ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations

Shoji Takada,*[,†] Ryo Kanada,[†] Cheng Tan,[†] Tsuyoshi Terakawa,[‡] Wenfei Li,[§] and Hiroo Kenzaki[⊥]

[†]Department of Biophysics, Graduate School of Science, Kyoto University, Sakyo, Kyoto 6068502, Japan

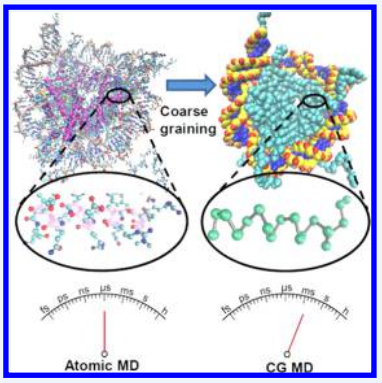
[‡]Department of Biochemistry and Molecular Biophysics, Columbia University, 650 W 168 Street New York, New York 10032, United States

[§]Department of Physics, Nanjing University, Nanjing 210093, China

[⊥]Advanced Center for Computing and Communication, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan

Supporting Information

CONSPECTUS: Due to hierarchic nature of biomolecular systems, their computational modeling calls for multiscale approaches, in which coarse-grained (CG) simulations are used to address long-time dynamics of large systems. Here, we review recent developments and applications of CG modeling methods, focusing on our methods primarily for proteins, DNA, and their complexes. These methods have been implemented in the CG biomolecular simulator, CafeMol.



Our CG model has resolution such that ∼10 non-hydrogen atoms are grouped into one CG particle on average. For proteins, each amino acid is represented by one CG particle. For DNA, one nucleotide is simplified by three CG particles, representing sugar, phosphate, and base. The protein modeling is based on the idea that proteins have a globally funnel-like energy landscape, which is encoded in the structure-based potential energy function. We first describe two representative minimal models of proteins, called the elastic network model and the classic Gō model. We then present a more elaborate protein model, which extends the minimal model to incorporate sequence and context dependent local flexibility and nonlocal contacts. For DNA, we describe a model developed by de Pablo's group that was tuned to well reproduce sequence-dependent structural and thermodynamic experimental data for single- and double-stranded DNAs. Protein−DNA interactions are modeled either by the structure-based term for specific cases or by electrostatic and excluded volume terms for nonspecific cases.

We also discuss the time scale mapping in CG molecular dynamics simulations. While the apparent single time step of our CGMD is about 10 times larger than that in the fully atomistic molecular dynamics for small-scale dynamics, large-scale motions can be further accelerated by two-orders of magnitude with the use of CG model and a low friction constant in Langevin dynamics.

Next, we present four examples of applications. First, the classic Gō model was used to emulate one ATP cycle of a molecular motor, kinesin. Second, nonspecific protein−DNA binding was studied by a combination of elaborate protein and DNA models. Third, a transcription factor, p53, that contains highly fluctuating regions was simulated on two perpendicularly arranged DNA segments, addressing intersegmental transfer of p53. Fourth, we simulated structural dynamics of dinucleosomes connected by a linker DNA finding distinct types of internucleosome docking and salt-concentration-dependent compaction.

Finally, we discuss many of limitations in the current approaches and future directions. Especially, more accurate electrostatic treatment and a phospholipid model that matches our CG resolutions are of immediate importance.

1. INTRODUCTION

One of the long-term challenges in computational chemistry is to accurately model structures, dynamics, and functions of complex molecules, such as soft matter and biological molecules. Major success therein was achieved by multiscale approaches.[1] At the finest level, chemical reactions are treated by quantum chemical calculations. Conformational changes of flexible molecules and intermolecular assemblies have been addressed by classical molecular-mechanics molecular dynamics (MD) simulations. Longer time-scale dynamics of larger molecular complexes have been approached by coarse-grained (CG) simulations.

Currently, CG simulations seem to be established to a lesser extent than the former two. This is primarily because there are inherently diverse formulations of the CG models; the average number of atoms that are grouped together, solvent−water treatment, and so on are all varied in different CG models. Moreover, deriving CG models, we often need to give priority

Received: July 20, 2015
Published: November 17, 2015

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

to reproduce some properties, such as thermodynamics, structures, or specific experiments, which often sacrifices others. The choice of priority is dependent on what we want to know from the simulation.

Given these, here we review recent developments and applications of CG modeling methods for biomolecular systems, primarily proteins, DNA, and their complexes, focusing on those we have been studying over years. These methods have been implemented in the CG biomolecular simulator, CafeMol.[2] The primary purpose of this Account is to provide concise yet reasonably self-explanatory descriptions of these methods and software, together with a few representative applications. For many other types of CG modeling methods and softwares, refer to recent excellent reviews[3−9] and in particular Table 1 in ref 10.

2. COARSE-GRAINED MOLECULAR DYNAMICS

First, we set the resolution in our CG model such that ∼10 non-hydrogen atoms are grouped into one CG particle on average (Figure 1A). In practice, for proteins, each amino acid is represented by one CG particle located at its Cα position. For DNA, one nucleotide is simplified by three CG particles, representing sugar, phosphate, and base. The solvent molecule is not explicitly treated; instead, the solvent effect is implicitly taken into account via effective potential energy between CG particles.

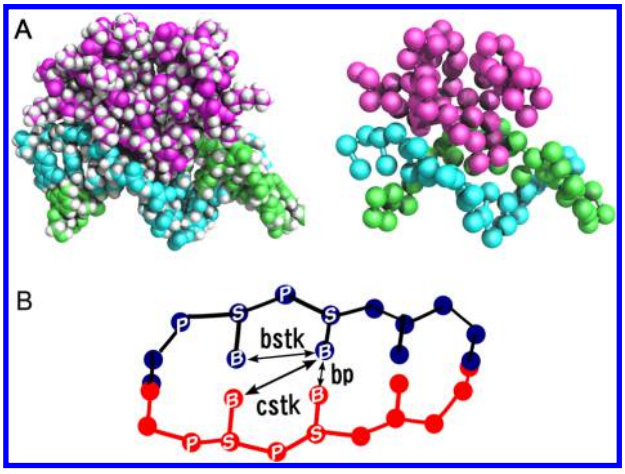


Figure 1. Schematics of coarse-grained (CG) protein and DNA models. (A) The atomistic (left) and a CG (right) representation of a protein−DNA complex. The protein is Sso7d (purple) and dsDNA (cyan and green) contains 12 bps. (B) A schematic view of the CG dsDNA model, 3SPN.2C. Each nucleotide is modeled by sugar (S), phosphate (P), and base (B). Three types of base−base interactions are denoted; base stacking (bstk), base pairing (bp), and cross stacking (cstk).

The basic idea in our protein modeling is that proteins have evolved their sequences so that their interactions are minimally frustrated at their native structures.[11,12] This leads to the idea that proteins have globally funnel-like energy landscape.[13] The structures and interactions in the native state of proteins are sequence specific. Thus, these interactions should be primarily realized by the side chain interactions. When we use the Cα CG models, however, these side chain interactions cannot be fully reproduced. To compensate these interactions in the Cα CG models, we need the native-structure-based model where the potential energy function is explicitly biased to the reference native structure.[10]

We also note that there are two goals in the development of CG models.[14] The first goal is to make it "as simple as possible, but not simpler", after Einstein's epigram. With its simplicity, one can easily see what is in the model and what is not, from which one can obtain crisp insights. The other goal is to make them as accurate as possible, which is more standard in general modeling methods. Here, we describe both types of approaches.

All the models described below are implemented in CafeMol[2] (See http://www.cafemol.org for updates).

2.1. As Simple As Possible: Minimal Modeling

We begin with an extremely simple model of proteins, the elastic network model of proteins first proposed by Tirion,[15] which is defined by

$$V_{\mathrm{ENM}}(R|R_0) = \sum_{ij,s.t.r_{ij}<r_c} k(r_{ij} - r_{ij,0})^2 \tag{1}$$

where r_{ij} is the distance between the ith and jth particles. Often, a CG particle represents an amino acid located at its Cα position. The $r_{ij,0}$ is the corresponding distance at the native (reference) structure R_0 (the meaning of subscript 0 is the same throughout the paper). The k represents the elasticity and is usually a uniform value independent of amino acids. The summation is taken over amino acid pairs that are physically "in contact" in R_0. The precise definition of "in contact" can vary but, in a simple treatment, uses a threshold distance, r_c, between the two amino acids. The elastic network model has been broadly applied to predict native fluctuations and conformational changes in giant biomolecules.[16−18] Many of its extensions and applications can be found in ref 19.

One clear limitation of the elastic network model is that, due to its harmonic nature, it cannot approximate large-amplitude motions including local and global unfolding. A class of protein models that allow break of the harmonic bonds in the elastic network potential are often called Gō models after seminal lattice models of similar idea.[20,21] Gō models have widely been used in studies of protein folding as well as protein function that involves large-amplitude motions.[22−26] Here, we write down the potential energy function of the classic Cα Gō model proposed by Clementi, Nymeyer, and Onuchic,[22]

$$\begin{aligned} V_{\mathrm{classic}}(R|R_0) = {} & \sum_i k_{\mathrm{b}}(r_{i,i+1} - r_{i,i+1,0})^2 + \sum_i k_\theta(\theta_i - \theta_{i,0})^2 \\ & + \sum_i \{k_{\phi 1}[1 - \cos(\phi_i - \phi_{i,0})] \\ & + k_{\phi 3}[1 - \cos 3(\phi_i - \phi_{i,0})]\} \\ & + \sum_{i<j-3}^{\text{native contact}} \varepsilon_{\mathrm{go}}\left[5\left(\frac{r_{ij,0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij,0}}{r_{ij}}\right)^{10}\right] \\ & + \sum_{i<j-3}^{\text{non-native}} \varepsilon_{\mathrm{ev}}\left(\frac{d}{r_{ij}}\right)^{12} \end{aligned} \tag{2}$$

where θ_i is the ith virtual angle made by two consecutive virtual bonds and ϕ_i is the ith dihedral angle defined by three consecutive virtual bonds. In the energy function, the first, second, and third terms represent restraint potentials for virtual bond lengths, virtual bond angles, and virtual dihedral angles,

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

respectively. The fourth term is the nonlocal contact potential that stabilizes amino acid pairs, where the summation is restricted to amino acid pairs that are in proximity at the reference structure. The last term represents a generic excluded volume effect. The coefficients k and ε are parameters that modulate weights among these terms.

Similar modeling can be done for nucleic acids and protein−nucleic acid complexes.[4,16,27]

2.2. As Accurate As Possible

While minimal CG modeling has been successful in many applications, it also revealed clear limitations in accuracy. These limitations motivated us to improve the model. A promising way is to make it a higher resolution model. In fact, an atomistic Gō model has proven to be useful.[28] If we go further, however, the model becomes closer to standard atomistic molecular mechanics, which diminishes the merit of CG modeling, that is, speed. Alternatively, here we seek to improve the accuracy as much as possible, while retaining the resolution.

One caveat in the classic Gō model, eq 2, is local interactions being too stiff; the bond-angle and dihedral-angle potential energy terms restrain these angles around the reference values. In reality, potential energies of both angles should have at least two distinct energy minima that correspond to α-helix and β-strand/loop. Restraining to one angle value is particularly problematic when (parts of) proteins are flexible or disordered. To overcome this limitation, we introduced a generic flexible local potential, $V_{\mathrm{loc}}^{\mathrm{flp}}(R)$ as the inverse Boltzmann statistical potentials,[29] which do not rely on any reference structure,

$$V_{\mathrm{loc}}^{\mathrm{flp}}(R) = -k_{\mathrm{B}}T\sum_i \ln\frac{P_\theta(\theta_i|i)}{\sin\theta_i} - k_{\mathrm{B}}T\sum_i \ln P_\phi(\phi_i|i) \tag{3}$$

where $P_\theta(\theta_i|i)$ ($P_\phi(\phi_i|i)$) is the angle (dihedral-angle) probability distribution dependent on the residue type of the ith amino acid (and partly the type of $(i+1)$th amino acid). Parameters k_{B} and T are the Boltzmann constant and the temperature, respectively. The denominator in the first term comes from Jacobian determinant for bond angles. As a generic knowledge-based approach, we used the probability distribution of the corresponding angle obtained from a loop-structure library that was collected from the Protein Data Bank (PDB). As a more-accurate approach specific to a target protein, we can conduct atomistic MD simulations for short fragments of the target, from which we can generate the probability distributions. This tailor-made flexible local potential applied to the p53 N-terminal disordered region showed a more accurate ensemble than the generic PDB knowledge-based potential.[29]

Another limitation in the classic Gō model of proteins is a uniform parameter $\varepsilon_{\mathrm{go}}$ in the native contact potential. On looking into the allosteric conformational change between two structures, we found that contacts specific to one structure tend to be clearly weaker than those common in the two structures.[30] To discriminate them in the CG model, we generalized the contact strength to be sequence- and context-dependent.[30] The newly introduced nonuniform parameters were determined by the multiscale algorithms, based partly on the atomic interaction estimated by the AMBER force field in the reference structure and also on the fluctuation-matching protocol.[31] Including this generalization as well as the flexible local potential, we defined the atomic interaction-based CG model, version 2+ (AICG2+)[32,33] as

$$\begin{aligned} V_{\mathrm{AICG2+}}(R|R_0) = {} & \sum_i k_{\mathrm{b},i}(r_{,i,i+1} - r_{i,i+1,0})^2 + V_{\mathrm{loc}}^{\mathrm{flp}}(R) \\ & + \sum_i \varepsilon_{\mathrm{loc2},i}\exp\left(-\frac{(r_{i-1,i+1} - r_{i-1,i+1,0})^2}{2W_{i-1,i+1}{}^2}\right) \\ & + \sum_i \varepsilon_{\mathrm{loc3},i}\exp\left(-\frac{(\phi_i - \phi_{i0})^2}{2W_{\phi,i}{}^2}\right) \\ & + \sum_{i<j-3}^{\text{nat contact}} \varepsilon_{\mathrm{go},ij}\left[5\left(\frac{r_{ij0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij0}}{r_{ij}}\right)^{10}\right] \\ & + \sum_{i<j-3}^{\text{non-native}} \varepsilon_{\mathrm{ev}}\left(\frac{d}{r_{ij}}\right)^{12} \end{aligned} \tag{4}$$

The third and fourth terms represent the structure-based-angle and dihedral-angle potentials, respectively. The parameters W define widths of the Gaussian. The fourth term was originally represented by the distance $r_{i,i+3}$ in the AICG2 model, which turned out to generate, albeit rarely, quasi-mirror-image misfolded structures. This problem was solved by rewriting this term in terms of the corresponding dihedral angles, ϕ_i, in the AICG2+ model.[33] In the fifth term, nonuniform parameters $\varepsilon_{\mathrm{go},ij}$, which are tuned by multiscale algorithms, depend on the sequence. RNA that can fold to tertiary structures was modeled in a somewhat similar way.[34]

The AICG2+ model is superior to the elastic network model and the classic Gō model in accuracy of predictions of small fluctuations in the native state, allosteric conformational change direction, and folding pathway ensembles.[30,32] As an example, in Figure 2A, we compare the AICG2+ model and the atomistic model in terms of the root-mean-square fluctuations of each residue in protein G. The atomistic simulation was performed for 50 ns with AMBER force field ff99SB and TIP3P. The correlation coefficient between the AICG2+ and atomistic simulation results was 0.86. Interestingly, for the same protein, the correlation coefficient between atomistic simulation results with explicit and implicit water models (a GBSA model) was 0.85, implying that the AICG2+ model is as accurate as the GBSA implicit solvent atomistic model in terms of the small fluctuations around the native state. We note that the computational cost of the AICG2+ model and the classic Gō model is nearly the same.

Next, we briefly discuss the last term in eqs 2 and 4, the excluded volume term. As far as reference structure-based interactions dominate the system, as in the case of protein folding simulations, the excluded volume term, which is the non-native interaction, does not play important roles, by design. Thus, results are insensitive to the parameter d in this term. On the other hand, when generic, that is, non-structure-based, protein−protein or protein−DNA interactions play major roles, this excluded volume parameter, which controls the accessible distance between the molecules, is important. To address the excluded volume parameter, we estimated nonstructure-based protein−protein interaction energy, which contains a hydrophobic, an electrostatic (described below), and the excluded volume interactions, $\sum 0.2(d/r_{ij})^{12}$. Starting from the native complex structure, we shifted the relative position along one-dimensional coordinate δ ($\delta = 0$ is the native, while $\delta > 0$ ($\delta < 0$) corresponds to the dissociation (the collapse)) and calculated the intermolecular energy. Figure 2B depicts the

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

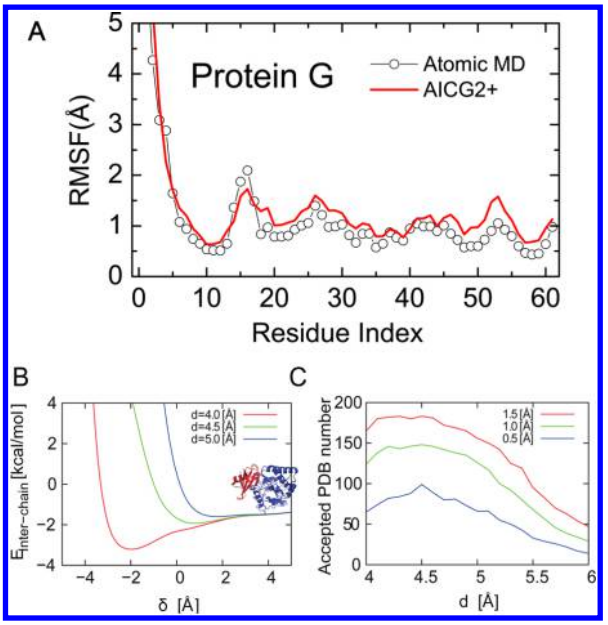


Figure 2. Accuracy in CG protein models. (A) The root-mean-square fluctuations (RMSF) around the native state in protein G (PDB ID 2IGD) at 300 K calculated by atomistic (open circles) and AICG2+ (red curve) MDs. (B) The protein−protein interaction energy as a function of displacement δ with three excluded volume parameters, d, for uracil−DNA glycosylase (blue in the inset cartoon) and its inhibitor complex (red) (PDB ID 2J8X). Of the three d's, the shift at the energy minima $|\delta_{min}|$ is the smallest with d = 4.5 Å. (C) Of the 251 protein−protein complexes tested, the numbers of complexes that satisfy $|\delta_{min}|$ < threshold are plotted against d for three thresholds = 1.5 (red), 1.0 (green), and 0.5 (blue) Å.

result for uracil−DNA glycosylase (blue in the inset cartoon) and its inhibitor complex (red) (PDB ID 2J8X), where $d \approx 4.5$ Å would be optimal to have the energy minimum δ_{min} closest to zero. Indeed, a statistical survey over 251 protein−protein complexes suggested that $d \approx 4.5$ Å is a good choice for interprotein excluded volumes (Figure 2C).

For DNA, we have employed a series of CG models developed in de Pablo's group, which take three CG particles per nucleotide.[35−38] This resolution seems to match that of CG protein models described above. The first model was reported in 2007,[35] which was updated to 3SPN.1,[36] and to 3SPN.2[37] and its variant 3SPN.2C[38] (Figure 1B). Here, we briefly describe the 3SPN.2C model.

The 3SPN.2C model uses three CG particles per nucleotide, representing the phosphate (P), sugar (S), and base (N). The potential energy function is written as

$$\begin{aligned} V_{3SPN.2C} = & \sum_{ij\in bonded} [K_b(r_{ij} - r_{ij,0})^2 + 100K_b(r_{ij} - r_{ij,0})^4] \\ & + \sum_i K_{\theta,i}(\theta_i - \theta_{i,0})^2 \\ & - \sum_{i\in SPSP,PSPS} K_{\phi\text{-Gaussian}} \exp\left[\frac{-(\phi_i - \phi_{i,0})^2}{2\sigma_{\phi,i}^2}\right] \\ & + \sum_{i\in all\ dihedral} K_{\phi\text{-periodic}}[1 + \cos(\phi_i - \phi_{i,0})] \\ & + V_{bstk} + V_{cstk} + V_{bp} + V_{elec} + V_{ev} \end{aligned} \quad (5)$$

Here, the first and the second terms are for virtual bonds and virtual angles, respectively. The third term stabilizes backbone dihedral angles at their reference values, while the fourth term weakly biases to the reference values for all the dihedral-angles. Note that the reference structure, R_0, is not the straight B-type DNA but takes sequence-dependent curved B-type DNA generated by the software 3DNA.[39]

The rest in eq 5 represents nonlocal interactions. The fifth to seventh terms are for base−base interactions, each representing the base stacking (V_{bstk}), the cross-base stacking (V_{cstk}), and the base pairing (V_{bp}) (Figure 1B for the definition of three base−base interactions). Both in 3SPN.2 and 3SPN.2C, these interactions are modeled as angle-dependent attractive potentials. For example, V_{bstk} takes the form,

$$V_{bstk} = \sum u_{ij}^{rep}(r_{ij}) + f_{filter}\, u_{ij}^{attr}(r_{ij})$$

where $u_{ij}^{rep}(r_{ij})$ and $u_{ij}^{attr}(r_{ij})$ are repulsive and attractive parts of Morse potential, respectively. The coefficient f_{filter} in the attractive part is an angle-dependent filter that depends on the angle defined by stacking bases and a neighboring particle. The filter f_{filter} takes the maximum value when the angles are close to those in the reference structures and smaller otherwise. The angle dependence makes the base-stacking sensitive to local geometry of the stacking bases, which is crucial to make dsDNA rigid and single-stranded DNA flexible. The base-pairing and cross-base-stacking interactions take the same form although the precise forms of f_{filter} and parametrization are rather different. See the original paper for more details.[37]

The eighth term in eq 5 is the electrostatics, modeled by the Debye−Hückel theory[40] as

$$V_{elec} = \sum_{i<j} \frac{1}{4\pi\varepsilon(T, C)} \frac{q_i q_j\, e^{-r_{ij}/\lambda_D}}{r_{ij}} \quad (6)$$

where $\varepsilon(T,C)$ is the dielectric constant that depends on the temperature and solvent ionic concentration, q_i is the charge of ith particle, and λ_D is the Debye length, which controls ionic screening depending on the solvent salt concentration. The last term in eq 5 is a generic excluded volume term.

Protein−protein as well as protein−DNA intermolecular interactions can be modeled in multiple ways. We often divide the intermolecular interactions into either specific or nonspecific although many interactions may fall in between these two limits. Usually, nanomolar order dissociation constants are interpreted as specific. Specific interactions can simply be modeled via structure-based contact potentials such as the fifth term in eq 4. The coefficients $\varepsilon_{go,ij}$ can be modeled either based on atomistic force fields or by high-throughput experimental data represented by the position−weight matrix for protein−DNA interactions.[41]

Nonspecific protein−protein interactions contain two major contributions; hydrophobic interactions and electrostatic interactions. In CafeMol, we estimate hydrophobic interactions as an empirical nonadditive potential that measures buriedness of amino acids.[2]

The electrostatic interactions are usually modeled by the Debye−Hückel model, eq 6, where determination of the charges for every CG particle is of crucial importance. In many early protein modeling efforts, we used either +1e, 0, or −1e charge for amino acids based on the standard pK_a values. Note that while charged amino acids carry their charge at the termini of their side chains, our CG modeling contains only Cα

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

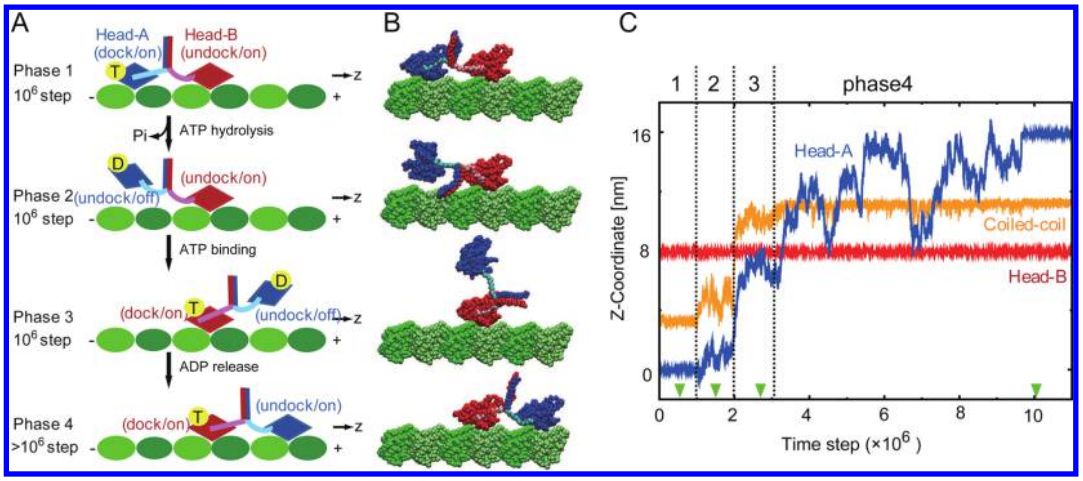


Figure 3. Kinesin stepping in CGMD simulations. (A) The simulation scheme for one ATP hydrolysis cycle based on a traditional hand-over-hand model. The blue and red rhombuses are the kinesin head A and head B, respectively. The cyan and pink strings near the heads are the neck-linkers. Two heads dimerize at the coiled-coil regions drawn as blue and red bars. The "T" and "D" stand for ATP and ADP bound on kinesin heads, respectively. The light green and green ellipsoids correspond to the $\alpha\beta$-tubulins. The neck-linker state is indicated by either dock or undock. The affinity to tubulin is indicated by either on (strong affinity) or off (no affinity). Our simulation consists of four phases defined by the combination of states of two heads. (B) Representative snapshots in the CGMD simulation. The color scheme is the same as panel A. (C) A representative time course of the coordinate along which kinesin proceeds. The positions of head A are drawn in blue, head B in red, and the root of coiled-coil helix in orange. The vertical dotted lines indicate boundaries of four phases. The green triangles at the bottom indicate the times when snapshots in panel B are taken.

coordinates. Thus, the simple integer charge assignment could be relatively poor. To overcome this issue, we recently proposed the RESPAC method.[42] This method derives from the RESP method broadly used in the partial charge determination in atomistic force field. For a given protein, we first compute the reference electrostatic potential around the protein by the Poisson−Boltzmann equation with the atomistic model. Then, a set of partial charges of CG amino acids are determined to minimize the difference between the Debye−Hückel-based CG and the reference atomic electrostatic potentials.[42]

There are different natures of electrostatic interaction between protein−DNA and DNA−DNA interactions. DNA−DNA electrostatic interactions are highly repulsive in short-range and thus do not involve direct atomic contacts. Thus, counterion condensation shields bare charges. In 3SPN.2 and 3SPN.2C models, a charge of $-0.6e$ was assigned for the phosphate groups in DNA−DNA interaction. On the other hand, for protein−DNA interactions, we are normally interested in short-ranged attractive interactions in which counterion condensation may not occur. Thus, a charge of $-1e$ for the phosphate group of DNA might be reasonable. We will discuss this issue in section 3.2, again.

2.3. Mapping Time Scales

We employ, by default, the standard Langevin equation for CG particles, $\mathbf{r}_i$,

$$m_i \frac{\mathrm{d}^2\mathbf{r}_i}{\mathrm{d}t^2} = -\frac{\partial V_{\mathrm{total}}}{\partial \mathbf{r}_i} - m_i\gamma\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} + m_i\boldsymbol{\xi}_i$$

as the equation of motion. Here m_i is the mass of ith CG particle and γ is the friction constant. The random noise vector, $\boldsymbol{\xi}_i$, has its element $\xi_{i,\mu}$ being the Gaussian white noise with the mean and variance,

$$\langle \xi_{i,\mu}(t)\rangle = 0, \quad \langle \xi_{i,\mu}(t)\xi_{j,\nu}(t')\rangle = \frac{2\gamma k_{\mathrm{B}}T}{m_i}\delta(t-t')\delta_{i,j}\delta_{\mu,\nu}$$

This fluctuation−dissipation relation guarantees that the system converges to the canonical ensemble at the temperature T. With this form, the hydrodynamic effect is ignored.

One of the most frequently asked questions (not easy to answer, though) for CGMD simulation is about the time scale. Unfortunately, it depends on the target molecules and the quantity we look into. In the past, we mapped the time scale case-by-case.

If we literally calculate the time scale using the Langevin equation, masses of CG particles, and the spring constants for virtual bond stretching, the single MD step corresponds to ∼10 fs. Compared with atomistic MD, one can use ∼10 times longer time step with Cα-based CGMD. This can be understood because the CG effective mass for the bond stretching is approximately 100 times as large as that in the atomistic MD and the period of oscillation scales as the square root of the mass.

Next, we discuss the time scale mapping with the longer time scales for larger motions, conformational dynamics of allosteric proteins. In a recent study on adenylate kinase, we looked into conformational relaxation dynamics with the principal component analysis.[43] For the largest-fluctuation mode, principal component 1, the decay times in the autocorrelation function were computed by fully atomistic MD and by CGMD simulations. By comparing these two time scales, we estimated that the single MD step corresponds to ∼1 ps. Notably, in this CGMD simulation, we used a small friction constant, $\gamma = 0.02$, which is 2 orders of magnitude smaller than that estimated from water viscosity. This may partly explain the difference from the microscopic mapping.

In another context, we addressed the time scale mapping from the diffusion coefficient of protein domains. For globular molecules, we can roughly estimate the diffusion coefficient by the Einstein−Stokes equation. Separately, we estimate the diffusion coefficient from CGMD simulations that used $\gamma = 0.02$. Comparison between these two estimates gives a time scale mapping. In the case of p53 core domain, the single MD

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

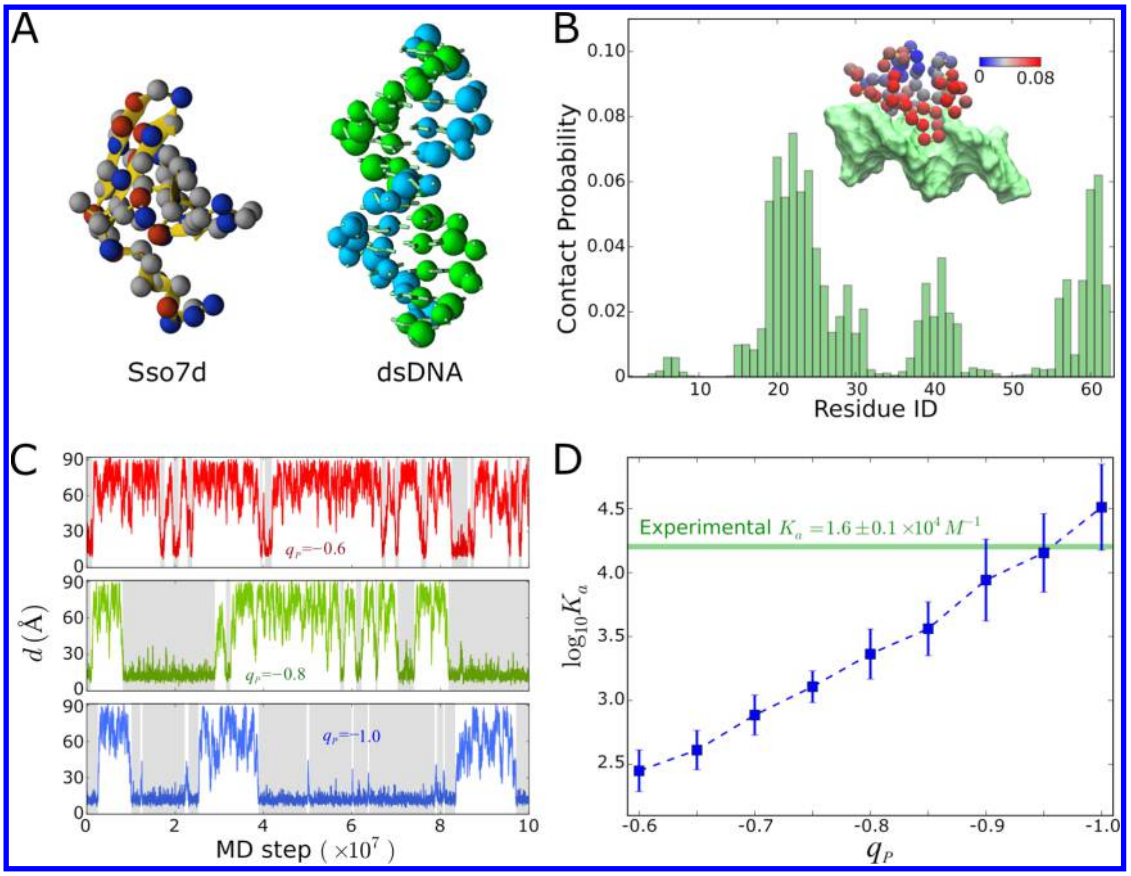


Figure 4. Nonspecific protein−DNA interactions. (A) The reference and initial structures of a DNA-binding protein Sso7d and the 12 bp dsDNA (PDB ID 1BBX). Positively (negatively) charged amino acid residues are colored blue (red). The two strands of DNA are colored cyan and green. (B) Probability for Sso7d residues to form contacts (<7.0 Å) with DNA in the simulations with $q_P = -1.0e$. Inset, protein residues are shown in colors according to their contact probabilities. (C) Representative time series of the distance between centers of mass of Sso7d and DNA in CGMD simulations with different values of phosphate charges ($q_P = -0.6e$, $-0.8e$, and $-1.0e$) in protein−DNA interactions. The shadowed regions represent the snapshots with $d < 30$ Å, which are considered as the bound state. (D) The logarithm of the simulated association constant (K_a, in the unit of M^{-1}) of Sso7d-DNA binding is shown as a function of the phosphate charge. The horizontal green line indicates the experimental K_a value.

step corresponds to ∼1 ps (Terakawa and Takada, *Sci. Rep.* **2015**, *5*:17107, DOI: 10.1038/srep17107).

3. APPLICATIONS

3.1. Molecular Machines: Two-Headed Kinesin

The first example of the application is a molecular motor, kinesin. Kinesin represents a family of ATP-driven molecular motors that proceed along microtubules. The conventional kinesin forms a homodimer where the two-ATPase domains are connected by the coiled-coil regions. It is known to proceed 8 nm per ATP hydrolysis by the hand-over-hand mechanism (Figure 3A). One ATP-cycle corresponds to one stepping of a head: Starting from the two-head bound state (the top in Figure 3A), the rear head (blue) detaches from tubulins (green) (the second cartoon), followed by its diffusive motions (the third) and attachment to the 16 nm forward binding site (the bottom). Here, assuming a particular chemical pathway shown in Figure 3A, we performed simulations of this one ATP cycle with CafeMol (see Kanada et al.[44] for more detail; watch Movie S1). The purpose here is to exemplify that we can emulate molecular motors at work rather easily by our CGMD method.

The simulation system contained eight protein subunits: Dimeric kinesin molecules (head A and head B) that moved dynamically and three copies of tubulin $\alpha\beta$ dimers that were fixed as a single protofilament of microtubule aligned along the z-axis (Figure 3A). All the proteins were modeled by the classic Gō model.

To emulate conformational changes of dimeric kinesin upon chemical reactions in the simulations, we employed the switching-Gō model, where reference structures in the Gō model are switched at certain time steps. Each head has three states characterized by the neck-linker docking (docked or undocked) and by the tubulin binding (on or off): “dock/on”, “undock/on”, and “undock/off”. The reference structure for the “dock/on” state was generated by superimposing the crystal structure for the neck-linker docked dimeric kinesin (PDB ID 3KIN) into the complex structure of single-head kinesin: KIF1A bound on tubulins (PDB ID 2HXF). The energy function for “undock/on” state was constructed just by deleting the native-contact interactions of the neck-linker with others. The energy function for “undock/off” state was made by deleting the interchain native-contact interaction between the kinesin head and tubulins from the “undock/on” state. The parameters used are the default values in CafeMol, except the strengths of the kinesin−tubulin interaction, which were calibrated based on kinesin experiments.

The simulation starts with the phase 1, where the heads A and B are bound to tubulin and take the “dock/on” state and “undock/on” state, respectively. Both heads as well as the

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

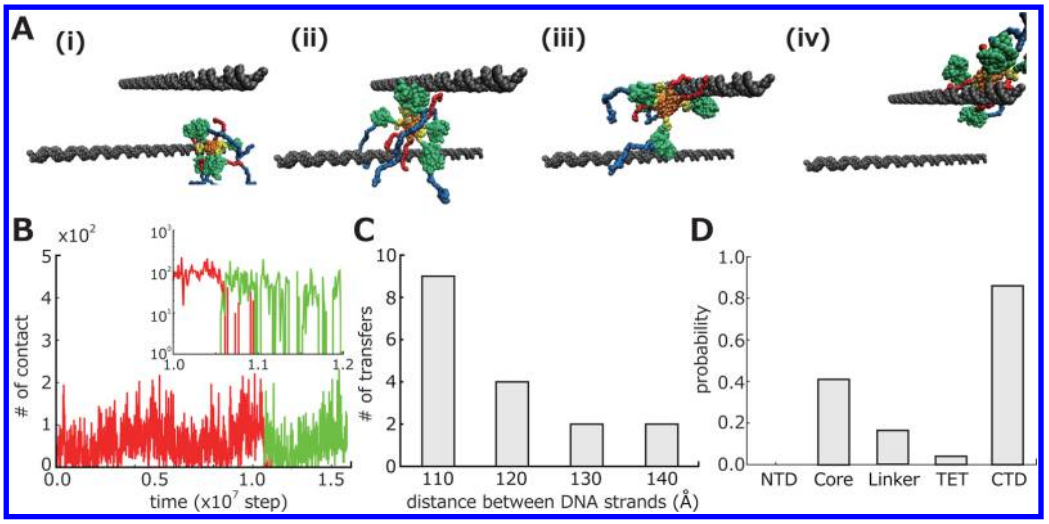


Figure 5. p53 intersegmental transfer realized in CGMD simulations. (A) Representative snapshots from a representative trajectory. A homotetrameric full-length p53 (colored) is simulated on two dsDNA segments that are perpendicularly aligned and frozen (gray). p53 is composed of five distinct domains: disordered N-terminal domain (NTD; blue), core domain (Core; green), flexible linker region (Linker; yellow), tetramerization domain (TET; orange), and disordered C-terminal domain (CTD; red). The p53 initially placed on the lower DNA (i) accomplishes intersegmental transfer to the upper DNA via two intermediate states ii and iii in which p53 binds both of the DNA segments. (B) The number of contacts between p53 CTDs and DNA segments from the same trajectory as panel A (red, contacts with the "initial" segment; green, contacts with the "final" segment). In the inset (log-plot), the transfer event is magnified. (C) The number of transfer events versus the distance between two DNA segments. The number of events decreases as two segments are separated away from each other. (D) Probability of each domain to contact (the minimal distance is less than 10 Å) the final segment in intermediate states.

coiled-coil rapidly fluctuated around their preset positions. We depict a snapshot (Figure 3B) and the time course of *z*-coordinate along which the kinesin proceeds (Figure 3C). After phase 1, we switched the state of the head A to the "undock/off" state. As a result, head A started diffusive motions around the initial position (phase 2, Figure 3B,C, Movie S1). Then, we switched the state of the head B to the "dock/on" state, emulating the ATP binding to the head B. Head A and the coiled-coil region moved forward with large fluctuation (phase 3, Figure 3C). Finally, we switched the state of head A to the "undock/on" state, emulating the ADP release from head A. Head A stochastically moved back and forth between 8 and 16 nm forward regions, before landing on the binding site located at the 16 nm forward position (phase 4, Figure 3C). Although further analysis is necessary, the current simulation illustrates the potential power of CGMD for studying molecular machines.

3.2. Protein–DNA Interactions: Nonspecific Binding of Sso7d to DNA

The second example is nonspecific protein–DNA binding. Nonspecific interactions play important roles in DNA packaging[45] and in maintenance of genomic architecture.[46] Even for many DNA binding proteins that recognize specific target DNA sequences, they also interact nonspecifically with other DNA sequences in their search processes.[47,48] Nonspecific protein–DNA interactions are dominated by electrostatics. Proteins often have positively charged motifs that approach negative phosphate groups of DNA.

Here, we studied the interactions between a small nonspecific DNA-binding protein Sso7d and 12 base-pair (bp) double stranded (ds) DNA (Figure 4A). We employed the AICG2+ model for protein, eq 4,[33] and the 3SPN.2C model for DNA, eq 5,[38] with all the default parameters. We included the excluded volume term and the electrostatic interactions, eq 6, as intermolecular interactions. Note that in the 3SPN.2C DNA model, the phosphate charge within the dsDNA interaction is set to $q_P = -0.6e$, taking the counterion condensation effect into account.[38] As mentioned in section 2.2, an attractive protein–DNA interaction is primed as a short-ranged direct contact, where the counterion may not stay between the molecules. We investigate an optimal value of phosphate charges in the protein–DNA electrostatic interactions.

For this purpose, we performed CGMD simulations in which Sso7d binds to DNA with different phosphate charge values (q_P), ranging from $-0.6e$ to $-1.0e$ (Figure 4C,D). For proteins, we used simple integer charges for amino acids. We then calculated the association constant (K_a) using the Scatchard equation: $r/[L] = (n - r)K_a$ where r is the average number of bound Sso7d per DNA molecule, [L] is the free Sso7d concentration, and n is the number of binding sites on DNA. Comparing with experimental results,[49] we found that $q_P \approx -1.0e$ gives the most accurate description of the Sso7d–DNA electrostatics (Figure 4D). Additionally, we analyzed the DNA contact probability for each amino acid residue in the simulations with $q_P = -1.0e$ (Figure 4B). The residues that contact DNA in the simulation with high probability localized at the protein–DNA interface in the NMR structure.[49]

For better accuracy, one can use partial charges of amino acids determined by the RESPAC method, which remains as a future challenge. Interestingly, for DNA, even if we use the RESPAC protocol, we obtain partial changes of $-1e$ for the phosphate groups and zero for sugars and bases.

3.3. Transcription Factor Dynamics on DNA: Roles of Intrinsically Disordered Regions of p53

The third example is a multifunctional transcription factor, p53, near two perpendicularly aligned dsDNAs. As a transcription factor, p53 needs to scan genomic DNA sequence, while searching its target position. In the search process, it has been suggested that transcription factors sometimes transit from one DNA segment to another in a monkey-bar fashion. Recently,

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026–3035

this intersegmental transfer for p53 was investigated by a CGMD simulation similar to ours.[50] Here, we extended their work further with CafeMol.

In the simulation set, two dsDNA segments are placed in a perpendicular arrangement and frozen for simplicity. A homotetrameric full-length p53 is placed on one DNA segment (Figure 5A(i)) initially. Protein p53 contains three intrinsically disordered regions, the N-terminal domain, the linker, and the C-terminal domain, each of which plays some important biological roles; p53 was modeled with AICG2+, in which domains that lack any reference structures are represented by flexible local potentials. The dsDNA was modeled by 3SPN.1[48] with the default parameters in CafeMol. Interaction between p53 and DNA is purely electrostatics approximated by the Debye–Hückel theory, in addition to the excluded volume term.

During 10^8 time steps in the trajectory, we observed multiple times of intersegmental transfer of p53 (Figure 5A for representative snapshots and Figure 5B for a time series; watch Movie S2). Repeating the simulations with different distances between the two DNA segments, we found that the frequency of the transfer decreases with the distance (Figure 5C). During the sliding along the initial DNA segment, the tetrameric p53 primarily uses its disordered C-terminal tails for the interaction with DNA. Interestingly, our previous study showed that, of the four C-terminal tails, only one-to-three tails interacted with DNA for most of the time near physiological salt concentrations,[48] which means that one-to-three C-terminal tails are free from the initial DNA. In the intermediate state of the intersegmental transfer, p53 grabbed the final DNA segment predominantly with these free C-terminal domains, which is reminiscent of the fly casting mechanism,[51] and subdominantly with its core domain (Figure 5D).

3.4. Chromatin/Nucleosome Structures: Dinucleosome

The last example is on chromatin/nucleosome structures. Eukaryotic genome is stored as hierarchical chromatin structures, of which the basic unit of folding is a nucleosome. A nucleosome consists of a core histone octamer and ∼147 bp dsDNA that are wrapped 1-3/4 turns around the histones. The chromatin structure regulates DNA replication and gene expression. Recently, we performed CGMD simulations of a single nucleosome to investigate the partial and global unwrapping of DNA from the histone and the role of histone tails in the unwrapping.[52] Here, extending that work, we briefly report dinucleosome simulation results.

Histone proteins were modeled by the classic Gō model, eq 2, for the core and by the flexible local potential, eq 3, for the histone tails, while DNA was approximated by 3SPN.1 model. We imposed protein–DNA interactions made of a structure-based contact potential and the Debye–Hückel model, eq 6, for electrostatics. For nucleosomes, we used an X-ray crystal structure (PDB ID 1KX5) as the reference. Two nucleosomes were connected by 25 base pair dsDNA. We performed 10^8 time step simulations at 300 K at the NaCl concentrations 100, 200, and 300 mM, repeating 10 times for each. The obtained structural ensemble contains a broad range of structures.

We plot representative time courses (Figure 6A) and the histogram (Figure 6B) of the distance between two nucleosomes (also watch Movie S3). We found that the distance, on average, increases with the salt concentration since the internucleosome electrostatic attraction mediated by histone tails is weakened by salt. The ensemble at 100 mM

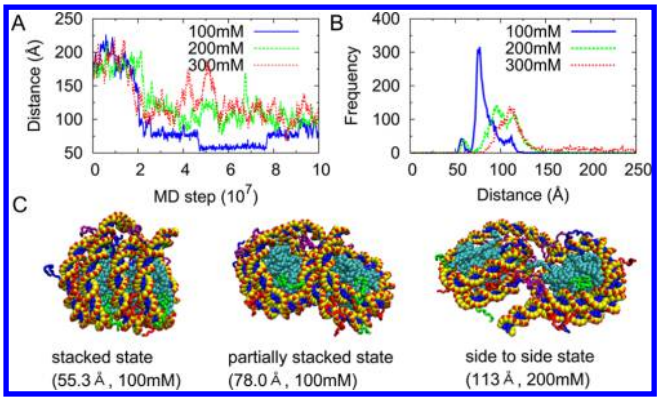


Figure 6. Structural dynamics of dinucleosome. (A) Representative time courses of the distance between the center of mass of two nucleosomes, d_{12}, at varying NaCl concentrations (100 (blue), 200 (green), and 300 (red) mM). (B) Histogram of the distance between two nucleosomes at each salt concentration sampled from the second halves of 10 runs of 10^8 MD steps. (C) Three characteristic structures of dinucleosomes: the stacked state (left), the partially stacked state (center), and the side-to-side state (right). The d_{12} and salt concentration given are in parentheses. Histone cores are in cyan, while histone tails are colored as H2A green, H2B red, H3 purple, and H4 blue. Phosphates, sugars, and bases of DNA are in red, yellow, and blue, respectively.

contained a stacked state (left in Figure 6C) as a minor and a partially stacked state (middle in Figure 6C) as the major populations. Interestingly, the stacked state seems to be stabilized primarily by histone H4 tails sandwiched between two nucleosome cores. At higher salt concentrations (200 and 300 mM), a side-to-side arranged state (right in Figure 6C) becomes a major state.

4. CONCLUDING REMARKS: FUTURE DIRECTIONS

Although useful, our current approach has many limitations that can be overcome in future. Here, we discuss some of the points.

The Debye–Hückel model for electrostatics is a rather poor approximation for strongly charged systems, such as protein–DNA complexes. First, although the theory relies on series expansion for low ionic concentration, the physiological salt concentration is not so low. Second, local counterion concentration around highly charged molecules, which is *a priori* unknown, can be much higher than the bulk ionic concentration, which corresponds to the experimental measurement. Third, the Debye–Hückel model cannot represent solvation by divalent ions, which often has great impact to DNA conformations. One way to improve accuracy of counterion effect is to use explicit ions (still using an implicit water model), which has been exploited in refs 53–55. To account for a heterogeneous dielectric environment around charged molecules, the generalized Born modeling developed in atomistic force fields could be extended to CG models. The generalized Born model is attractive especially when it is combined with the solvent accessible surface area model, the latter of which is known to be very accurate in Cα CG models.[56]

Related to this, one hot issue may be modeling of sequence-specific protein–DNA interactions, given the recent expansion of high throughput data by next-generation sequencing technology, such as protein-binding microarrays (PBMs)[57] and HT-SELEX.[58] These experiments lead to the position–

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026–3035

weight matrix. To take advantage of these big data, we need to develop experimental-data-driven modeling of sequence-specific protein−DNA interactions.

Another missing element in our current modeling is a phospholipid model that matches resolutions of our protein and DNA models, perhaps with five or six CG particles per phospholipid. Although numerous good CG models for membrane have been developed,[5] most of them employ somewhat higher resolution with >10 CG particles per phospholipid. Few others employed lower resolution with approximately three CG particles, which, geometrically, do not approximate amphiphilic interactions between proteins and lipids. Given the molecular architecture of phospholipids, accurately modeling one phospholipid with five or six beads is challenging.

Another promising direction to extend CafeMol is to introduce slightly higher resolution than that used here. In fact, putting one bead to the side chain center is a popular way of coarse graining. For giant molecular complexes, one may employ a hybrid-resolution modeling where some portion of the system uses two beads and another part uses one bead per amino acid.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.accounts.5b00338.

Kinesin stepping trajectory in CGMD simulations (Movie S1) (MPG)

p53 intersegmental transfer realized in CGMD simulations.(Movie S2) (AVI)

Structural dynamics of dinucleosome at NaCl 100 mM (Movie S3) (AVI)

■ AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

Biographies

Shoji Takada received his Ph.D. in 1994 from The Graduate University For Advance Studies under supervision of Prof. Hiroki Nakamura. In 1995−1998, he was a JSPS postdoctoral fellow studying at the group of Prof. Peter Wolynes at University of Illinois at Urbana−Champaign. Then, he became a faculty member at Kobe University. In 2007, he moved to Department of Biophysics, Graduate School of Science, Kyoto University as Associate Professor and was later promoted to Professor. He is a computational biophysicist working on protein folding, molecular machines, chromatin and transcriptional regulation, and so forth.

Ryo Kanada received his Ph.D. in 2003 from Tohoku University under supervision of Prof. Kazuo Sasaki. After working at a company as a system engineer, he spent two years as a postdoctoral researcher at the laboratory of Prof. Makoto Kikuchi at Osaka University. In 2007, he joined the laboratory of Shoji Takada at Kyoto University as a postdoctoral researcher.

Cheng Tan received his Ph.D. in 2014 from Nanjing University under the guidance of Prof. Wei Wang and Wenfei Li. He has started his postdoctoral research in Kyoto University since 2014.

Tsuyoshi Terakawa received his Ph.D. in 2014 from Kyoto University under the guidance of Shoji Takada. Since 2014, he has joined the laboratory of Dr. Eric C. Greene at Columbia University as a research fellow (JSPS). His primary interests are functional mechanisms of protein machineries on DNA from theoretical, computational, and experimental points of view.

Wenfei Li received his Ph.D. in 2004 from Chinese Academy of Science. After postdoctoral work with Prof. Wei Wang at Nanjing University, he became a faculty member in 2006. From 2008 to 2010, he conducted postdoctoral research at Kyoto University with Shoji Takada. He is now Professor of Physics at Nanjing University.

Hiroo Kenzaki received his Ph.D. in Physics from Osaka University in 2007. He spent about five years at Kyoto University as a postdoctoral fellow with Shoji Takada. He has been a technical scientist in RIKEN since 2013. His research interests include the development of coarse-grained model and the simulation of biomolecules.

■ ACKNOWLEDGMENTS

Parts of the simulations were performed by using the RIKEN Integrated Cluster of Clusters (RICC) facility.

■ REFERENCES

(1) The Nobel Prize in Chemistry 2013, Advanced Information Nobelprize.org.

(2) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K. I.; Yao, X. Q.; Takada, S. CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J. Chem. Theory Comput.* **2011**, *7* (6), 1979−1989.

(3) Tozzini, V. Multiscale Modeling of Proteins. *Acc. Chem. Res.* **2010**, *43* (2), 220−230.

(4) Hyeon, C.; Thirumalai, D. Capturing the Essence of Folding and Functions of Biomolecules Using Coarse-Grained Models. *Nat. Commun.* **2011**, *2*, 487.

(5) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (3), 225−248.

(6) Doye, J. P. K.; Ouldridge, T. E.; Louis, A. A.; Romano, F.; Sulc, P.; Matek, C.; Snodin, B. E. K.; Rovigatti, L.; Schreck, J. S.; Harrison, R. M.; Smith, W. P. J. Coarse-Graining DNA for Simulations of DNA Nanotechnology. *Phys. Chem. Chem. Phys.* **2013**, *15* (47), 20395−20414.

(7) Liwo, A.; He, Y.; Scheraga, H. A. Coarse-Grained Force Field: General Folding Theory. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16890−16901.

(8) Ozer, G.; Luque, A.; Schlick, T. The Chromatin Fiber: Multiscale Problems and Approaches. *Curr. Opin. Struct. Biol.* **2015**, *31*, 124−139.

(9) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116* (29), 8494−8503.

(10) Takada, S. Coarse-Grained Molecular Simulations of Large Biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22* (2), 130−137.

(11) Bryngelson, J. D.; Wolynes, P. G. Spin Glasses and the Statistical Mechanics of Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (21), 7524−7528.

(12) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167−195.

(13) Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein Folding Funnels: A Kinetic Approach to the Sequence-Structure Relationship. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (18), 8721−8725.

(14) Clementi, C. Coarse-Grained Models of Protein Folding: Toy Models or Predictive Tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10−15.

(15) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035

(16) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (16), 9319−9323.
(17) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global Ribosome Motions Revealed with Elastic Network Model. *J. Struct. Biol.* **2004**, *147* (3), 302−314.
(18) Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein Structural Change upon Ligand Binding: Linear Response Theory. *Phys. Rev. Lett.* **2005**, *94* (7), 078102.
(19) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. Global Dynamics of Proteins: Bridging between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.
(20) Go, N. Theoretical Studies of Protein Folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.
(21) Takada, S. Go-Ing for the Prediction of Protein Folding Mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (21), 11698−11700.
(22) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and "En-Route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298* (5), 937−953.
(23) Koga, N.; Takada, S. Roles of Native Topology and Chain-Length Scaling in Protein Folding: A Simulation Study with a Go-like Model. *J. Mol. Biol.* **2001**, *313* (1), 171−180.
(24) Karanicolas, J.; Brooks, C. L. Improved G??-like Models Demonstrate the Robustness of Protein Folding Mechanisms towards Non-Native Interactions. *J. Mol. Biol.* **2003**, *334* (2), 309−325.
(25) Koga, N.; Takada, S. Folding-Based Molecular Simulations Reveal Mechanisms of the Rotary Motor F1-ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (14), 5367−5372.
(26) Hoang, T. X.; Cieplak, M. Sequencing of Folding Events in Go-Type Proteins. *J. Chem. Phys.* **2000**, *113* (18), 8319−8328.
(27) Hyeon, C.; Thirumalai, D. Mechanical Unfolding of RNA Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6789−6794.
(28) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. SMOG@ctbp: Simplified Deployment of Structure-Based Models in GROMACS. *Nucleic Acids Res.* **2010**, *38* (SUPPL. 2), W657.
(29) Terakawa, T.; Takada, S. Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: P53 N-Terminal Domain. *Biophys. J.* **2011**, *101* (6), 1450−1458.
(30) Li, W.; Wolynes, P. G.; Takada, S. Frustration, Specific Sequence Dependence, and Nonlinearity in Large-Amplitude Fluctuations of Allosteric Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (9), 3504−3509.
(31) Chu, J.-W.; Izveko, S.; Voth, G. A. The Multiscale Challenge for Biomolecular Systems: Coarse-Grained Modeling. *Mol. Simul.* **2006**, *32*, 211−218.
(32) Li, W.; Terakawa, T.; Wang, W.; Takada, S. Energy Landscape and Multiroute Folding of Topologically Complex Proteins Adenylate Kinase and 2ouf-Knot. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (44), 17789−17794.
(33) Li, W.; Wang, W.; Takada, S. Energy Landscape Views for Interplays among Folding, Binding, and Allostery of Calmodulin Domains. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10550−10555.
(34) Hori, N.; Takada, S. Coarse-Grained Structure-Based Model for RNA-Protein Complexes Developed by Fluctuation Matching. *J. Chem. Theory Comput.* **2012**, *8* (9), 3384−3394.
(35) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. A Coarse Grain Model for DNA. *J. Chem. Phys.* **2007**, *126* (8), 084901.
(36) Sambriski, E. J.; Schwartz, D. C.; De Pablo, J. J. A Mesoscale Model of DNA and Its Renaturation. *Biophys. J.* **2009**, *96* (5), 1675−1690.
(37) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; De Pablo, J. J. An Experimentally-Informed Coarse-Grained 3-Site-per-Nucleotide Model of DNA: Structure, Thermodynamics, and Dynamics of Hybridization. *J. Chem. Phys.* **2013**, *139* (14), 144903.
(38) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; de Pablo, J. J. Coarse-Grained Modeling of DNA Curvature. *J. Chem. Phys.* **2014**, *141* (16), 165103.
(39) Lu, X. J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31* (17), 5108−5121.
(40) Jackson, M. B. *Molecular and Cellular Biophysics*; Cambridge University Press: Cambridge, U.K., 2006.
(41) Stormo, G. D. *Introduction to Protein-DNA Interactions: Structure, Thermodynamics, and Bioinformatics*; Cold Spring Harbor Laboratory Press: New York, 2013.
(42) Terakawa, T.; Takada, S. RESPAC: Method to Determine Partial Charges in Coarse-Grained Protein Model and Its Application to DNA-Binding Proteins. *J. Chem. Theory Comput.* **2014**, *10* (2), 711−721.
(43) Ono, J.; Takada, S.; Saito, S. Couplings between Hierarchical Conformational Dynamics from Multi-Time Correlation Functions and Two-Dimensional Lifetime Spectra: Application to Adenylate Kinase. *J. Chem. Phys.* **2015**, *142*, 212404.
(44) Kanada, R.; Kuwata, T.; Kenzaki, H.; Takada, S. Structure-Based Molecular Simulations Reveal the Enhancement of Biased Brownian Motions in Single-Headed Kinesin. *PLoS Comput. Biol.* **2013**, *9* (2), e1002907.
(45) Luijsterburg, M. S.; Noom, M. C.; Wuite, G. J. L.; Dame, R. T. The Architectural Role of Nucleoid-Associated Proteins in the Organization of Bacterial Chromatin: A Molecular Perspective. *J. Struct. Biol.* **2006**, *156* (2), 262−272.
(46) Murphy, F. V.; Churchill, M. E. Nonsequence-Specific DNA Recognition: A Structural Perspective. *Structure* **2000**, *8* (4), 83−89.
(47) Marcovitz, A.; Levy, Y. Frustration in Protein-DNA Binding Influences Conformational Switching and Target Search Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (44), 17957−17962.
(48) Terakawa, T.; Kenzaki, H.; Takada, S. P53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains. *J. Am. Chem. Soc.* **2012**, *134* (35), 14555−14562.
(49) Agback, P.; Baumann, H.; Knapp, S.; Ladenstein, R.; Härd, T. Architecture of Nonspecific Protein-DNA Interactions in the Sso7d-DNA Complex. *Nat. Struct. Biol.* **1998**, *5* (7), 579−584.
(50) Khazanov, N.; Levy, Y. Sliding of p53 along DNA Can Be Modulated by Its Oligomeric State and by Cross-Talks between Its Constituent Domains. *J. Mol. Biol.* **2011**, *408*, 335−355.
(51) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. Fly-Casting in Protein-DNA Binding: Frustration between Protein Folding and Electrostatics Facilitates Target Recognition. *J. Am. Chem. Soc.* **2007**, *129* (4), 738−739.
(52) Kenzaki, H.; Takada, S. Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLoS Comput. Biol.* **2015**, *11* (8), e1004443.
(53) Savelyev, A.; Papoian, G. A. Chemically Accurate Coarse Graining of. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20340−20345.
(54) Freeman, G. S.; Hinckley, D. M.; Pablo, J. J. De; Freeman, G. S.; Hinckley, D. M.; Pablo, J. J. De A Coarse-Grain Three-Site-per-Nucleotide Model for DNA with Explicit Ions. *J. Chem. Phys.* **2012**, *135*, 165104.
(55) Fan, Y.; Korolev, N.; Lyubartsev, A. P.; Nordenskiöld, L. An Advanced Coarse-Grained Nucleosome Core Particle Model for Computer Simulations of Nucleosome-Nucleosome Interactions under Varying Ionic Conditions. *PLoS One* **2013**, *8* (2), e54228.
(56) Fraternali, F.; Cavallo, L. Parameter Optimized Surfaces (POPS): Analysis of Key Interactions and Conformational Changes in the Ribosome. *Nucleic Acids Res.* **2002**, *30* (13), 2950−2960.
(57) Berger, M. F.; Philippakis, A. A.; Qureshi, A. M.; He, F. S.; Estep, P. W.; Bulyk, M. L. Compact, Universal DNA Microarrays to Comprehensively Determine Transcription-Factor Binding Site Specificities. *Nat. Biotechnol.* **2006**, *24* (11), 1429−1435.
(58) Zhao, Y.; Granas, D.; Stormo, G. D. Inferring Binding Energies from Selected Binding Sites. *PLoS Comput. Biol.* **2009**, *5* (12), e1000590.

DOI: 10.1021/acs.accounts.5b00338
Acc. Chem. Res. 2015, 48, 3026−3035