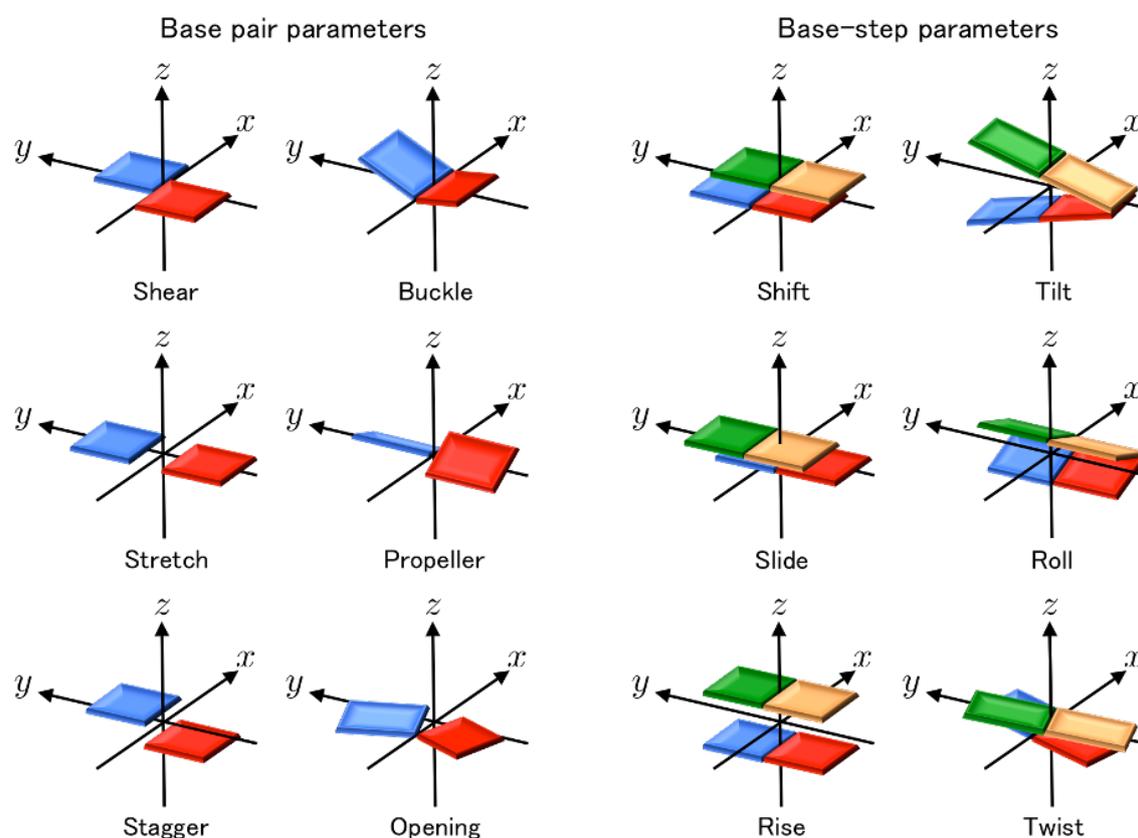


Figure S1. Base-Pair and Base-Step Parameters [232,233]. Schematics of base-pair and base-step parameters are drawn. Bases are shown by planes.

7.3 Theory of Adaptive Biasing Force (ABF) Method

In Section 3, adaptive biasing force (ABF) method was employed to estimate free energy profiles in terms of defined reaction coordinates [113, 262, 263]. The procedure of ABF is briefly explained below [262].

In multi-body systems (e.g. biomolecular systems), their correlating dynamics often play essential roles in molecular dynamics. The dynamics are characterized by "order parameters", "reaction coordinates" etc. in statistical physics of biomolecular dynamics. Dynamic behavior of the reaction coordinates, noted as ξ in the following discussion, is characterized by free energy profile $G(\xi^*)$ obeying

$$G(\xi^*) = -k_B T \int \exp\left(\frac{-\mathcal{H}(\mathbf{p}, \mathbf{x})}{k_B T}\right) \delta(\xi(\mathbf{x}) - \xi^*) d\mathbf{p} d\mathbf{x} \quad (7.1)$$

and then, by separation of hamiltonian $\mathcal{H}(\mathbf{p}, \mathbf{x})$ into kinetic energy term and enthalpy term $H(\mathbf{x})$ assuming that integration covers enough phase space (\mathbf{x}, \mathbf{p}) , Eq. 7.1 is regarded as

$$G(\xi^*) = -k_B T \int \exp\left(\frac{-H(\mathbf{x})}{k_B T}\right) \delta(\xi(\mathbf{x}) - \xi^*) d\mathbf{x}. \quad (7.2)$$

This equation is an enthalpy-based (potential-based) free energy evaluation. However, this formula is often not useful. Therefore, to obtain a free energy profile $H(\xi^*)$ based on its derivative function is better. The formula is as follow.

$$\frac{dG}{d\xi}(\xi^*) = \frac{\int \frac{dH}{d\xi} \exp\left(\frac{-H}{k_B T}\right) \delta(\xi(\mathbf{x}) - \xi^*) d\mathbf{x}}{\int \exp\left(\frac{-H}{k_B T}\right) \delta(\xi(\mathbf{x}) - \xi^*) d\mathbf{x}} := \left\langle \frac{\partial H}{\partial \xi} \middle| \xi^* \right\rangle \quad (7.3)$$

Eq. 7.3 means that derivative function of A can be evaluated as potential of mean force function (average of $dH/d\xi$). Then, the potential of mean force is equivalent to the average of acting force on ξ as Eq. 7.4.

$$\langle F_\xi \mid \xi^* \rangle = -\frac{dH(\xi^*)}{d\xi} \quad (7.4)$$

Thus, the force acting on ξ averaged over enough sampling is expected to be close to 0 when reverse force $-\langle F_\xi \mid \xi^* \rangle$ is applied. In other words, to estimate the profile of A , enough sampling with acting reverse force makes ξ as diffusion-like motion. This is an essential logic of ABF.

The averaged force (Eq. 7.4) is estimated as follows. First, the phase space of ξ^* was divided (each bin is labeled by k). In the bin k , mean force is historically evaluated as

$$\mathbf{F}_\xi(N_{\text{step}}, k) = \frac{1}{L(N_{\text{step}}, k)} \sum_{i=1}^{L(N_{\text{step}}, k)} \mathbf{F}_i(t_i^k) \quad (7.5)$$

and

$$\mathbf{F}_i(t_i^k) = \frac{d}{dt} \left(M_\xi \frac{d\xi}{dt} \right) \Big|_{t_i^k} \quad (7.6)$$

where M_ξ is mass matrix of generalized coordinates ξ , $L(N_{\text{step}}, k)$ is count at bin k obtained from simulation trajectory, and N_{step} is threshold (i.e. necessary count to start to apply biasing force). After the integration of \mathbf{F}_ξ , deviation of A is estimated [262].

For the detailed settings of ABF, please see Collective Variables (CV) modules of NAMM [137,138].

7.4 Supplemental Figures of Section 2

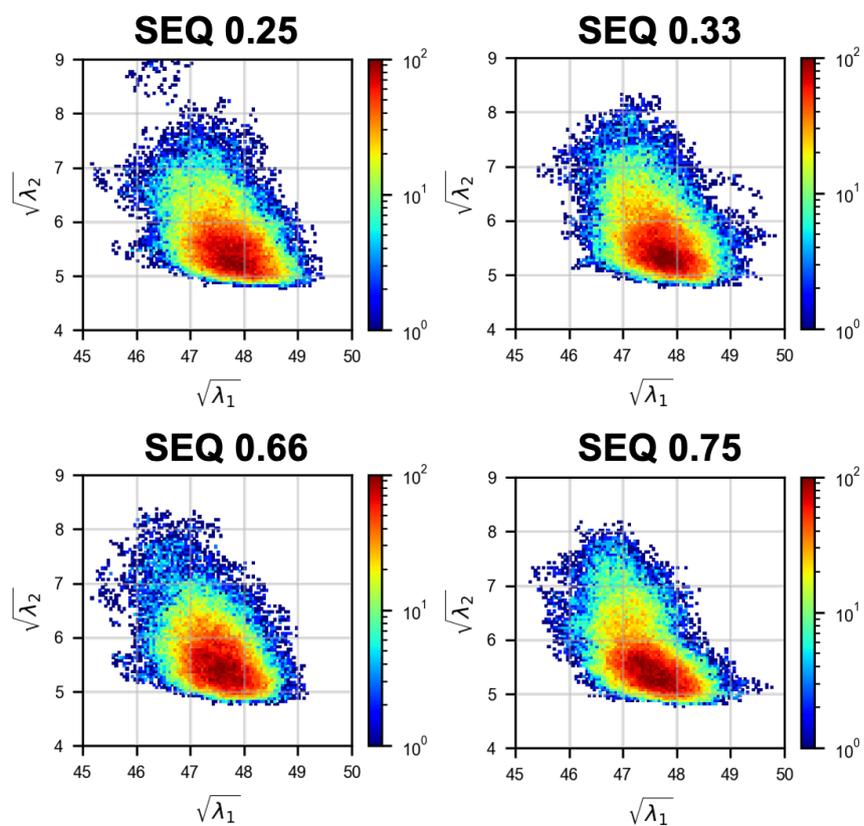


Figure S2. Profiles of Overall Geometry of DNA. The distribution of $(\sqrt{\lambda_1}, \sqrt{\lambda_2})$ summed over ten trials is shown.

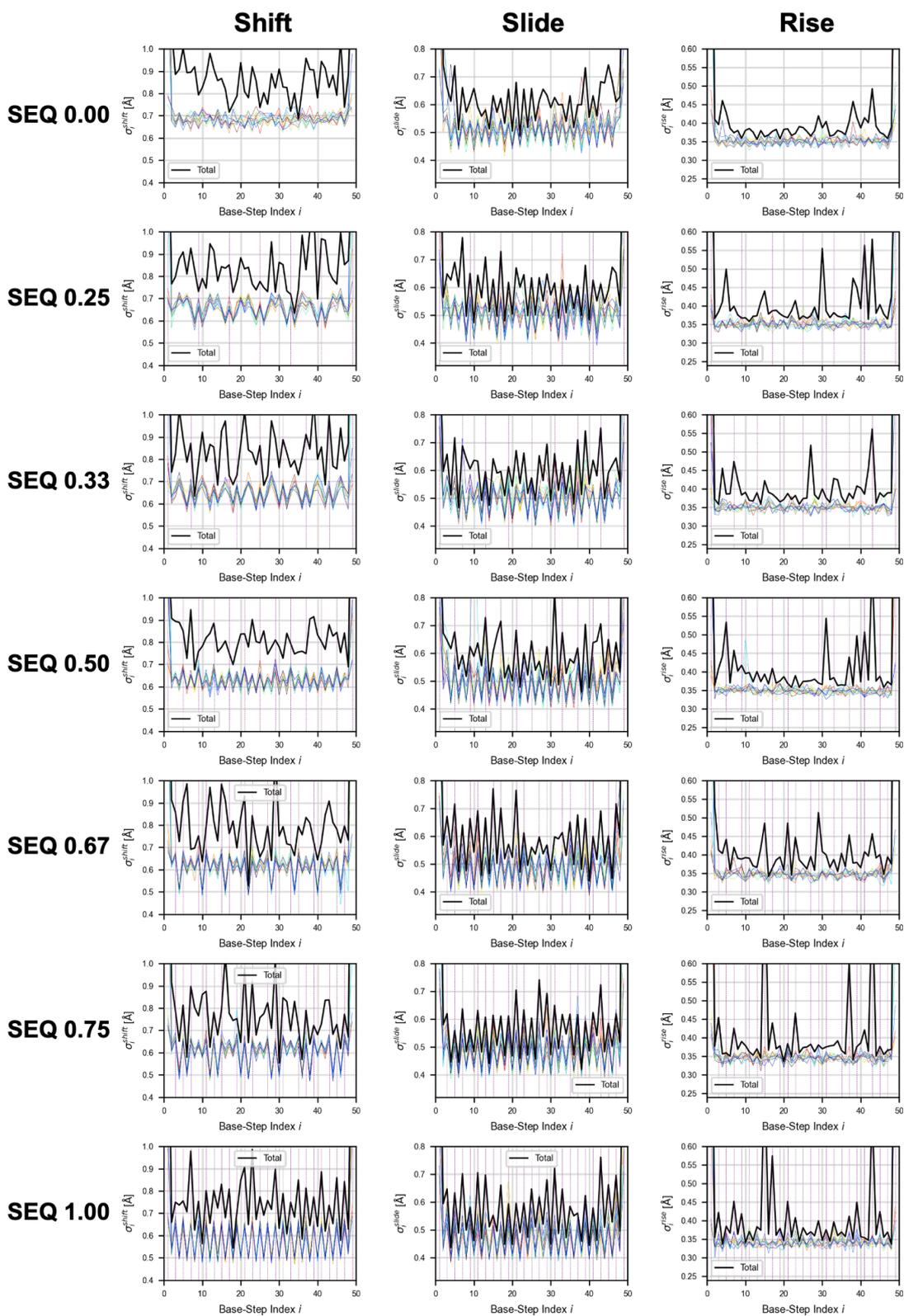
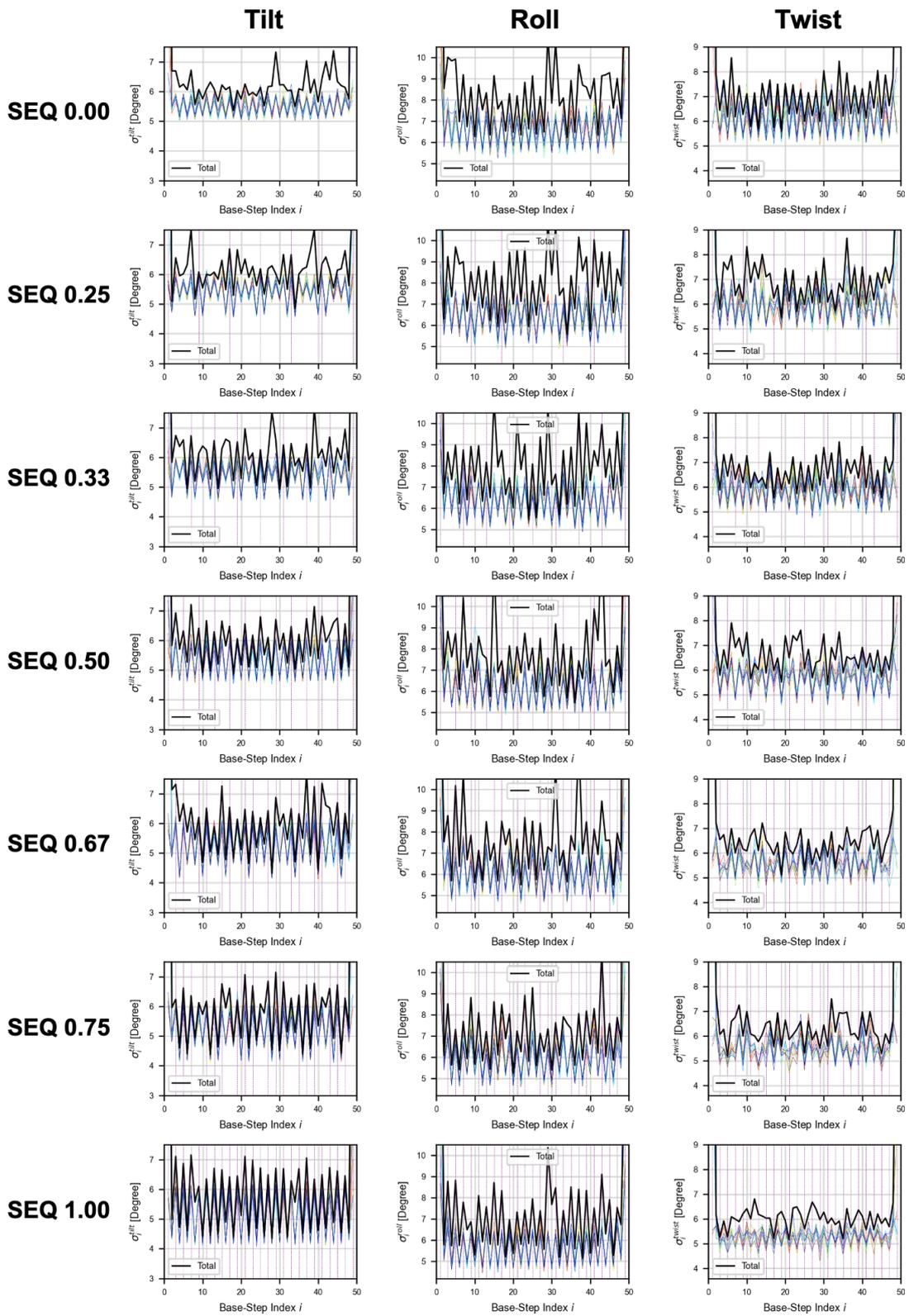
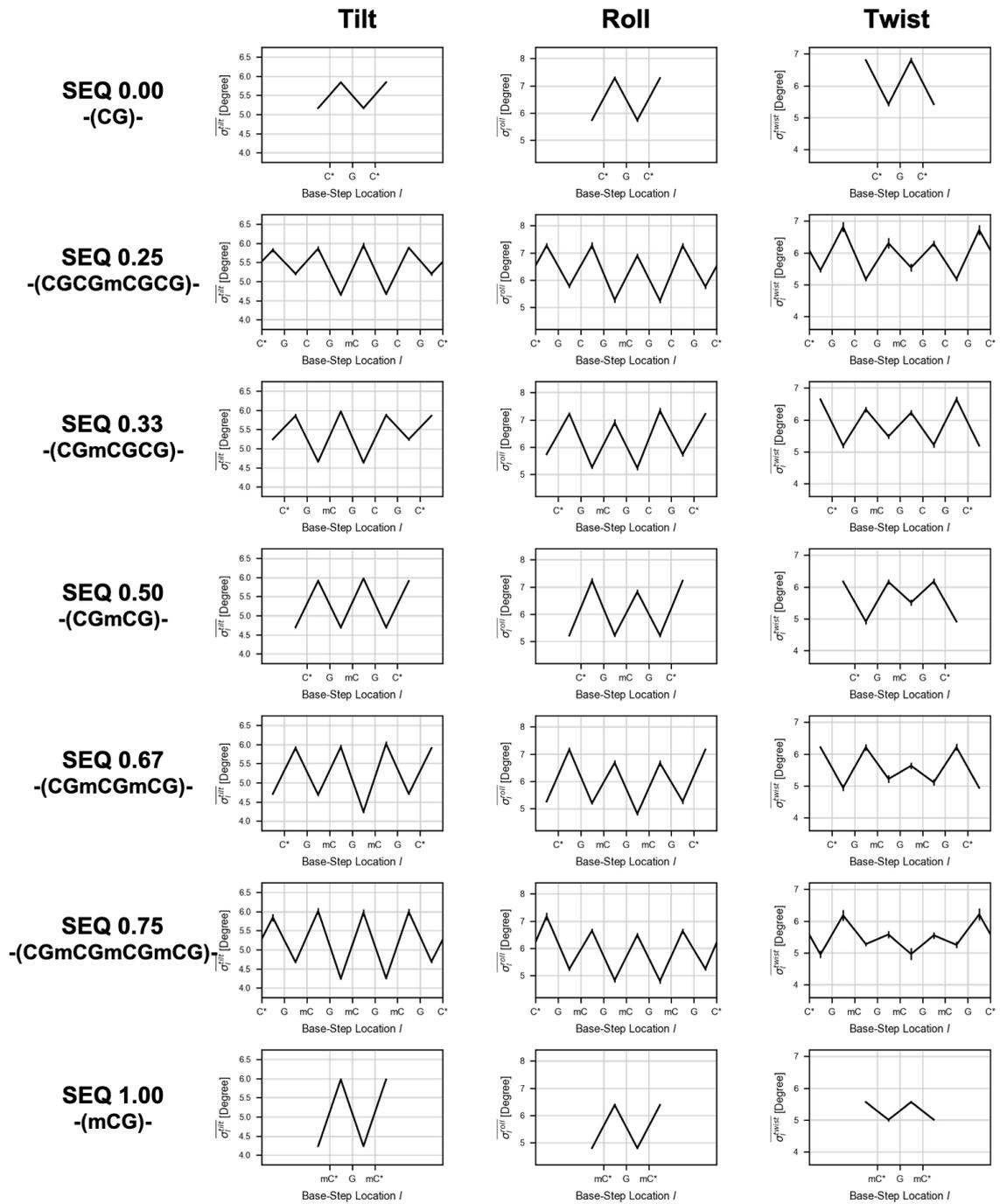


Figure S3. Profiles of σ_i^* . Colors show different simulation trials and the black line shows the overall flexibility (structural variation) in ten trials. In the horizontal axis, odd and even numbers correspond to base-steps of C→G and G→C, respectively (Fig. 1 (c)). Purple lines show mC→G (C→G in the case of mC) base-steps.





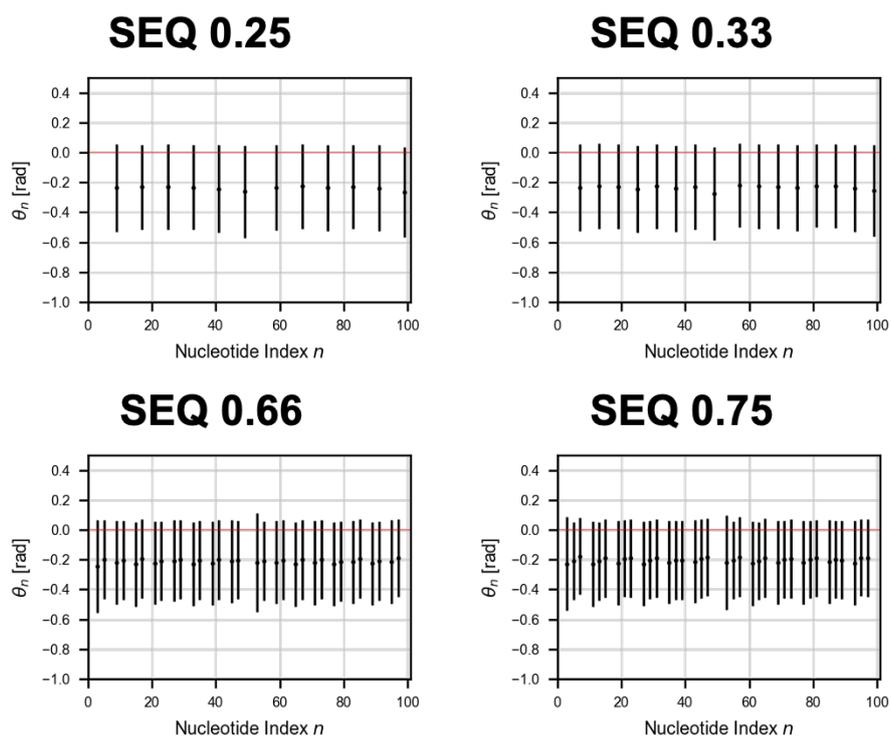


Figure S5. Orientation of Methyl Groups. The average \pm the standard deviation (S.D.) of θ_n for each methyl group are shown.

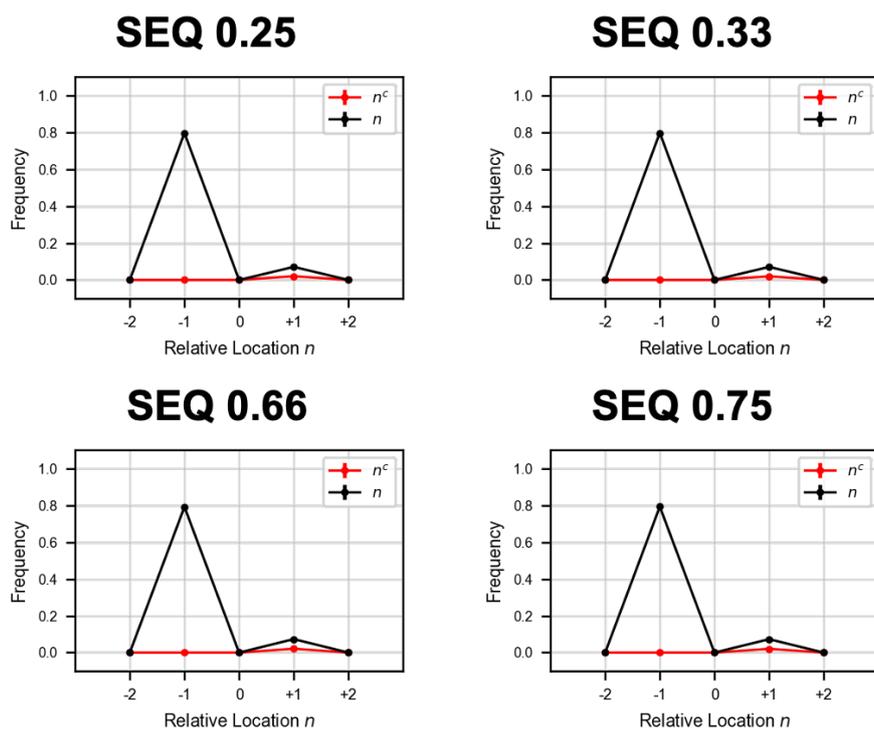


Figure S6. Interaction Between Methyl Groups and DNA. The average \pm the standard deviation (S.D.) of contact frequency for all methyl groups is shown.

7.5 Supplemental Figures of Section 3

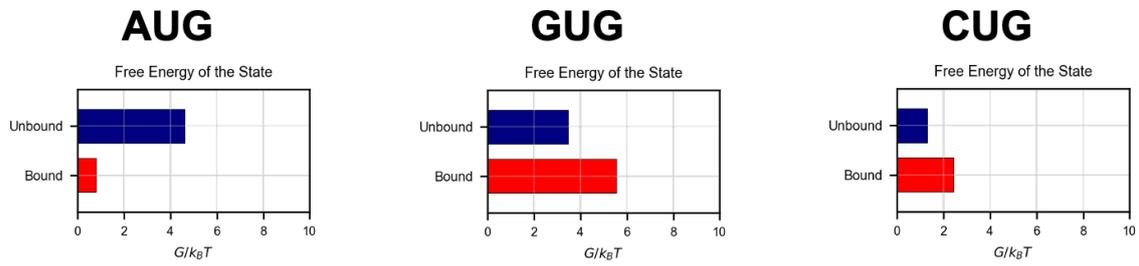


Figure S7. Estimated Free Energy of Bound and Unbound States. Red and blue bars correspond to G_{bound} and G_{unbound} (Eq. 3.5), respectively (see the schematics in Fig. 11). The scores were obtained from $P(d_1, d_2, d_3)$ and $G(d_1, d_2, d_3)$ profiles (Eq. 3.2) over five simulation trials for each model.

7.6 Supplemental Figures of Section 4

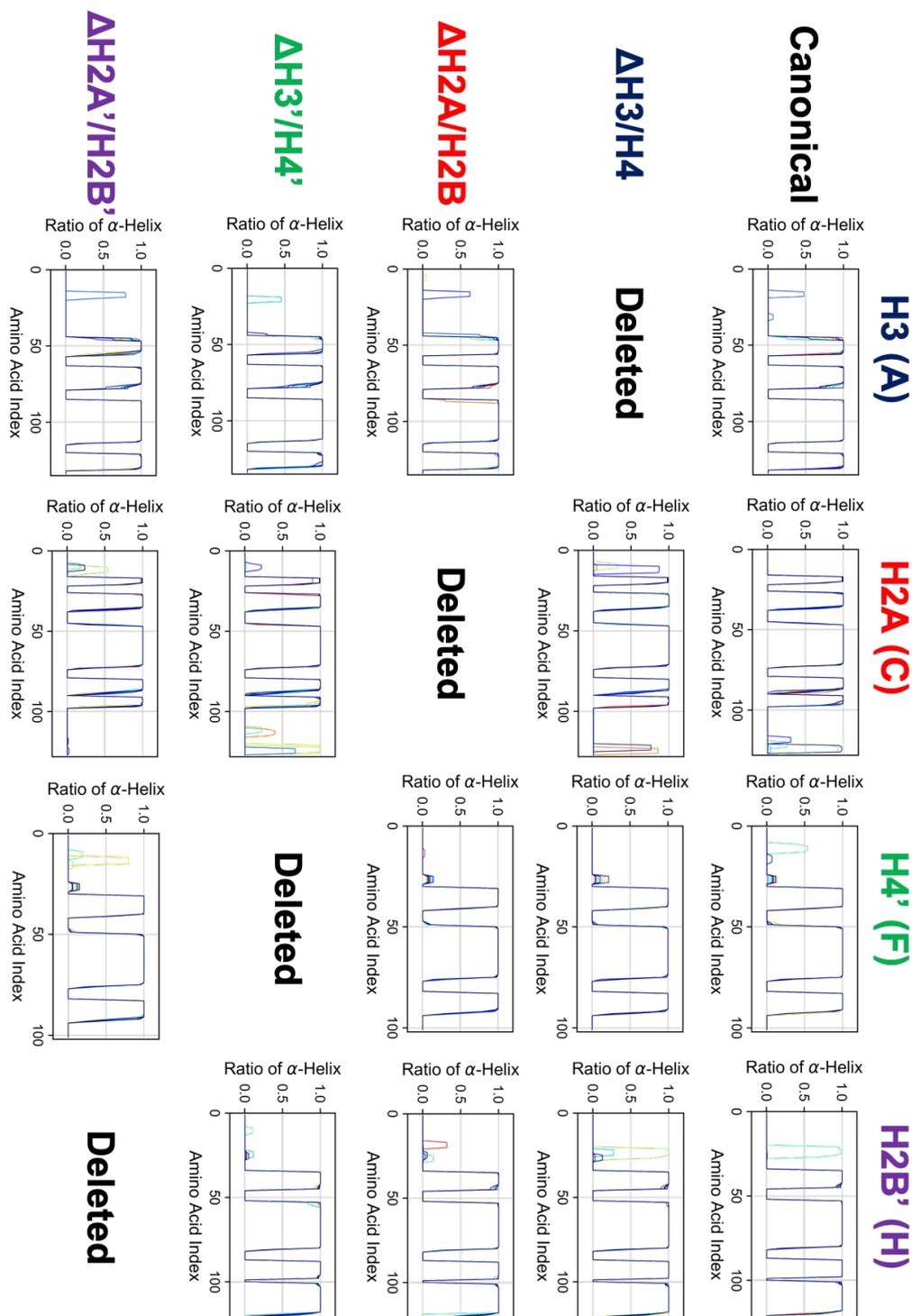
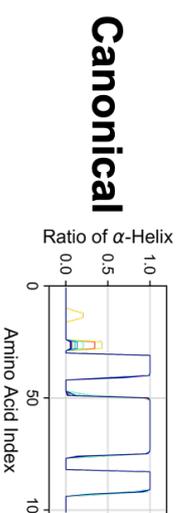
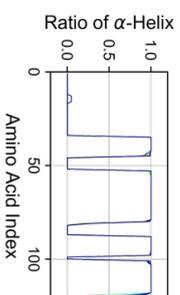


Figure S8. Secondary Structure Stability of Histones in Each Nucleosome. Horizontal and vertical axes of each graph show amino acid index and ratio of α -helix formation in each histone, respectively. Colors show different simulation trajectories (total ten).

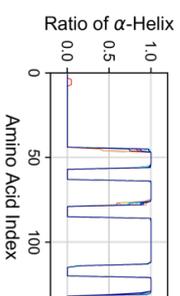
H4 (B)



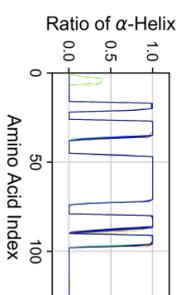
H2B (D)



H3' (E)

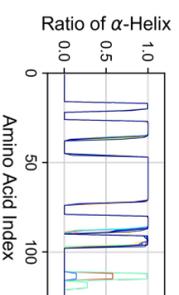
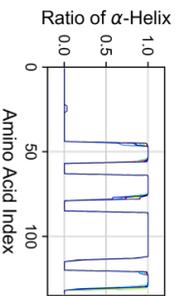
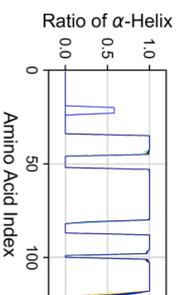


H2A' (G)



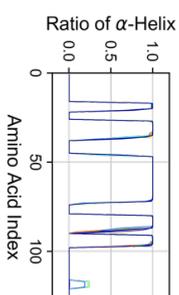
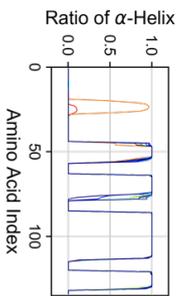
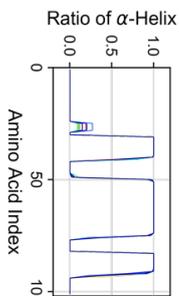
Δ H3'/H4

Deleted



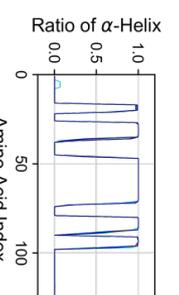
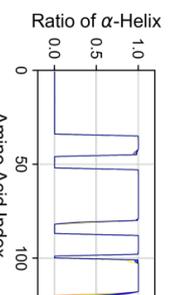
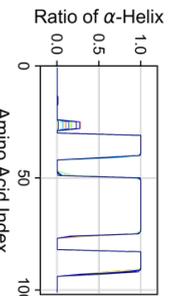
Δ H2A'/H2B'

Deleted



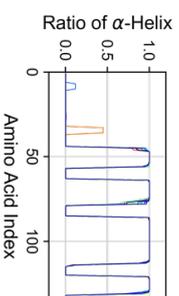
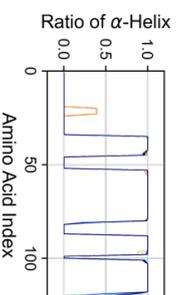
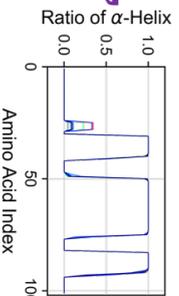
Δ H3'/H4'

Deleted



Δ H2A'/H2B'

Deleted



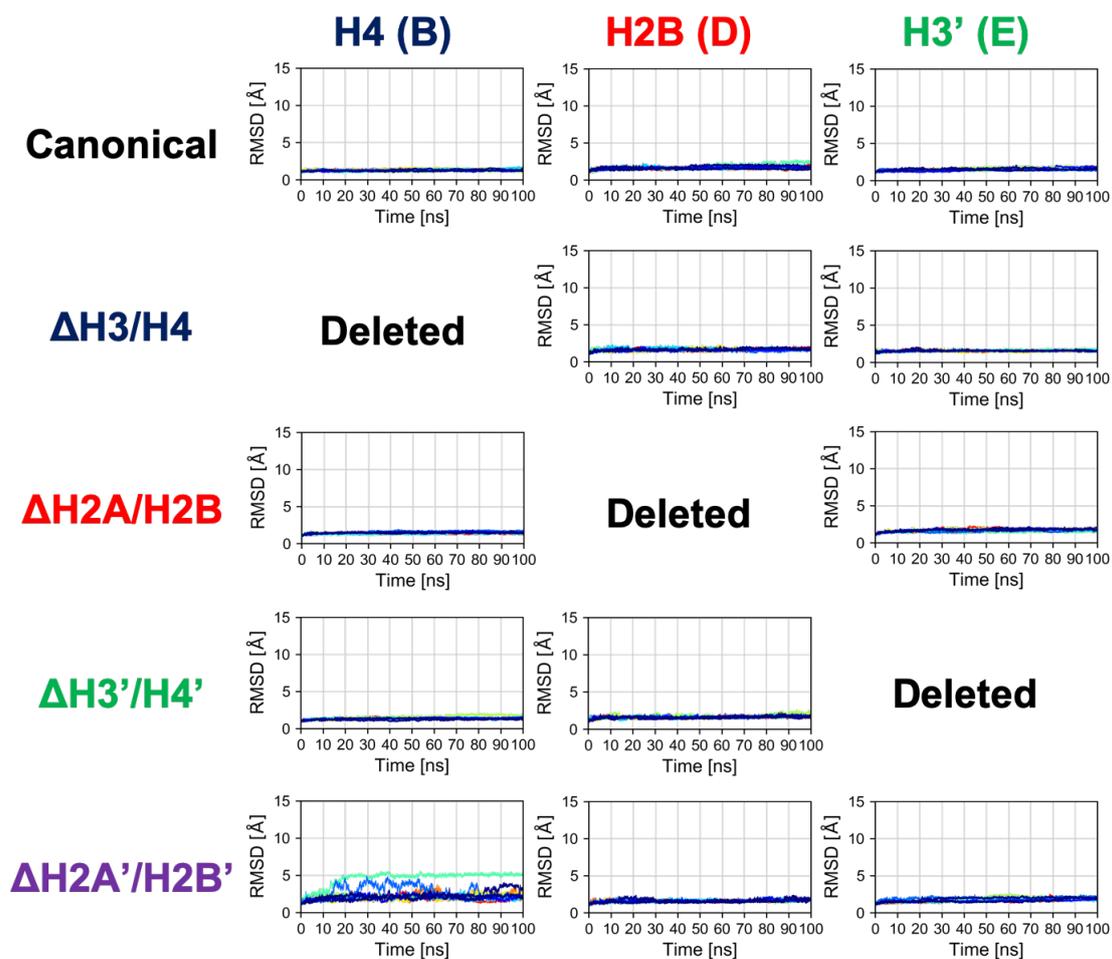
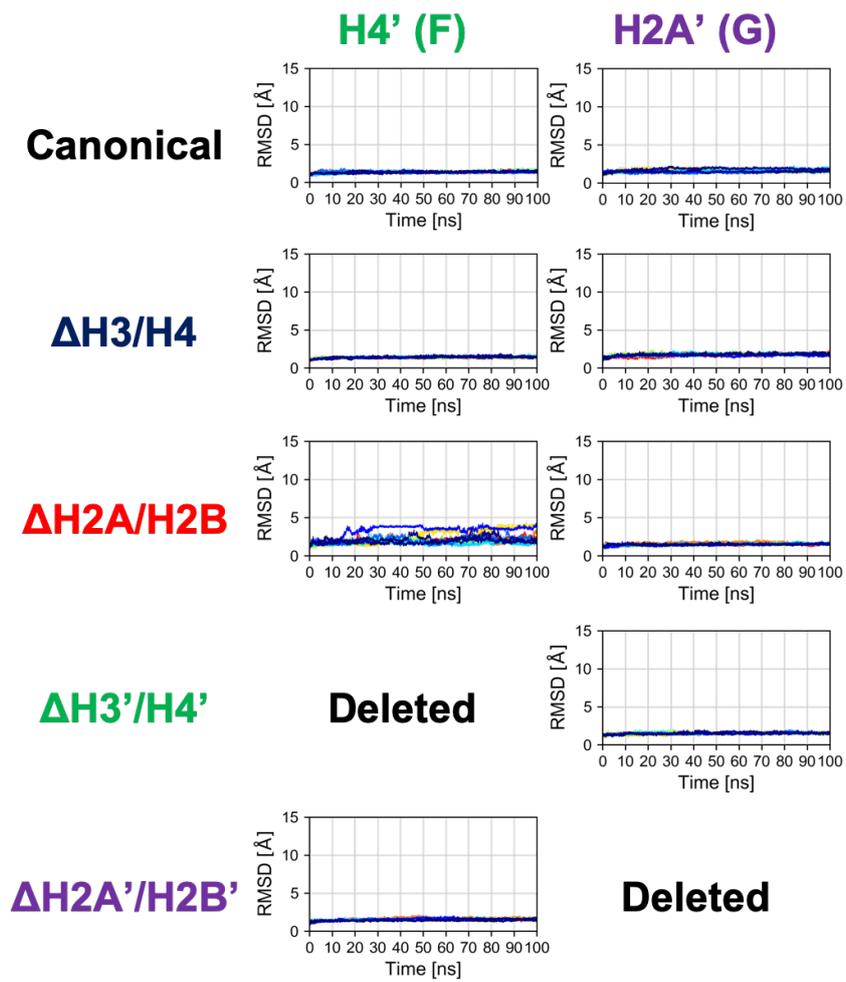


Figure S9. Time Series of RMSD of Histone in Each Nucleosome. Colors show different simulation trajectories (total ten).



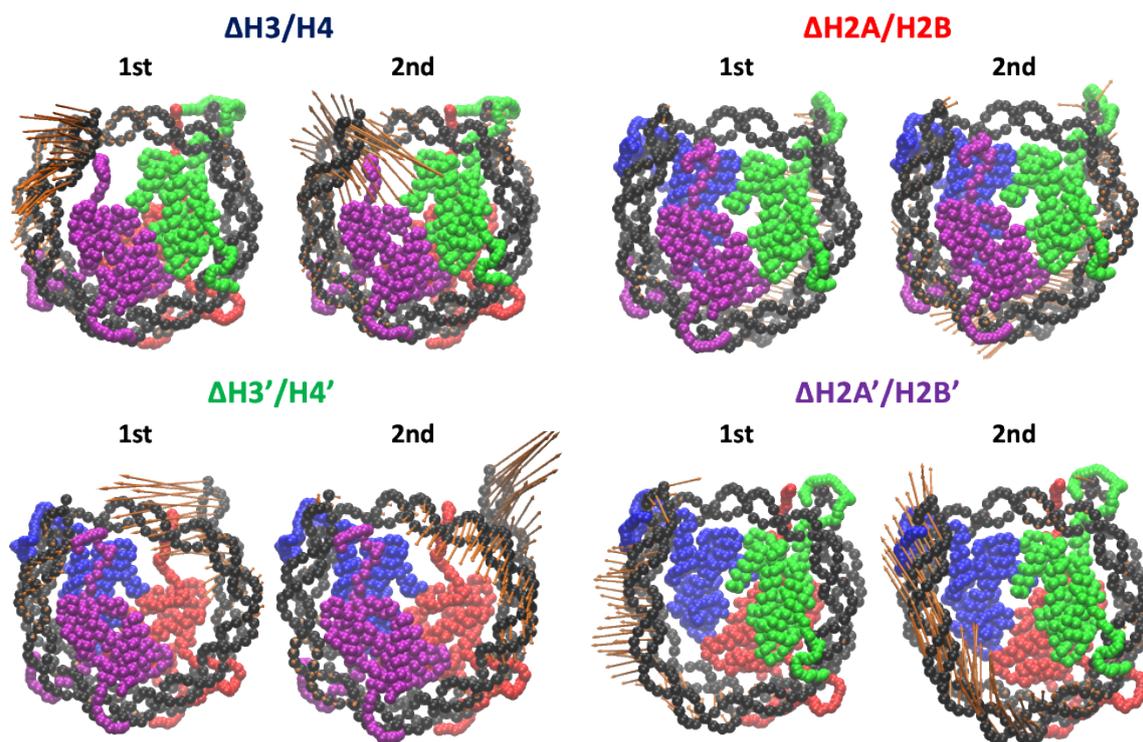


Figure S10. Effective Directions of DNA Deformation (Back View). Visualized from the side of chains E to H (H3', H4', H2A', and H2B'). Nucleotide (C1' of each nucleotide) and amino acid (C^α of each amino acid) coordinates are shown by spheres. Orange arrows show the vectors of the 1st and 2nd PC modes. Arrow lengths indicate the norm of the vector.

References

- [1] Raphael Falk. *Genetic analysis: A history of genetic thinking*. Cambridge University Press Cambridge, 2009.
- [2] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.
- [3] Stephen Cooper. The central dogma of cell biology. *Cell Biol. Int. Rep.*, 5(6):539–549, 1981.
- [4] Francis HC Crick. On protein synthesis. In *Symp Soc Exp Biol*, volume 12, page 8, 1958.
- [5] Javier De Las Rivas and Celia Fontanillo. Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell. *Brief Funct Genomics*, 11(6):489–496, 2012.
- [6] Pavel Dolezal, Vladimir Likic, Jan Tachezy, and Trevor Lithgow. Evolution of the molecular machines for protein import into mitochondria. *Science*, 313(5785):314–318, 2006.
- [7] Thomas Köcher and Giulio Superti-Furga. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods*, 4(10):807–815, 2007.
- [8] International Human Genome Sequencing Consortium et al. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931, 2004.
- [9] SDM Brown and JM Hancock. The mouse genome. In *Vertebrate Genomes*, volume 2, pages 33–45. Karger Publishers, 2006.
- [10] Feng Yang, Jun He, Xionghui Lin, Qin Li, Deng Pan, Xiaobo Zhang, and Xun Xu. Complete genome sequence of the shrimp white spot bacilliform virus. *J. Virol.*, 75(23):11811–11820, 2001.
- [11] Elisabeth A Herniou, Julie A Olszewski, Jennifer S Cory, and David R O’Reilly. The genome sequence and evolution of baculoviruses. *Annu. Rev. Entomol.*, 48(1):211–234, 2003.
- [12] Bernard Dujon. The yeast genome project: what did we learn? *Trends Genet.*, 12(7):263–270, 1996.
- [13] Daniel G Gibson, Gwynedd A Benders, Cynthia Andrews-Pfannkoch, Evgeniya A Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B Stockwell, Anushka Brownley, David W Thomas, Mikkel A Algire, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 319(5867):1215–1220, 2008.
- [14] Stephan C Schuster. Next-generation sequencing transforms today’s biology. *Nat. Methods*, 5(1):16–18, 2008.

- [15] Erwin L Van Dijk, Hélène Auger, Yan Jaszczyszyn, and Claude Thermes. Ten years of next-generation sequencing technology. *Trends Genet.*, 30(9):418–426, 2014.
- [16] Michael L Metzker. Sequencing technologies—the next generation. *Nat. Rev. Genet.*, 11(1):31–46, 2010.
- [17] Matthieu Le Gallo, Fred Lozy, and Daphne W Bell. Next-generation sequencing. In *Molecular Genetics of Endometrial Carcinoma*, pages 119–148. Springer, 2017.
- [18] Elaine R Mardis. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9:387–402, 2008.
- [19] Michael Z Ludwig, Casey Bergman, Nipam H Patel, and Martin Kreitman. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*, 403(6769):564–567, 2000.
- [20] Robin Andersson and Albin Sandelin. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.*, 21(2):71–87, 2020.
- [21] George Khoury and Peter Gruss. Enhancer elements. *Cell*, 33(2):313–314, 1983.
- [22] Abraham S Weintraub, Charles H Li, Alicia V Zamudio, Alla A Sigova, Nancy M Hannett, Daniel S Day, Brian J Abraham, Malkiel A Cohen, Behnam Nabet, Dennis L Buckley, et al. Yy1 is a structural regulator of enhancer-promoter loops. *Cell*, 171(7):1573–1588, 2017.
- [23] Adam C Bell, Adam G West, and Gary Felsenfeld. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science*, 291(5503):447–450, 2001.
- [24] Kevin Van Bortle and Victor G Corces. The role of chromatin insulators in nuclear architecture and genome function. *Curr. Opin. Genet. Dev.*, 23(2):212–218, 2013.
- [25] Dahong Chen and Elissa P Lei. Function and regulation of chromatin insulators in dynamic genome organization. *Curr. Opin. Cell Biol.*, 58:61–68, 2019.
- [26] Jingping Yang and Victor G Corces. Insulators, long-range interactions, and genome function. *Curr. Opin. Genet. Dev.*, 22(2):86–92, 2012.
- [27] Matteo Vietri Rudan, Christopher Barrington, Stephen Henderson, Christina Ernst, Duncan T Odom, Amos Tanay, and Suzana Hadjur. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.*, 10(8):1297–1309, 2015.
- [28] Kyle P Eagen. Principles of chromosome architecture revealed by Hi-C. *Trends Biochem. Sci.*, 43(6):469–478, 2018.

- [29] Chin-Tong Ong and Victor G Corces. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.*, 15(4):234–246, 2014.
- [30] Anders S Hansen. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus*, 11(1):132–148, 2020.
- [31] ENCODE Project Consortium et al. The ENCODE (ENCyclopedia of DNA elements) project. *Science*, 306(5696):636–640, 2004.
- [32] John A Stamatoyannopoulos, Michael Snyder, Ross Hardison, Bing Ren, Thomas Gingeras, David M Gilbert, Mark Groudine, Michael Bender, Rajinder Kaul, Theresa Canfield, et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.*, 13(8):1–5, 2012.
- [33] Carrie A Davis, Benjamin C Hitz, Cricket A Sloan, Esther T Chan, Jean M Davidson, Idan Gabdank, Jason A Hilton, Kriti Jain, Ulugbek K Baymurov, Aditi K Narayanan, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.*, 46(D1):D794–D801, 2018.
- [34] Chung-Chau Hon and Piero Carninci. Expanded ENCODE delivers invaluable genomic encyclopedia, 2020.
- [35] Craig J Benham and Steven P Mielke. DNA mechanics. *Annu. Rev. Biomed. Eng.*, 7:21–53, 2005.
- [36] Zev Bryant, Florian C Oberstrass, and Aakash Basu. Recent developments in single-molecule DNA mechanics. *Curr. Opin. Struct. Biol.*, 22(3):304–312, 2012.
- [37] David Levens, Laura Baranello, and Fedor Kouzine. Controlling gene expression by DNA mechanics: emerging insights and challenges. *Biophys Rev*, 8(3):259–268, 2016.
- [38] Anthony N Imbalzano, Hyockman Kwon, Michael R Green, and Robert E Kingston. Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature*, 370(6489):481–485, 1994.
- [39] John P Thomson, Peter J Skene, Jim Selfridge, Thomas Clouaire, Jacky Guy, Shaun Webb, Alastair RW Kerr, Aimée Deaton, Rob Andrews, Keith D James, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, 464(7291):1082–1086, 2010.
- [40] François Spitz and Eileen EM Furlong. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, 13(9):613–626, 2012.
- [41] Bing Ren, François Robert, John J Wyrick, Oscar Aparicio, Ezra G Jennings, Itamar Simon, Julia Zeitlinger, Jörg Schreiber, Nancy Hannett, Elenita Kanin, et al. Genome-wide location and function of DNA binding proteins. *Science*, 290(5500):2306–2309, 2000.

- [42] Pamela J Mitchell and Robert Tjian. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, 245(4916):371–378, 1989.
- [43] Wilma K Olson, Andrey A Gorin, Xiang-Jun Lu, Lynette M Hock, and Victor B Zhurkin. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, 95(19):11163–11168, 1998.
- [44] Matthias Rief, Hauke Clausen-Schaumann, and Hermann E Gaub. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.*, 6(4):346–349, 1999.
- [45] Martha L Bulyk, Erik Gentalen, David J Lockhart, and George M Church. Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nat. Biotechnol.*, 17(6):573–577, 1999.
- [46] Ronald Hancock. Structures and functions in the crowded nucleus: new biophysical insights. *Front Phys*, 2:53, 2014.
- [47] Alexey A Shvets and Anatoly B Kolomeisky. Crowding on DNA in protein search for targets. *J. Phys. Chem. Lett.*, 7(13):2502–2506, 2016.
- [48] J Andrew McCammon. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.*, 8(2):245–249, 1998.
- [49] Bernhard Rupp. *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science, 2009.
- [50] Georg E Schulz and R Heiner Schirmer. *Principles of protein structure*. Springer Science & Business Media, 2013.
- [51] Sébastien Boutet, Lukas Lomb, Garth J Williams, Thomas RM Barends, Andrew Aquila, R Bruce Doak, Uwe Weierstall, Daniel P DePonte, Jan Steinbrener, Robert L Shoeman, et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science*, 337(6092):362–364, 2012.
- [52] JCH Spence, Uwe Weierstall, and HN Chapman. X-ray lasers for structural and dynamic biology. *Rep. Prog. Phys.*, 75(10):102601, 2012.
- [53] Jose M Martin-Garcia, Chelsie E Conrad, Jesse Coe, Shatabdi Roy-Chowdhury, and Petra Fromme. Serial femtosecond crystallography: A revolution in structural biology. *Arch. Biochem. Biophys.*, 602:32–47, 2016.
- [54] John Cavanagh, Wayne J Fairbrother, Arthur G Palmer III, and Nicholas J Skelton. *Protein NMR spectroscopy: principles and practice*. Elsevier, 1995.
- [55] Timothy DW Claridge. *High-resolution NMR techniques in organic chemistry*, volume 27. Elsevier, 2016.

- [56] Rieko Ishima and Dennis A Torchia. Protein dynamics from NMR. *Nat. Struct. Biol.*, 7(9):740–743, 2000.
- [57] Maya Topf, Keren Lasker, Ben Webb, Haim Wolfson, Wah Chiu, and Andrej Sali. Protein structure fitting and refinement guided by cryo-EM density. *Structure*, 16(2):295–307, 2008.
- [58] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, 2020.
- [59] Dan Shi, Brent L Nannenga, Matthew G Iadanza, and Tamir Gonen. Three-dimensional electron crystallography of protein microcrystals. *Elife*, 2:e01345, 2013.
- [60] Ammasi Periasamy and Richard Day. *Molecular imaging: FRET microscopy and spectroscopy*. Elsevier, 2011.
- [61] Tomasz Heyduk. Measuring protein conformational changes by FRET/LRET. *Curr. Opin. Biotechnol.*, 13(4):292–296, 2002.
- [62] Igor L Medintz and Niko Hildebrandt. *FRET-Förster resonance energy transfer: from theory to applications*. John Wiley & Sons, 2013.
- [63] Eva Nogales. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods*, 13(1):24–27, 2016.
- [64] Phuong T Ho and Vijay S Reddy. Rapid increase of near atomic resolution virus capsid structures determined by cryo-electron microscopy. *J. Struct. Biol.*, 201(1):1–4, 2018.
- [65] John H Van Drie and Liang Tong. Cryo-EM as a powerful tool for drug discovery. *Bioorg. Med. Chem.*, page 127524, 2020.
- [66] Takeshi Kawabata. Gaussian-input Gaussian mixture model for representing density maps and atomic models. *J. Struct. Biol.*, 203(1):1–16, 2018.
- [67] Martin Karplus and Gregory A Petsko. Molecular dynamics simulations in biology. *Nature*, 347(6294):631–639, 1990.
- [68] Tomas Hansson, Chris Oostenbrink, and WilfredF van Gunsteren. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 12(2):190–196, 2002.
- [69] Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol*, 9(9):646–652, 2002.
- [70] Erik M Boczko and Charles L Brooks. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*, 269(5222):393–396, 1995.

- [71] Manfred J Sippl. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput. Aided Mol. Des.*, 7(4):473–501, 1993.
- [72] Mark Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2010.
- [73] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.
- [74] Daan Frenkel and Berend Smit. *Understanding molecular simulation: From algorithms to applications*, 2002.
- [75] Christopher D Snow, Houbi Nguyen, Vijay S Pande, and Martin Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, 420(6911):102–106, 2002.
- [76] Ugo Mayor, Christopher M Johnson, Valerie Daggett, and Alan R Fersht. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.*, 97(25):13518–13522, 2000.
- [77] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F Van Gunsteren, and Alan E Mark. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.*, 38(1-2):236–240, 1999.
- [78] Ignasi Buch, Toni Giorgino, and Gianni De Fabritiis. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*, 108(25):10184–10189, 2011.
- [79] Adam Hospital, Josep Ramon Goñi, Modesto Orozco, and Josep L Gelpí. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.*, 8:37, 2015.
- [80] Jacob D Durrant and J Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biol.*, 9(1):1–9, 2011.
- [81] Sergio Decherchi and Andrea Cavalli. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.*, 2020.
- [82] Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.*, 59(9):4035–4061, 2016.
- [83] Benjamin R Jagger, Sarah E Kochanek, Susanta Halder, Rommie E Amaro, and Adrian J Mulholland. Multiscale simulation approaches to modeling drug-protein binding. *Curr. Opin. Struct. Biol.*, 61:213–221, 2020.
- [84] Xuwei Liu, Danfeng Shi, Shuangyan Zhou, Hongli Liu, Huanxiang Liu, and Xiaojun Yao. Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.*, 13(1):23–37, 2018.

- [85] Massimiliano Bonomi and Carlo Camilloni. *Biomolecular Simulations*, 2019.
- [86] David J Huggins, Philip C Biggin, Marc A Dämgen, Jonathan W Essex, Sarah A Harris, Richard H Henchman, Syma Khalid, Antonija Kuzmanic, Charles A Laughton, Julien Michel, et al. *Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity*. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 9(3):e1393, 2019.
- [87] Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.
- [88] Norman L Allinger, Xuefeng Zhou, and John Bergsma. Molecular mechanics parameters. *J. Mol. Struct.*, 312(1):69–83, 1994.
- [89] Kenno Vanommeslaeghe, Olgun Guvench, et al. Molecular mechanics. *Curr. Pharm. Des.*, 20(20):3281–3292, 2014.
- [90] Sigeru Huzinaga, Jan Andzelm, Es Radzio-Andzelm, Y Sakai, H Tatewaki, and M Klobukowski. *Gaussian basis sets for molecular calculations*. Elsevier, 2012.
- [91] Thom H Dunning and P Jeffrey Hay. Gaussian basis sets for molecular calculations. In *Methods of electronic structure theory*, pages 1–27. Springer, 1977.
- [92] Thomas Steinbrecher and Marcus Elstner. QM and QM/MM simulations of proteins. In *Biomolecular Simulations*, pages 91–124. Springer, 2013.
- [93] Hans Martin Senn and Walter Thiel. QM/MM methods for biological systems. In *Atomistic approaches in modern biology*, pages 173–290. Springer, 2006.
- [94] PA Bash, UC Singh, R Langridge, and PA Kollman. Free energy calculations by computer simulation. *Science*, 236(4801):564–568, 1987.
- [95] M Germana Paterlini and David M Ferguson. Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chem. Phys.*, 236(1-3):243–252, 1998.
- [96] MBBJM Tuckerman, Bruce J Berne, and Glenn J Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97(3):1990–2001, 1992.
- [97] Giovanni Gallavotti. Ergodicity, ensembles, irreversibility in Boltzmann and beyond. *J. Stat. Phys.*, 78(5-6):1571–1589, 1995.
- [98] Jérôme Hénin and Christophe Chipot. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.*, 121(7):2904–2914, 2004.

- [99] Rafael C Bernardi, Marcelo CR Melo, and Klaus Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta*, 1850(5):872–877, 2015.
- [100] Tomas Rodinger and Régis Pomès. Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Curr. Opin. Struct. Biol.*, 15(2):164–170, 2005.
- [101] Thomas Hamelryck, John T Kent, and Anders Krogh. Sampling realistic protein conformations using local structural bias. *PLoS Comput. Biol.*, 2(9):e131, 2006.
- [102] Donald Hamelberg, John Mongan, and J Andrew McCammon. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120(24):11919–11929, 2004.
- [103] Yuko Okamoto. Protein structure predictions by enhanced conformational sampling methods. *Biophysics and Physicobiology*, 16:344–366, 2019.
- [104] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, 1999.
- [105] Marcus B Kubitzki and Bert L De Groot. Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. *Biophys. J.*, 92(12):4262–4270, 2007.
- [106] Jaegil Kim, John E Straub, and Tom Keyes. Replica exchange statistical temperature molecular dynamics algorithm. *J. Phys. Chem. B*, 116(29):8646–8653, 2012.
- [107] Ruhong Zhou. Replica exchange molecular dynamics method for protein folding simulation. In *Protein Folding Protocols*, pages 205–223. Springer, 2007.
- [108] Young Min Rhee and Vijay S Pande. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.*, 84(2):775–786, 2003.
- [109] Fan Jiang and Yun-Dong Wu. Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *J. Am. Chem. Soc.*, 136(27):9536–9539, 2014.
- [110] Hiraku Oshima, Suyong Re, and Yuji Sugita. Replica-exchange umbrella sampling combined with Gaussian accelerated molecular dynamics for free-energy calculation of biomolecules. *J. Chem. Theory Comput.*, 15(10):5199–5208, 2019.
- [111] Ayori Mitsutake and Hiroshi Takano. Relaxation mode analysis for molecular dynamics simulations of proteins. *Biophys Rev*, 10(2):375–389, 2018.

- [112] Yasuhiro Matsunaga, Hiroshi Fujisaki, Tohru Terada, Tadaomi Furuta, Kei Moritsugu, and Akinori Kidera. Minimum free energy path of ligand-induced transition in adenylate kinase. *PLoS Comput. Biol.*, 8(6):e1002555, 2012.
- [113] Jeffrey Comer, James C Gumbart, Jérôme Hénin, Tony Lelièvre, Andrew Pohorille, and Christophe Chipot. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B*, 119(3):1129–1151, 2015.
- [114] Lingle Wang, Richard A Friesner, and BJ Berne. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B*, 115(30):9431–9438, 2011.
- [115] Raquel Dias, Jr de Azevedo, and F Walter. Molecular docking algorithms. *Curr Drug Targets*, 9(12):1040–1047, 2008.
- [116] Juan Fernandez-Recio, Maxim Totrov, and Ruben Abagyan. Identification of protein–protein interaction sites from docking energy landscapes. *J. Mol. Biol.*, 335(3):843–865, 2004.
- [117] Jacqueline Cherfils, Stephane Duquerroy, and Joel Janin. Protein-protein recognition analyzed by docking simulation. *Proteins*, 11(4):271–280, 1991.
- [118] Gerard Pujadas, Montserrat Vaque, Anna Ardevol, Cinta Blade, MJ Salvado, Mayte Blay, Juan Fernandez-Larrea, and Lluís Arola. Protein-ligand docking: A review of recent advances and future perspectives. *Curr. Pharm. Anal.*, 4(1):1–19, 2008.
- [119] Venkatraman Mohan, Alan C Gibbs, Maxwell D Cummings, Edward P Jaeger, and Renee L DesJarlais. Docking: successes and challenges. *Curr. Pharm. Des.*, 11(3):323–333, 2005.
- [120] Paweł Śledź and Amedeo Caflisch. Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.*, 48:93–102, 2018.
- [121] Masahito Ohue, Yuri Matsuzaki, Takehiro Shimoda, Takashi Ishida, and Yutaka Akiyama. Highly precise protein-protein interaction prediction based on consensus between template-based and de novo docking methods. In *BMC Proc.*, volume 7, page S6. Springer, 2013.
- [122] Zhifeng Liu, Yujie Liu, Guangming Zeng, Binbin Shao, Ming Chen, Zhigang Li, Yilin Jiang, Yang Liu, Yu Zhang, and Hua Zhong. Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review. *Chemosphere*, 203:139–150, 2018.
- [123] Janaina Cruz Pereira, Ernesto Raul Caffarena, and Cicero Nogueira dos Santos. Boosting docking-based virtual screening with deep learning. *J Chem Inf Model*, 56(12):2495–2506, 2016.

- [124] Mohamed A Khamis, Walid Gomaa, and Walaa F Ahmed. Machine learning in computational docking. *Artif Intell Med*, 63(3):135–152, 2015.
- [125] Mauro S Nogueira and Oliver Koch. The development of target-specific machine learning models as scoring functions for docking-based target prediction. *J. Chem. Inf. Model.*, 59(3):1238–1252, 2019.
- [126] Tanay Chandak, John P Mayginnes, Howard Mayes, and Chung F Wong. Using machine learning to improve ensemble docking for drug discovery. *Proteins*, 88(10):1263–1270, 2020.
- [127] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The Amber biomolecular simulation programs. *J Comput Chem*, 26(16):1668–1688, 2005.
- [128] Romelia Salomon-Ferrer, David A Case, and Ross C Walker. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 3(2):198–210, 2013.
- [129] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S a Swaminathan, and Martin Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.
- [130] Bernard R Brooks, Charles L Brooks III, Alexander D Mackerell Jr, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, 30(10):1545–1614, 2009.
- [131] Herman JC Berendsen, David van der Spoel, and Rudi van Drunen. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun*, 91(1-3):43–56, 1995.
- [132] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.
- [133] Jaewoon Jung, Takaharu Mori, Chigusa Kobayashi, Yasuhiro Matsunaga, Takao Yoda, Michael Feig, and Yuji Sugita. GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 5(4):310–323, 2015.
- [134] Jaewoon Jung, Akira Nourse, Chigusa Kobayashi, and Yuji Sugita. Graphics processing unit acceleration and parallelization of GENESIS for large-scale molecular dynamics simulations. *J. Chem. Theory Comput.*, 12(10):4947–4958, 2016.

- [135] Yoshimichi Andoh, Noriyuki Yoshii, Kazushi Fujimoto, Keisuke Mizutani, Hidekazu Kojima, Atsushi Yamada, Susumu Okazaki, Kazutomo Kawaguchi, Hidemi Nagao, Kensuke Iwahashi, et al. MODYLAS: A highly parallelized general-purpose molecular dynamics simulation program for large-scale systems with long-range forces calculated by fast multipole method (FMM) and highly scalable fine-grained new parallel processing algorithms. *J. Chem. Theory Comput.*, 9(7):3201–3209, 2013.
- [136] Noriyuki Yoshii, Yoshimichi Andoh, Kazushi Fujimoto, Hidekazu Kojima, Atsushi Yamada, and Susumu Okazaki. MODYLAS: A highly parallelized general-purpose molecular dynamics simulation program. *Int. J. Quant. Chem.*, 115(5):342–348, 2015.
- [137] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.
- [138] James C Phillips, David J Hardy, Julio DC Maia, John E Stone, João V Ribeiro, Rafael C Bernardi, Ronak Buch, Giacomo Fiorin, Jérôme Hémin, Wei Jiang, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.*, 153(4):044130, 2020.
- [139] Scott J Weiner, Peter A Kollman, David A Case, U Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106(3):765–784, 1984.
- [140] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *J Comput Chem*, 25(9):1157–1174, 2004.
- [141] Yushan Zhang, Yong Zhang, Mark J McCready, and Edward J Maginn. Evaluation and refinement of the general AMBER Force Field for nineteen pure organic electrolyte solvents. *J. Chem. Eng. Data*, 63(9):3488–3502, 2018.
- [142] Katarina Hart, Nicolas Foloppe, Christopher M Baker, Elizabeth J Denning, Lennart Nilsson, and Alexander D MacKerell Jr. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.*, 8(1):348–362, 2012.
- [143] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro EM Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.*, 8(9):3257–3273, 2012.
- [144] Walter RP Scott, Philippe H Hünenberger, Ilario G Tironi, Alan E Mark, Salomon R Billeter, Jens Fennen, Andrew E Torda, Thomas Huber, Peter

- Krüger, and Wilfred F van Gunsteren. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, 103(19):3596–3607, 1999.
- [145] Devleena Shivakumar, Edward Harder, Wolfgang Damm, Richard A Friesner, and Woody Sherman. Improving the prediction of absolute solvation free energies using the next generation OPLS force field. *J. Chem. Theory Comput.*, 8(8):2553–2558, 2012.
- [146] William George Noid. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.*, 139(9):09B201_1, 2013.
- [147] Sergei Izvekov and Gregory A Voth. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, 109(7):2469–2473, 2005.
- [148] Valentina Tozzini. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, 15(2):144–150, 2005.
- [149] Marissa G Saunders and Gregory A Voth. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.*, 42:73–93, 2013.
- [150] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007.
- [151] Shoji Takada. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.*, 22(2):130–137, 2012.
- [152] Shoji Takada, Ryo Kanada, Cheng Tan, Tsuyoshi Terakawa, Wenfei Li, and Hiroo Kenzaki. Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc. Chem. Res.*, 48(12):3026–3035, 2015.
- [153] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chem. Rev.*, 116(14):7898–7936, 2016.
- [154] Sereina Riniker, Jane R Allison, and Wilfred F van Gunsteren. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.*, 14(36):12423–12430, 2012.
- [155] Thomas E Ouldridge, Ard A Louis, and Jonathan PK Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, 134(8):02B627, 2011.
- [156] Gordon S Freeman, Daniel M Hinckley, Joshua P Lequieu, Jonathan K Whitmer, and Juan J de Pablo. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.*, 141(16):165103, 2014.
- [157] Shuhei Isami, Naoaki Sakamoto, Hiraku Nishimori, and Akinori Awazu. Simple elastic network models for exhaustive analysis of long double-stranded DNA dynamics with sequence geometry dependence. *PLoS One*, 10(12):e0143760, 2015.

- [158] Takeru Kameda, Shuhei Isami, Yuichi Togashi, Hiraku Nishimori, Naoaki Sakamoto, and Akinori Awazu. The 1-particle-per-k-nucleotides (1PkN) elastic network model of DNA dynamics with sequence-dependent geometry. *Front. Physiol.*, 8:103, 2017.
- [159] Holger Flechsig and Yuichi Togashi. Designed elastic networks: Models of complex protein machinery. *Int. J. Mol. Sci.*, 19(10):3152, 2018.
- [160] Romain Amyot, Yuichi Togashi, and Holger Flechsig. Analyzing Fluctuation Properties in Protein Elastic Networks with Sequence-Specific and Distance-Dependent Interactions. *Biomolecules*, 9(10):549, 2019.
- [161] Li L Duan, Guo Q Feng, and Qing G Zhang. Large-scale molecular dynamics simulation: Effect of polarization on thrombin-ligand binding energy. *Sci. Rep.*, 6(1):1–11, 2016.
- [162] Michael Feig and Yuji Sugita. Whole-cell models and simulations in molecular detail. *Annu. Rev. Cell Dev. Biol.*, 35:191–211, 2019.
- [163] Hisashi Ishida and Hidetoshi Kono. H4 tails potentially produce the diversity in the orientation of two nucleosomes. *Biophys. J.*, 113(5):978–990, 2017.
- [164] Atsushi Matsumoto, Masaaki Sugiyama, Zhenhai Li, Anne Martel, Lionel Porcar, Rintaro Inoue, Daiki Kato, Akihisa Osakabe, Hitoshi Kurumizaka, and Hidetoshi Kono. Structural studies of overlapping dinucleosomes in solution. *Biophys. J.*, 118(9):2209–2219, 2020.
- [165] Takefumi Yamashita. Toward rational antibody design: recent advancements in molecular dynamics simulations. *Int. Immunol.*, 30(4):133–140, 2018.
- [166] Thomas E Cheatham III and Matthew A Young. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, 56(4):232–256, 2000.
- [167] Dennis W Grogan. Hyperthermophiles and the problem of DNA instability. *Mol. Microbiol.*, 28(6):1043–1049, 1998.
- [168] Richard J Lipton. DNA solution of hard computational problems. *Science*, 268(5210):542–545, 1995.
- [169] James D Watson and Francis HC Crick. The structure of DNA. In *Cold Spring Harbor symposia on quantitative biology*, volume 18, pages 123–131. Cold Spring Harbor Laboratory Press, 1953.
- [170] Richard R Sinden. *DNA structure and function*. Gulf Professional Publishing, 1994.
- [171] Thomas E Creighton. Protein folding. *Biochem. J.*, 270(1):1–16, 1990.

- [172] Ken A Dill and Justin L MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110):1042–1046, 2012.
- [173] Ning Ma and Arjan van der Vaart. Anisotropy of B-DNA groove bending. *J. Am. Chem. Soc.*, 138(31):9951–9958, 2016.
- [174] AA Travers. The structural basis of DNA flexibility. *Phil. Trans. R. Soc. Lond. A*, 362(1820):1423–1438, 2004.
- [175] Davood Norouzi, Farshid Mohammad-Rafiee, and Ramin Golestanian. Effect of bending anisotropy on the 3D conformation of short DNA loops. *Phys. Rev. Lett.*, 101(16):168103, 2008.
- [176] Victor B Zhurkin, Yury P Lysov, and Valery I Ivanov. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.*, 6(3):1081–1096, 1979.
- [177] Satoshi Fujii, Hidetoshi Kono, Shigeori Takenaka, Nobuhiro Go, and Akinori Sarai. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.*, 35(18):6063–6074, 2007.
- [178] Yoshiteru Yonetani and Hidetoshi Kono. Sequence dependencies of DNA deformability and hydration in the minor groove. *Biophys. J.*, 97(4):1138–1147, 2009.
- [179] Peter L Privalov, Anatoly I Dragan, and Colyn Crane-Robinson. Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. *Nucleic Acids Res.*, 39(7):2483–2491, 2011.
- [180] Diane Lejeune, Nicolas Delsaux, Benoît Charloteaux, Annick Thomas, and Robert Brasseur. Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. *Proteins*, 61(2):258–271, 2005.
- [181] Carl Schildkraut and Shneior Lifson. Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*, 3(2):195–208, 1965.
- [182] Kirk M Ririe, Randy P Rasmussen, and Carl T Wittwer. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.*, 245(2):154–160, 1997.
- [183] Sukanya Halder and Dhananjay Bhattacharyya. RNA structure and dynamics: a base pairing perspective. *Prog. Biophys. Mol. Biol.*, 113(2):264–283, 2013.
- [184] Michiaki Hamada. Fighting against uncertainty: an essential issue in bioinformatics. *Brief. Bioinformatics*, 15(5):748–767, 2014.
- [185] Jérôme Waldispühl, Srinivas Devadas, Bonnie Berger, and Peter Clote. Efficient algorithms for probing the RNA mutation landscape. *PLoS Comput. Biol.*, 4(8):e1000124, 2008.

- [186] Thomas Derrien, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, David Martin, Angelika Merkel, David G Knowles, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.*, 22(9):1775–1789, 2012.
- [187] Mathew W Wright. A short guide to long non-coding RNA gene nomenclature. *Hum. Genomics*, 8(1):1–4, 2014.
- [188] Giulia Palermo, Lorenzo Casalino, Alessandra Magistrato, and J Andrew McCammon. Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *J. Struct. Biol.*, 206(3):267–279, 2019.
- [189] Irina V Novikova, Scott P Hennesly, Chang-Shung Tung, and Karissa Y Sanbonmatsu. Rise of the RNA machines: exploring the structure of long non-coding RNAs. *J. Mol. Biol.*, 425(19):3731–3746, 2013.
- [190] Jochen Erler, Ruihan Zhang, Loukas Petridis, Xiaolin Cheng, Jeremy C Smith, and Jörg Langowski. The role of histone tails in the nucleosome: a computational study. *Biophys. J.*, 107(12):2911–2922, 2014.
- [191] Mithun Biswas, Karine Voltz, Jeremy C Smith, and Jörg Langowski. Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.*, 7(12):e1002279, 2011.
- [192] Shantanu Sharma, Feng Ding, and Nikolay V Dokholyan. Multiscale modeling of nucleosome dynamics. *Biophys. J.*, 92(5):1457–1470, 2007.
- [193] Mithun Biswas, Jörg Langowski, and Thomas C Bishop. Atomistic simulations of nucleosomes. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 3(4):378–392, 2013.
- [194] Hidetoshi Kono and Hisashi Ishida. Nucleosome unwrapping and unstacking. *Curr. Opin. Struct. Biol.*, 64:119–125, 2020.
- [195] Artemi Bendandi, Alessandro S Patelli, Alberto Diaspro, and Walter Rocchia. The role of histone tails in nucleosome stability: An electrostatic perspective. *Comput. Struct. Biotechnol. J.*, 18:2799–2809, 2020.
- [196] Zhenhai Li and Hidetoshi Kono. Distinct roles of histone H3 and H2A tails in nucleosome stability. *Sci. Rep.*, 6:31437, 2016.
- [197] Hidetoshi Kono, Kazuyoshi Shirayama, Yasuhiro Arimura, Hiroaki Tachikawa, and Hitoshi Kurumizaka. Two arginine residues suppress the flexibility of nucleosomal DNA in the canonical nucleosome core. *PLoS One*, 10(3):e0120635, 2015.
- [198] Jinzen Ikebe, Shun Sakuraba, and Hidetoshi Kono. H3 histone tail conformation within the nucleosome and the impact of K14 acetylation studied

- using enhanced sampling simulation. *PLoS Comput. Biol.*, 12(3):e1004788, 2016.
- [199] Zhenhai Li and Hidetoshi Kono. Investigating the influence of arginine dimethylation on nucleosome dynamics using all-atom simulations and kinetic analysis. *J. Phys. Chem. B*, 122(42):9625–9634, 2018.
- [200] Hidetoshi Kono, Shun Sakuraba, and Hisashi Ishida. Free energy profiles for unwrapping the outer superhelical turn of nucleosomal DNA. *PLoS Comput. Biol.*, 14(3):e1006024, 2018.
- [201] Hisashi Ishida and Hidetoshi Kono. Torsional stress can regulate the unwrapping of two outer half superhelical turns of nucleosomal DNA. *Proc. Natl. Acad. Sci. USA*, 118(7), 2021.
- [202] Takeru Kameda, Miho M. Suzuki, Akinori Awazu, and Yuichi Togashi. Structural dynamics of DNA depending on methylation pattern. *Phys. Rev. E*, 103:012404, 2021.
- [203] Takeru Kameda, Katsura Asano, and Yuichi Togashi. Free energy landscape of RNA binding dynamics in start codon recognition by eukaryotic ribosomal pre-initiation complex. *bioRxiv*, 2021.
- [204] Takeru Kameda, Akinori Awazu, and Yuichi Togashi. Histone tail dynamics in partially disassembled nucleosomes during chromatin remodeling. *Front. Mol. Biosci.*, 6:133, 2019.
- [205] Takeru Kameda, Akinori Awazu, and Yuichi Togashi. Molecular dynamics analysis of partially disassembled nucleosomes. *Biophysics*, 60(5):288–290, 2020.
- [206] Izuru Ohki, Nobuya Shimotake, Naoyuki Fujita, Jun-Goo Jee, Takahisa Ikegami, Mitsuyoshi Nakao, and Masahiro Shirakawa. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, 105(4):487–497, 2001.
- [207] Xueqing Zou, Wen Ma, Ilia A Solov’Yov, Christophe Chipot, and Klaus Schulten. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res.*, 40(6):2747–2758, 2012.
- [208] Ayako Furukawa, Erik Walinda, Kyohei Arita, and Kenji Sugase. Structural dynamics of double-stranded DNA with epigenome modification. *Nucleic Acids Res.*, 2020.
- [209] David Neil Cooper. Eukaryotic DNA methylation. *Human genetics*, 64(4):315–333, 1983.
- [210] Adrian Bird. The essentials of DNA methylation. *Cell*, 70(1):5–8, 1992.
- [211] Zachary D Smith and Alexander Meissner. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, 14(3):204–220, 2013.

- [212] Peter A Jones. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, 13(7):484–492, 2012.
- [213] Peter A Jones and Daiya Takai. The role of DNA methylation in mammalian epigenetics. *Science*, 293(5532):1068–1070, 2001.
- [214] Melanie Ehrlich. Expression of various genes is controlled by DNA methylation during mammalian development. *J. Cell. Biochem.*, 88(5):899–910, 2003.
- [215] Robert J Klose and Adrian P Bird. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, 31(2):89–97, 2006.
- [216] Miho M Suzuki and Adrian Bird. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, 9(6):465–476, 2008.
- [217] Thuy TM Ngo, Jehoong Yoo, Qing Dai, Qiucen Zhang, Chuan He, Aleksei Aksimentiev, and Taekjip Ha. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.*, 7(1):1–9, 2016.
- [218] Philip MD Severin, Xueqing Zou, Klaus Schulten, and Hermann E Gaub. Effects of cytosine hydroxymethylation on DNA strand separation. *Biophys. J.*, 104(1):208–215, 2013.
- [219] Jehoong Yoo, Hajin Kim, Aleksei Aksimentiev, and Taekjip Ha. Direct evidence for sequence-dependent attraction between double-stranded DNA controlled by methylation. *Nat. Commun.*, 7:11045, 2016.
- [220] Isabel Jimenez-Useche, Jiaying Ke, Yuqing Tian, Daphne Shim, Steven C Howell, Xiangyun Qiu, and Chongli Yuan. DNA methylation regulated nucleosome dynamics. *Sci. Rep.*, 3(1):1–5, 2013.
- [221] Alexandra TP Carvalho, Leonor Gouveia, Charan Raju Kanna, Sebastian KTS Wärmländer, Jamie A Platts, and Shina Caroline Lynn Kamerlin. Understanding the structural and dynamic consequences of DNA epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation. *Epigenetics*, 9(12):1604–1612, 2014.
- [222] Alberto Pérez, Chiara Lara Castellazzi, Federica Battistini, Kathryn Collinet, Oscar Flores, Ozgen Deniz, Maria Luz Ruiz, David Torrents, Ramon Eritja, Montserrat Soler-López, et al. Impact of methylation on the physical properties of DNA. *Biophys. J.*, 102(9):2140–2148, 2012.
- [223] Cécilia Hognon, Vanessa Besancenot, Arnaud Gruez, Stephanie Grandemange, and Antonio Monari. Cooperative effects of cytosine methylation on DNA structure and dynamics. *J. Phys. Chem. B*, 123(34):7365–7371, 2019.
- [224] Johanna Hörberg and Anna Reymer. A sequence environment modulates the impact of methylation on the torsional rigidity of DNA. *Chem. Commun.*, 54(84):11885–11888, 2018.

- [225] Xiaojing Teng and Wonmuk Hwang. Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.*, 114(8):1791–1803, 2018.
- [226] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–38, 1996.
- [227] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh Ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.
- [228] Henrik G Petersen. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.*, 103(9):3668–3679, 1995.
- [229] Oded Farago. Langevin thermostat for robust configurational and kinetic sampling. *Physica A: Statistical Mechanics and its Applications*, 534:122210, 2019.
- [230] D Quigley and MIJ Probert. Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.*, 120(24):11432–11441, 2004.
- [231] Paul J Flory and M Volkenstein. Statistical mechanics of chain molecules. *Biopolymers*, 8(5):699–700, 1969.
- [232] Xiang-Jun Lu and Wilma K Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, 31(17):5108–5121, 2003.
- [233] Xiang-Jun Lu and Wilma K Olson. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, 3(7):1213, 2008.
- [234] Peter Gross, Niels Laurens, Lene B Oddershede, Ulrich Bockelmann, Erwin JG Peterman, and Gijs JL Wuite. Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.*, 7(9):731–736, 2011.
- [235] Jan Bednar, Patrick Furrer, Vsevolod Katritch, Alicja Stasiak, Jacques Dubochet, and Andrzej Stasiak. Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.*, 254(4):579–594, 1995.
- [236] Gerald S Manning. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys. J.*, 91(10):3607–3616, 2006.
- [237] Csaba I Pongor, Pasquale Bianco, György Ferenczy, Richárd Kellermayer, and Miklós Kellermayer. Optical trapping nanometry of hypermethylated CpG-island DNA. *Biophys. J.*, 112(3):512–522, 2017.

- [238] Ramakrishna K Chodavarapu, Suhua Feng, Yana V Bernatavichute, Pao-Yang Chen, Hume Stroud, Yanchun Yu, Jonathan A Hetzel, Frank Kuo, Jin Kim, Shawn J Cokus, et al. Relationship between nucleosome positioning and DNA methylation. *Nature*, 466(7304):388–392, 2010.
- [239] Curt A Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.*, 319(5):1097–1113, 2002.
- [240] Petr Šulc, Flavio Romano, Thomas E Ouldridge, Lorenzo Rovigatti, Jonathan PK Doye, and Ard A Louis. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.*, 137(13):135101, 2012.
- [241] K Asano. Translational control. *Encyclopedia of systems biology*. Springer, New York, NY, pages 2278–2282, 2013.
- [242] K Asano and K Ito. Translational elongation. *Encyclopedia of systems biology*. Springer, New York, NY, pages 2259–2263, 2013.
- [243] T Martin Schmeing and Venki Ramakrishnan. What recent ribosome structures have revealed about the mechanism of translation. *Nature*, 461(7268):1234–1242, 2009.
- [244] Nick Goldman and Ziheng Yang. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.*, 11(5):725–736, 1994.
- [245] MARILYN KozAK. Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc. Natl. Acad. Sci. USA*, 92(7):2662–2666, 1995.
- [246] Richard J Jackson, Christopher UT Hellen, and Tatyana V Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, 11(2):113–127, 2010.
- [247] Katsura Asano, Jason Clayton, Anath Shalev, and Alan G Hinnebusch. A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA^{Met} is an important translation initiation intermediate in vivo. *Genes Dev.*, 14(19):2534–2546, 2000.
- [248] Chingakham Ranjit Singh, Bumjun Lee, Tsuyoshi Udagawa, Sarah S Mohammad-Qureshi, Yasufumi Yamamoto, Graham D Pavitt, and Katsura Asano. An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation. *EMBO J.*, 25(19):4537–4546, 2006.
- [249] Mikkel A Algire, David Maag, and Jon R Lorsch. Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell*, 20(2):251–262, 2005.

- [250] Adam Ben-Shem, Nicolas Garreau de Loubresse, Sergey Melnikov, Lasse Jenner, Gulnara Yusupova, and Marat Yusupov. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, 334(6062):1524–1529, 2011.
- [251] Venki Ramakrishnan. Ribosome structure and the mechanism of translation. *Cell*, 108(4):557–572, 2002.
- [252] Heena Khatter, Alexander G Myasnikov, S Kundhavi Natchiar, and Bruno P Klaholz. Structure of the human 80S ribosome. *Nature*, 520(7549):640–645, 2015.
- [253] Alan G Hinnebusch, Thomas E Dever, and Katsura Asano. Mechanism of translation initiation in the yeast *Saccharomyces cerevisiae*. *COLD SPRING HARBOR MONOGRAPH SERIES*, 48:225, 2007.
- [254] Nahum Sonenberg and Alan G Hinnebusch. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4):731–745, 2009.
- [255] T Meinnel, C Sacerdot, M Graffe, S Blanquet, and M Springer. Discrimination by *Escherichia coli* initiation factor IF3 against initiation on non-canonical codons relies on complementarity rules. *J. Mol. Biol.*, 290(4):825–837, 1999.
- [256] Nicholas T Ingolia, Liana F Lareau, and Jonathan S Weissman. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4):789–802, 2011.
- [257] Katsura Asano and Matthew S Sachs. Translation factor control of ribosome conformation during start codon selection. *Genes Dev.*, 21(11):1280–1287, 2007.
- [258] Katsura Asano. Why is start codon selection so precise in eukaryotes? *Translation*, 2(1):e28387, 2014.
- [259] Lori A Passmore, T Martin Schmeing, David Maag, Drew J Applefield, Michael G Acker, Mikkel A Algire, Jon R Lorsch, and V Ramakrishnan. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell*, 26(1):41–50, 2007.
- [260] Adesh K Saini, Jagpreet S Nanda, Jon R Lorsch, and Alan G Hinnebusch. Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA_i^{Met} binding to the ribosome. *Genes Dev.*, 24(1):97–110, 2010.
- [261] Christoffer Lind and Johan Åqvist. Principles of start codon recognition in eukaryotic translation initiation. *Nucleic Acids Res.*, 44(17):8425–8432, 2016.
- [262] Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.*, 128(14):144120, 2008.

- [263] Jerome Henin, Giacomo Fiorin, Christophe Chipot, and Michael L Klein. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.*, 6(1):35–47, 2010.
- [264] Tanweer Hussain, Jose L Llácer, Israel S Fernández, Antonio Munoz, Pilar Martin-Marcos, Christos G Savva, Jon R Lorsch, Alan G Hinnebusch, and V Ramakrishnan. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell*, 159(3):597–607, 2014.
- [265] David J Hardy, Zhe Wu, James C Phillips, John E Stone, Robert D Skeel, and Klaus Schulten. Multilevel summation method for electrostatic force evaluation. *J. Chem. Theory Comput.*, 11(2):766–779, 2015.
- [266] Yaser Hashem, Amedee des Georges, Vidya Dhote, Robert Langlois, Hstau Y Liao, Robert A Grassucci, Christopher UT Hellen, Tatyana V Pestova, and Joachim Frank. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell*, 153(5):1108–1119, 2013.
- [267] Jiajie Wei, Ying Zhang, Ivaylo P Ivanov, and Matthew S Sachs. The stringency of start codon selection in the filamentous fungus *Neurospora crassa*. *J. Biol. Chem.*, 288(13):9549–9562, 2013.
- [268] Jose L Llácer, Tanweer Hussain, Laura Marler, Colin Echeverría Aitken, Anil Thakur, Jon R Lorsch, Alan G Hinnebusch, and V Ramakrishnan. Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell*, 59(3):399–412, 2015.
- [269] Jose Luis Llácer, Tanweer Hussain, Adesh K Saini, Jagpreet Singh Nanda, Sukhvir Kaur, Yuliya Gordiyenko, Rakesh Kumar, Alan G Hinnebusch, Jon R Lorsch, and Venki Ramakrishnan. Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *Elife*, 7:e39273, 2018.
- [270] Pilar Martin-Marcos, Jagpreet Nanda, Rafael E Luna, Gerhard Wagner, Jon R Lorsch, and Alan G Hinnebusch. β -Hairpin loop of eukaryotic initiation factor 1 (eIF1) mediates 40S ribosome binding to regulate initiator tRNA^{Met} recruitment and accuracy of AUG selection in vivo. *J. Biol. Chem.*, 288(38):27546–27562, 2013.
- [271] Sarah E Kolitz, Julie E Takacs, and Jon R Lorsch. Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA*, 15(1):138–152, 2009.
- [272] Leiming Tang, Jacob Morris, Ji Wan, Chelsea Moore, Yoshihiko Fujita, Sarah Gillaspie, Eric Aube, Jagpreet Nanda, Maud Marques, Maika Jangal, et al. Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Res.*, 45(20):11941–11953, 2017.

- [273] JT Finch, LC Lutter, D Rhodes, RS Brown, B Rushton, M Levitt, and A Klug. Structure of nucleosome core particles of chromatin. *Nature*, 269(5623):29–36, 1977.
- [274] Amber R Cutter and Jeffrey J Hayes. A brief review of nucleosome structure. *FEBS letters*, 589(20):2914–2922, 2015.
- [275] James D McGhee and Gary Felsenfeld. Nucleosome structure. *Annu. Rev. Biochem.*, 49(1):1115–1156, 1980.
- [276] Jayanth V Chodaparambil, Andrew J Barbera, Xu Lu, Kenneth M Kaye, Jeffrey C Hansen, and Karolin Luger. A charged and contoured surface on the nucleosome regulates chromatin compaction. *Nat. Struct. Mol. Biol.*, 14(11):1105–1107, 2007.
- [277] Katalin Tóth, Nathalie Brun, and Jörg Langowski. Chromatin compaction at the mononucleosome level. *Biochemistry*, 45(6):1591–1598, 2006.
- [278] Răzvan V Chereji and Alexandre V Morozov. Functional roles of nucleosome stability and dynamics. *Briefings in functional genomics*, 14(1):50–60, 2015.
- [279] Anna A Kalashnikova, Mary E Porter-Goff, Uma M Muthurajan, Karolin Luger, and Jeffrey C Hansen. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface*, 10(82):20121022, 2013.
- [280] Eva Bártová, Jana Krejčí, Andrea Harničarová, Gabriela Galiová, and Stanislav Kozubek. Histone modifications and nuclear architecture: a review. *J. Histochem. Cytochem.*, 56(8):711–721, 2008.
- [281] Yingming Zhao and Benjamin A Garcia. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.*, 7(9):a025064, 2015.
- [282] Patrick A Grant. A tale of histone modifications. *Genome Biol.*, 2(4):reviews0003–1, 2001.
- [283] Tamaki Suganuma and Jerry L Workman. Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.*, 80:473–499, 2011.
- [284] Peter Cheung, C David Allis, and Paolo Sassone-Corsi. Signaling to chromatin through histone modifications. *Cell*, 103(2):263–271, 2000.
- [285] Anna A Igolkina, Arsenii Zinkevich, Kristina O Karandasheva, Aleksey A Popov, Maria V Selifanova, Daria Nikolaeva, Victor Tkachev, Dmitry Penzar, Daniil M Nikitin, and Anton Buzdin. H3K4me3, H3K9ac, H3K27ac, H3K27me3 and H3K9me3 histone tags suggest distinct regulatory evolution of open and condensed chromatin landmarks. *Cells*, 8(9):1034, 2019.

- [286] Rosa Karlič, Ho-Ryun Chung, Julia Lasserre, Kristian Vlahoviček, and Martin Vingron. Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. USA*, 107(7):2926–2931, 2010.
- [287] Gabriel E Zentner and Steven Henikoff. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.*, 20(3):259, 2013.
- [288] Kevin Struhl and Eran Segal. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.*, 20(3):267, 2013.
- [289] Cizhong Jiang and B Franklin Pugh. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, 10(3):161–172, 2009.
- [290] Carmen Aguilar-Gurrieri, Amédé Larabi, Vinesh Vinayachandran, Nisha A Patel, Kuangyu Yen, Rohit Reja, Ima-O Ebong, Guy Schoehn, Carol V Robinson, B Franklin Pugh, et al. Structural evidence for Nap1-dependent H2A/H2B deposition and nucleosome assembly. *EMBO J.*, 35(13):1465–1482, 2016.
- [291] Tim Formosa. FACT and the reorganized nucleosome. *Mol. Syst. Biol.*, 4(11):1085–1093, 2008.
- [292] Kristin M Keck and Lucy F Pemberton. Histone chaperones link histone nuclear import and chromatin assembly. *Biochim. Biophys. Acta Gene Regul. Mech.*, 1819(3-4):277–289, 2012.
- [293] Keda Zhou, Yang Liu, and Karolin Luger. Histone chaperone fact facilitates chromatin transcription: mechanistic and structural insights. *Curr. Opin. Struct. Biol.*, 65:26–32, 2020.
- [294] Georgy N Rychkov, Andrey V Ilatovskiy, Igor B Nazarov, Alexey V Shvetsov, Dmitry V Lebedev, Alexander Y Konev, Vladimir V Isaev-Ivanov, and Alexey V Onufriev. Partially assembled nucleosome structures at atomic detail. *Biophys. J.*, 112(3):460–472, 2017.
- [295] Maria L Kireeva, Wendy Walter, Vladimir Tchernajenko, Vladimir Bondarenko, Mikhail Kashlev, and Vasily M Studitsky. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol. Cell*, 9(3):541–552, 2002.
- [296] Hiroshi Kimura and Peter R Cook. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.*, 153(7):1341–1354, 2001.
- [297] Ryo Natsume, Masamitsu Eitoku, Yusuke Akai, Norihiko Sano, Masami Horikoshi, and Toshiya Senda. Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature*, 446(7133):338–341, 2007.

- [298] Lisa M Gloss and Brandon J Placek. The Effect of Salts on the Stability of the H2A-H2B Histone Dimer. *Biochemistry*, 41(50):14951–14959, 2002.
- [299] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12):2577–2637, 1983.
- [300] Xiaoying Wang, Susan C Moore, Mario Laszczak, and Juan Ausi6. Acetylation increases the α -helical content of the histone tails of the nucleosome. *J. Biol. Chem.*, 275(45):35013–35020, 2000.
- [301] Yoshihito Moriwaki, Tsutomu Yamane, Hideaki Ohtomo, Mitsunori Ikeguchi, Jun-ichi Kurita, Masahiko Sato, Aritaka Nagadoi, Hideaki Shimajo, and Yoshifumi Nishimura. Solution structure of the isolated histone H2A/H2B heterodimer. *Sci. Rep.*, 6:24999, 2016.
- [302] Jeffrey C Hansen, Xu Lu, Eric D Ross, and Robert W Woody. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.*, 281(4):1853–1856, 2006.
- [303] Zhenling Peng, Marcin J Mizianty, Bin Xue, Lukasz Kurgan, and Vladimir N Uversky. More than just tails: intrinsic disorder in histone proteins. *Mol. Biosyst.*, 8(7):1886–1901, 2012.
- [304] Yasuhiro Arimura, Hiroaki Tachiwana, Takashi Oda, Mamoru Sato, and Hitoshi Kurumizaka. Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry*, 51(15):3302–3309, 2012.
- [305] Silvija Bilokapic, Mike Strauss, and Mario Halic. Histone octamer rearranges to adapt to DNA unwrapping. *Nat. Struct. Mol. Biol.*, 25(1):101, 2018.
- [306] Manel A Balsera, Willy Wriggers, Yoshitsugu Oono, and Klaus Schulten. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100(7):2567–2572, 1996.
- [307] Daiki Kato, Akihisa Osakabe, Yasuhiro Arimura, Yuka Mizukami, Naoki Horikoshi, Kazumi Saikusa, Satoko Akashi, Yoshifumi Nishimura, Sam-Yong Park, Jumpei Nogami, et al. Crystal structure of the overlapping dinucleosome composed of hexasome and octasome. *Science*, 356(6334):205–208, 2017.
- [308] Maurice Kleman and Oleg D Laverntovich. *Soft matter physics: an introduction*. Springer Science & Business Media, 2007.
- [309] Masao Doi. *Soft matter physics*. Oxford University Press, 2013.
- [310] David Maag, Mikkel A Algire, and Jon R Lorsch. Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *J. Mol. Biol.*, 356(3):724–737, 2006.

- [311] Vineet Bafna, S Muthukrishnan, and R Ravi. Computing similarity between RNA strings. In *Annual Symposium on Combinatorial Pattern Matching*, pages 1–16. Springer, 1995.
- [312] Dan Gusfield, Krishnan Balasubramanian, and Dalit Naor. Parametric optimization of sequence alignment. *Algorithmica*, 12(4):312–326, 1994.
- [313] Francesca Chiaromonte, VB Yap, and W Miller. Scoring pairwise genomic sequence alignments. In *Biocomputing 2002*, pages 115–126. World Scientific, 2001.
- [314] Zhichao Miao and Eric Westhof. RNA structure: advances and assessment of 3D structure prediction. *Annu Rev Biophys*, 46:483–503, 2017.
- [315] Joerg Fallmann, Sebastian Will, Jan Engelhardt, Björn Grüning, Rolf Backofen, and Peter F Stadler. Recent advances in RNA folding. *J. Biotechnol.*, 261:97–104, 2017.
- [316] Jiri Sponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, et al. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.*, 118(8):4177–4338, 2018.
- [317] Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.
- [318] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.
- [319] James V Stone. Independent component analysis. *Encyclopedia of statistics in Behavioral Science*, 2005.
- [320] Yusuke Naritomi and Sotaro Fuchigami. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.*, 134(6):02B617, 2011.
- [321] Toshifumi Mori and Shinji Saito. Dynamic heterogeneity in the folding/unfolding transitions of FiP35. *J. Chem. Phys.*, 142(13):04B601_1, 2015.
- [322] Yutaka Maruyama, Hiroshi Takano, and Ayori Mitsutake. Analysis of molecular dynamics simulations of 10-residue peptide, chignolin, using statistical mechanics: Relaxation mode analysis and three-dimensional reference interaction site model theory. *Biophysics and Physicobiology*, 16:407–429, 2019.
- [323] Andrew P Allen, James H Brown, and James F Gillooly. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, 297(5586):1545–1548, 2002.

- [324] Eran Segal, Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Thåström, Yair Field, Irene K Moore, Ji-Ping Z Wang, and Jonathan Widom. A genomic code for nucleosome positioning. *Nature*, 442(7104):772–778, 2006.
- [325] Akinori Awazu. Prediction of nucleosome positioning by the incorporation of frequencies and distributions of three different nucleotide segment lengths into a general pseudo k-tuple nucleotide composition. *Bioinformatics*, 33(1):42–48, 2017.
- [326] Erinija Pranckeviciene, Sergey Hosid, Nathan Liang, and Ilya Ioshikhes. Nucleosome positioning sequence patterns as packing or regulatory. *PLoS Comput. Biol.*, 16(1):e1007365, 2020.
- [327] Dong Deng, Chuangye Yan, Xiaojing Pan, Magdy Mahfouz, Jiawei Wang, Jian-Kang Zhu, Yigong Shi, and Nieng Yan. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, 335(6069):720–723, 2012.
- [328] Naoya Tochio, Kohei Umehara, Jun-ichi Uewaki, Holger Flechsig, Masaharu Kondo, Takehisa Dewa, Tetsushi Sakuma, Takashi Yamamoto, Takashi Saitoh, Yuichi Togashi, et al. Non-RVD mutations that enhance the dynamics of the TAL repeat array along the superhelical axis improve TALEN genome editing efficacy. *Sci. Rep.*, 6:37887, 2016.
- [329] Timothy R Blosser, Luuk Loeff, Edze R Westra, Marnix Vlot, Tim Künne, Małgorzata Sobota, Cees Dekker, Stan JJ Brouns, and Chirlmin Joo. Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell*, 58(1):60–70, 2015.
- [330] Patrick D Hsu, Eric S Lander, and Feng Zhang. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262–1278, 2014.
- [331] Jennifer A Doudna and Emmanuelle Charpentier. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), 2014.
- [332] Hunter B Fraser, Aaron E Hirsh, Lars M Steinmetz, Curt Scharfe, and Marcus W Feldman. Evolutionary rate in the protein interaction network. *Science*, 296(5568):750–752, 2002.
- [333] Joshua L Payne and Andreas Wagner. The robustness and evolvability of transcription factor binding sites. *Science*, 343(6173):875–877, 2014.
- [334] Sebastian J Maerkl and Stephen R Quake. A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, 315(5809):233–237, 2007.
- [335] Sunhwan Jo and Wei Jiang. A generic implementation of replica exchange with solute tempering (REST2) algorithm in NAMD for complex biophysical simulations. *Comput. Phys. Commun.*, 197:304–311, 2015.

- [336] Nicholas M Luscombe, Dov Greenbaum, and Mark Gerstein. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med*, 40(04):346–358, 2001.
- [337] Jin Xiong. *Essential bioinformatics*. Cambridge University Press, 2006.
- [338] Jacques Cohen. Bioinformatics—an introduction for computer scientists. *ACM Comput Surv*, 36(2):122–158, 2004.
- [339] Andrea M Chiariello, Carlo Annunziatella, Simona Bianco, Andrea Esposito, and Mario Nicodemi. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.*, 6:29775, 2016.
- [340] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of chromosomal domains by loop extrusion. *Cell Rep.*, 15(9):2038–2049, 2016.
- [341] Marco Di Stefano, Jonas Paulsen, Tonje G Lien, Eivind Hovig, and Cristian Micheletti. Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. *Sci. Rep.*, 6:35985, 2016.
- [342] Soya Shinkai, Masaki Nakagawa, Takeshi Sugawara, Yuichi Togashi, Hiroshi Ochiai, Ryuichiro Nakato, Yuichi Taniguchi, and Shuichi Onami. PHi-C: deciphering Hi-C data into polymer dynamics. *NAR Genomics and Bioinformatics*, 2(2):lqaa020, 2020.
- [343] Clifford A Meyer and X Shirley Liu. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.*, 15(11):709–721, 2014.
- [344] Pedro Madrigal. On accounting for sequence-specific bias in genome-wide chromatin accessibility experiments: recent advances and contradictions. *Front. Bioeng. Biotechnol.*, 3:144, 2015.
- [345] Satoru Ishihara, Yohei Sasagawa, Takeru Kameda, Mana Umeda, Hayato Yamashita, Naoe Kotomura, Masayuki Abe, Yohei Shimono, and Itoshi Nikaido. The local state of chromatin compaction at transcription start sites controls transcription levels. *bioRxiv*, 2019.
- [346] Wilma K Olson, Andrew V Colasanti, Yun Li, Wei Ge, Guohui Zheng, and Victor B Zhurkin. DNA simulation benchmarks as revealed by X-ray structures. In *Computational Studies of RNA and DNA*, pages 235–257. Springer, 2006.
- [347] Alexandre V Morozov, Karissa Fortney, Daria A Gaykalova, Vasily M Studitsky, Jonathan Widom, and Eric D Siggia. Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res.*, 37(14):4707–4722, 2009.