

# Coarse-Grained Structure-Based Model for RNA-Protein Complexes Developed by Fluctuation Matching

Naoto Hori<sup>†</sup> and Shoji Takada<sup>\*,†,‡</sup>

<sup>†</sup>Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto, Japan

<sup>‡</sup>Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Japan

## Supporting Information

**ABSTRACT:** RNA and RNA–protein complexes have recently been intensively studied in experiments, but the corresponding molecular simulation work is much less abundant, primarily due to its large system size and the long time scale involved. Here, to overcome these bottlenecks, we develop a coarse-grained (CG) structure-based simulation model for RNA and RNA–protein complexes and test it for several molecular systems. The CG model for RNA contains three particles per nucleotide, each for phosphate, sugar, and a base. Focusing on RNA molecules that fold to well-defined native structures, we employed a structure-based potential, which is similar to the Go-like potential successfully used in CG modeling of proteins. In addition, we tested three means to approximate electrostatic interactions. Many parameters involved in the CG potential were determined via a multiscale method: We matched the native fluctuation of the CG model with that by all-atom simulations for 16 RNA molecules and 10 RNA–protein complexes, from which we derived a generic set of CG parameters. We show that the derived parameters can reproduce native fluctuations well for four RNA and two RNA–protein complexes. For tRNA, the native fluctuation in solution includes large-amplitude motions that reach conformations nearly corresponding to the hybrid state P/E and EF-Tu-bound state A/T seen in the complexes with ribosome. Finally, large-amplitude modes of ribosome are briefly described.

## 1. INTRODUCTION

Recently, various functional RNA molecules and RNA–protein complexes have been focused upon and studied experimentally due to their essential functions, especially related to genetic mechanisms such as transcription and translation.<sup>1–3</sup> To complement such experiments and to provide much time-dependent structural information, atomistic molecular dynamics (MD) simulations are highly desired, but they have been seldom or just partly applied to RNA–protein complexes primarily because of their large system size and the long time scale involved. Coarse-graining provides one solution to the problems.<sup>4,5</sup> Indeed, coarse-grained (CG) MD simulations have been developed and used for RNA as well as for proteins.<sup>6–23</sup> However, since CG models for proteins and RNAs were developed somewhat separately, CG models for RNA–protein complexes are less developed at the moment. Here, employing a consistent representation for RNA and proteins, we present a new CG model that can be used for RNA–protein complexes.

For proteins, CG molecular models have long been used to investigate long-time protein dynamics such as folding and conformational change related to allosteric regulation. In such coarse-graining, the polypeptide has often been treated as a chain of beads in which one amino acid corresponds to one to a few beads. With these simplified representations that lack the side-chain atomic view, physicochemical interactions alone cannot easily make the native state sufficiently stable. As an alternative, it has been popular to use structure-based models where the potential energy is a function of the native structure making the native state sufficiently stable. This is often called the Go-like model<sup>24</sup> or the topology-based model and is closely linked to energy landscape theory, which tells us the energy landscape of proteins having globally funnel-like shape.<sup>25</sup> Go-

like models have been widely applied, showing their promise.<sup>26–28</sup>

Coarse-graining techniques have been successful also in the field of nucleic acid simulations. In contrast to DNA molecules that are mostly present as the double-stranded helix, RNA molecules are mostly single-stranded, and some of them form compact folded tertiary structures and sometimes work even as enzymes. In this sense, applying a structure-based model to these RNA molecules would be a reasonable choice in the same way as in the case of proteins. Hyeon and Thirumalai have developed a structure-based CG model, referred to as the self-organized polymer model, which reproduces well the folding behavior and the experimental force–extension curves of short hairpin RNA.<sup>9,13</sup> Paliy et al. examined a series of CG models and applied them to the dynamics of an RNA nanoring structure. They concluded that the three-site-per-nucleotide representation is suitable to capture the dynamics precisely.<sup>22</sup> There are also many other CG models of RNA primarily intended for the prediction of the tertiary structure at several different resolutions.<sup>14,17,18,23</sup> We also mention that the structure-based approach is used even in the all-atom representation of RNA and their complexes with proteins, where, although their resolution is not coarse-grained, their interaction is coarse-grained, making the energy landscape smoother.<sup>29,30</sup> They, however, still need large-scale computational resources to investigate large-amplitude motions in ribosomes.

In the development of CG models in general, a major task is to determine parameters involved in the CG potential energy

Received: May 8, 2012

Published: July 25, 2012



function as accurately and systematically as possible. In some earlier structure-based potentials, parameters were defined quite empirically, and their parameters were validated by the agreement of the simulation results with experiments. More recent studies put forward multiscale methods to determine CG parameters in a more systematic manner from all-atom (AA) models.<sup>31</sup> For example, Voth and co-workers have developed systematic methods to determine parameters using AA simulations so that the CG parameters become more realistic and based on more physicochemical aspects.<sup>32,33</sup> One such method is the fluctuation matching method in which first AA-MD simulations are performed, and then that information is used to determine CG parameters by matching mean-square fluctuations through iterative CG-MD simulations.<sup>32</sup> In the current work, we employ the fluctuation matching method to derive CG parameters for generic RNA–protein complexes.

Given that CG models for RNA–protein complexes are not well developed, except for a specific target, ribosomes,<sup>7,8,11,12,15,21,34</sup> our purpose here is to construct a generic CG model for RNA and RNA–protein complexes and then to derive its CG parameters via the multiscale approach. In the model, each nucleotide is represented as three CG beads, each bead representing phosphate, sugar, and base, and interactions are modeled primarily with the structure-based potential. Because RNA molecules are highly charged, the electrostatic effect is important. To take it into account, we tested some different treatments of electrostatics in this study. One is to express the electrostatic interaction explicitly in the potential function, and the other is to assume that the electrostatic interaction implicitly inheres in the structure-based potential. For each of the treatments, to obtain CG parameters, we used the fluctuation matching method proposed by Voth and co-workers,<sup>32</sup> which has already been applied to a protein CG model of our group.<sup>35</sup> First, we performed 20 ns AA MD simulations for a total of 26 molecules and used those data to derive CG parameters. To confirm the accuracy and availability, we first tested the native-state dynamics comparing fluctuations around native structures and also compared dominant movements in the fluctuation of tRNA and ribosomes. The current RNA model and the derived parameters have been implemented in our CG-MD simulation package, CafeMol,<sup>36</sup> and thus are publicly available via the web page, [www.cafemol.org](http://www.cafemol.org).

## 2. MODEL AND METHODS

Here, we describe our CG model for RNA–protein complexes. We assume that our current target molecules are well-folded to particular structures, of which atomic coordinates are characterized experimentally. Given that the molecules fold to particular structures under physiological conditions, it is reasonable that their CG potential energy function contains structure-based terms.<sup>4,5,29</sup>

Among a few possible levels of coarse-graining, here we use one bead per an amino acid for proteins and three beads per a nucleotide for RNA. For each of the proteins<sup>26,27</sup> and RNA,<sup>9,13,17,22</sup> these levels of coarse-graining have been successful. For one thing, densities of their CG particles are roughly similar, which is advantageous in representing their interactions. Moreover, this level of modeling is simple enough to simulate long-time dynamics of giant complexes such as the ribosome, as is shown in this study.

Our potential energy function for RNA–protein complexes is composed of a structure-based part and an electrostatic contribution:

$$V_{\text{total}} = V_{\text{SB}} + V_{\text{ele}} \quad (1)$$

The structure-based potential can be further divided into three contributions:

$$V_{\text{SB}} = V_{\text{protein}} + V_{\text{RNA}} + V_{\text{pro-RNA}} \quad (2)$$

In the following, we first present each of the three structure-based terms, which is followed by an explanation of the electrostatic interaction. Then, details of computational methods are described in the latter part of this section.

**2.1. Proteins.** For proteins, we employ the off-lattice Go model proposed by Clementi et al.,<sup>26</sup> with multiscale parametrization from Li et al.<sup>35</sup> Protein molecules are represented as connected CG particles, of which one particle located on the C<sub>α</sub> atom corresponds to one amino acid. The potential is composed of local, contact, and excluded volume terms.

$$V_{\text{protein}} = V_{\text{local}} + V_{\text{contact}} + V_{\text{excluded}} \quad (3)$$

The local term,  $V_{\text{local}}$ , is the summation of potentials for virtual bond lengths, bond angles, and dihedral angles,

$$\begin{aligned} V_{\text{local}} = & \sum_{\text{ibd}} K_b(r_{\text{ibd}} - r_{\text{ibd}}^0)^2 + \sum_{\text{iba}} K_\theta(\theta_{\text{iba}} - \theta_{\text{iba}}^0)^2 \\ & + \sum_{\text{idih}} \left\{ K_\phi [1 - \cos(\phi_{\text{idih}} - \phi_{\text{idih}}^0)] \right. \\ & \left. + \frac{1}{2} K_{\phi 3} [1 - \cos 3(\phi_{\text{idih}} - \phi_{\text{idih}}^0)] \right\} \end{aligned} \quad (4)$$

where  $r_{\text{ibd}}$  is the length of ibd-th virtual bond connecting ibd-th and (ibd + 1)-th CG particles;  $\theta_{\text{iba}}$  is the iba-th virtual bond angle defined by iba-th, (iba + 1)-th, and (iba + 2)-th CG particles; and  $\phi_{\text{idih}}$  is the idih-th dihedral angle defined by idih-th, (idih + 1)-th, (idih + 2)-th, and (idih + 3)-th CG particles. Here and hereafter, the superscript “0” indicates the corresponding values in the native structure (in most cases, a PDB structure). All of the constants  $K$  are parameters that modulate local stiffness of the protein.

Next,  $V_{\text{contact}}$  is the nonlocal contact potential, essential for structure-based modeling, which is defined as

$$V_{\text{contact}} = \sum_{i < j-3}^{\text{nat-con}} \varepsilon_{\text{go}} \left[ 5 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{10} \right] \quad (5)$$

where  $r_{ij}$  is the distance between amino acids  $i$  and  $j$ , and the summation is taken only for natively contacting pairs (nat-con). Any residue pair is considered to be a natively contacting pair if those two residues are separated by more than three residues in sequence and the distances of the closest heavy atoms between them are within a cutoff distance in the native structure. In our model, the cutoff distance is chosen as 6.5 Å.

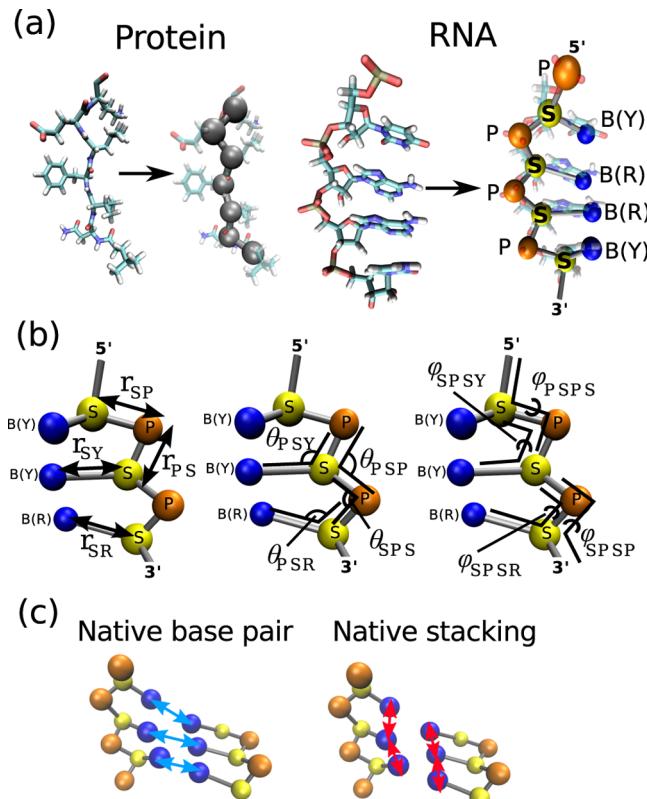
The third term,  $V_{\text{excluded}}$ , is a simple excluded volume term, applied for the non-native pairs:

$$V_{\text{excluded}} = \sum_{i < j-3}^{\text{non-native}} \varepsilon_{\text{ex}} \left( \frac{d}{r_{ij}} \right)^{12} \quad (6)$$

where  $d$  is 4.0 Å.

All of the coefficient parameters,  $K_b$ ,  $K_\theta$ ,  $K_\phi$ ,  $\varepsilon_{go}$ , and  $\varepsilon_{ex}$  of the protein were determined by Li et al. so that CG MD-based fluctuations match with those estimated by all-atom MD simulations with the AMBER force field.<sup>35</sup> The unit of energy is kcal/mol.

**2.2. RNA: Structure-Based Terms.** Our CG RNA model uses three particles per one nucleotide, each representing phosphate (P), sugar (S), and base (B) (Figure 1a). This three-



**Figure 1.** Coarse-grained representation. (a) Comparison of all-atom and coarse-grained representation for protein and RNA. (b) Local interactions of RNA, virtual bond length, bond angle, and dihedral angle. (c) Two types of RNA-specific nonlocal interactions, base pair and stacking. All of the molecular images in this article were created with VMD.<sup>37</sup>

beads-per-nucleotide representation has been successfully used in previous CG simulations of both RNA<sup>9,17,22</sup> and DNA.<sup>38</sup> The scale is thought to be a reasonable partition of RNA molecules in terms of chemical properties. It is, additionally, roughly consistent with the coarse-graining scale of our protein model (Figure 1a; CG particle density for RNA is slightly higher than that for protein, though). The CG particles of phosphates are placed at the positions of the phosphorus atom of phosphate groups. The CG particles of sugars are placed at the centers of sugar-ring atoms, and the CG particles of bases are placed at positions of N1 atoms for purine–base and N3 atoms for pyrimidine–base.

In a similar way to the CG potential for proteins, we write the potential energy function for RNA using the three terms:

$$V_{\text{RNA}} = V_{\text{local}} + V_{\text{contact}} + V_{\text{excluded}} \quad (7)$$

The local potential is conceptually the same as that for the protein expressed as

$$\begin{aligned} V_{\text{local}} = & \sum_{\eta(\text{ibd}) \in \{\text{PS,SP,SR,SY}\}} K_\eta^b (r_{\text{ibd}} - r_{\text{ibd}}^0)^2 \\ & + \sum_{\eta(\text{iba}) \in \{\text{PSP,SPS,SPR,SPY}\}} K_\theta^\eta (\theta_{\text{iba}} - \theta_{\text{iba}}^0)^2 \\ & + \sum_{\eta(\text{idih}) \in \{\text{PSPS,SPSP,SPSR,SPSY}\}} \left\{ K_\phi^\eta [1 - \cos(\phi_{\text{idih}} - \phi_{\text{idih}}^0)] \right. \\ & \left. + \frac{1}{2} K_\phi^\eta [1 - \cos 3(\phi_{\text{idih}} - \phi_{\text{idih}}^0)] \right\} \end{aligned} \quad (8)$$

From a preliminary statistical analysis of local structures of RNA using RNADB2005,<sup>39</sup> we found that local structures can be well classified when we treat bases as being of two types, purine (R) and pyrimidine (Y). (See Figure S1 in the Supporting Information.) Thus, base beads are denoted as either R or Y. Although it might be better to discriminate four types of bases in some cases, we consider that such specificity can be included implicitly by the structure-based potential because the interactions are defined using all of the heavy atoms. In the first term that represents the virtual bond length,  $\eta$  is one of PS, SP, SR, and SY (Figure 1b). In the second term,  $\eta$  represents one of PSP, SPS, PSR, and PSY. In the fourth term for dihedral angles,  $\eta$  is one of PSPS, SPSP, SPSR, and SPSY. For each of the types, we assign one coefficient parameter, which is determined by the fluctuation matching in the same way as the case of proteins. We note that, although we could, we did not include the virtual bond angles defined by  $B(i)-S(i)-P(i+1)$ , i.e., RSP or YSP, and the dihedral angles defined by  $B(i)-S(i)-P(i+1)-S(i+1)$ , i.e. RSPS or YSPS. Without including them, we could fit the all-atom MD-based fluctuation equally well, and we consider that, given the similar approximation of fluctuations, using fewer terms is better.

For the nonlocal contact potential of RNA molecules, we employ the same functional form as that of proteins but take a somewhat more elaborate scheme of parametrization:

$$V_{\text{contact}} = \sum_{i,j \in \{\text{S,B}\}}^{\text{native contact}} \varepsilon_\xi \left[ 5 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{10} \right] \quad (9)$$

where local pairs that can be connected by no more than three virtual bonds are excluded in the summation. First, the contact potential is introduced only for sugar and base particles but not for phosphate particles because the latter is much smaller. In contrast to proteins, nucleic acids have clearly distinct contact interactions which may play distinct roles in forming typical single- or double-stranded helix. Concretely, we treat base pair and base-stacking interactions distinctively from other and more general contact interactions using different coefficients. Namely, we classified natively contacting pairs into the three types: base pair, stacking, and others. In our model, two residues are considered to make the base pair interaction if two or more hydrogen bonds are formed between them in the native structure. The interaction type is further discriminated by the number of hydrogen bonds, two ( $\varepsilon_\xi = \varepsilon_{\text{BP2}}$ ) or three and more ( $\varepsilon_\xi = \varepsilon_{\text{BP3}}$ ), because the interaction strength would depend on the number of formed hydrogen bonds. Base stacking is another important feature ( $\varepsilon_\xi = \varepsilon_{\text{ST}}$ ). To exclude the flipped base from this interaction, we define base stacking

under the following two conditions, which were derived from a preliminary statistical survey (see Figure S2 in the Supporting Information): (1) The distance of the nearest neighbor atoms of them is less than 6.0 Å in the native structure. (2) The virtual dihedral angle formed by  $B(i)-S(i)-S(i+1)-B(i+1)$  is smaller than 40° in the native structure. If a natively interacting pair is not classified as the base pair or the base-stacking pair, this contact formed by the base and sugar particles is defined as a general “contact” interaction ( $\epsilon_{\xi} = \epsilon_{\text{con}}^{\eta}$  where  $\eta$  is either S–S, S–B, or B–B). The contact detection is atomic-based in the same way as that for proteins. The cutoff distance is 5.5 Å.

Nonlocal CG particle pairs that have no native contact interactions have simple excluded-volume interactions:

$$V_{\text{excluded}} = \sum_{\text{non-native}} \epsilon_{\text{ex}} \left( \frac{d}{r_{ij}} \right)^{12} \quad (10)$$

where  $d$  is 4.0 Å. The coefficient,  $\epsilon_{\text{ex}}$ , is set to 0.2 kcal/mol in the same manner as the protein model.

**2.3. RNA–Protein Interaction: Structure-Based Term.** Between protein and RNA molecules that form stable complexes, we include the same structure-based potential as those for proteins and RNA. Although there are three possible interaction types (i.e., pro–P, pro–S, and pro–B), we did not include pro–P interactions. The phosphate moiety of the RNA backbone is markedly smaller than those of the sugar, base, and amino acid. Thus, it may be better balanced if only pro–S and pro–B interactions are included in the structure-based potential. In addition, a statistical study has shown that pro–P interactions are not dominant in RNA–protein interactions, unlike the situation in DNA–protein interactions.<sup>40</sup> Thus, the potential is defined as

$$V_{\text{pro-RNA}} = \sum_{\substack{i \in \text{pro} \\ j \in S, B}} \epsilon_{\text{con}}^{\eta} \left[ 5 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{10} \right] \quad (11)$$

where  $i$  and  $j$  represent the CG particles in proteins and RNA, respectively, and  $\eta$  is either pro–S or pro–B. Nonlocal CG particle pairs have excluded-volume interactions expressed in 10, as usual.

**2.4. Electrostatic Interactions.** RNA molecules are highly negatively charged polyelectrolytes, and their structures and dynamics strongly depend on the solvent ion concentration, clearly suggesting an important contribution of the electrostatic interactions (see Figure S4 for distribution of charges in the ribosome, as an example). How to treat the electrostatic interactions in CG modeling of RNA–protein complex is a major and nontrivial issue. Although many treatments were proposed, there is no single and established way.<sup>41</sup> As a polyelectrolyte, nucleic acids must be surrounded by counterions.<sup>42</sup> Especially in RNA, the divalent ion plays crucial roles for its folding.<sup>43</sup> Yet, in CG modeling we rely on structure-based potentials which, in principle, can take into account the short-ranged part of electrostatic interactions. Thus, straightforward use of the electrostatic interactions as well as structure-based potentials may overestimate the interactions.

In this study, we take simple approaches and compare three different models varying in the treatment of electrostatic interactions.

**Model A.** The electrostatic interaction is explicitly modeled for all interactions of intra-RNA and between RNA and protein.

Every phosphate particle in RNA and negatively charged amino acid (Asp and Glu) has one negative charge, and every positively charged amino acid in proteins (Arg, His, and Lys) has one positive charge. All of these charged particles contribute to the electrostatic interaction.

**Model B.** The electrostatic interaction is not explicitly treated. This is purely a structure-based model. In this model, we assume that all physicochemical effects regarding electrostatic interactions are effectively included in the structure-based potential.

**Model C.** A hybrid model of models A and B. In this model, we assume that intramolecule interactions can be described by the structure-based potential completely, but for intermolecular interactions between RNA and the protein, the electrostatic term is explicitly included as in model A.

For the models A and C, the electrostatic potential is expressed by the Debye–Hückel model.

$$V_{\text{ele}} = \sum_{i,j} \frac{z_i z_j e^2}{4\pi\epsilon_0\epsilon_r r_{ij}} \exp\left(-\frac{r_{ij}}{l_D}\right) \quad (12)$$

$$l_D = \sqrt{\frac{\epsilon_0\epsilon_r k_B T}{2N_A e^2 I}} \quad (13)$$

where  $z$  is number of charge,  $e$  is the elementary electric charge,  $\epsilon_0$  is electric constant,  $\epsilon_r$  is dielectric constant,  $l_D$  is the so-called Debye length,  $N_A$  is Avogadro's number,  $k_B$  is the Boltzmann constant, and  $I$  is the ionic strength. In this study, we set the dielectric constant and the ionic strength to 78.0 and 0.15, respectively (corresponding to an aqueous solution of 150 mM NaCl). As described above, the summation is for all charged-particle pairs excepting intraprotein in model A whereas for only between phosphates and charged amino acids in the model C. Note that we do not include the contact (structure-based) energy between a phosphate and an amino acid. Thus, the two energy terms, structure-based and electrostatic, never work simultaneously between two of any particles.

**2.5. Target Molecules.** Target molecules for the fluctuation matching calculation are shown in the tables in the Supporting Information. Six canonical A-type double-helical structures are constructed with the NAB tool<sup>44</sup> of the AMBER11 suite.<sup>45</sup> Sequences are chosen so that various nucleotide-type dependencies are incorporated and all of the lengths are 20 bp. The other 10 structures are taken from the Protein Data Bank (PDB),<sup>46</sup> of which three are NMR<sup>47–49</sup> and the others are X-ray structures.<sup>50–56</sup> Some of them are typical small motifs, and others are some sort of functional RNA which contains not only typical double-helical structures but also various loops, kinks, turns, and many tertiary contacts. The sizes of RNA molecules are distributed from 12 nt to 398 nt. The 10 RNA–protein complex structures are also taken from the PDB. Of the 10, two are tRNA molecules bound to tRNA synthetase,<sup>57,58</sup> four are ribosome-related complexes,<sup>59–62</sup> and the others are various RNA-binding proteins.<sup>63–66</sup> In this manner, our modeling procedure attempted to guarantee the generality of the obtained parameter set.

For ribosome simulations, we used two structures reported by X-ray crystallography: one is PDB codes 2j00 and 2j01, which contain three tRNA molecules bound to tRNA-binding sites,<sup>67</sup> and the other is PDB codes 3i8h and 3i8i, which is a more recent structure. We removed tRNA molecules from the

latter complex to simulate the motion of the intrinsic ribosome.<sup>68</sup>

**2.6. MD Simulations.** All of the AA MD simulations were performed with the AMBER11 software package.<sup>45</sup> The parameter set ff99bsc0 was used as a force field for standard protein and RNA molecules.<sup>69</sup> Solvent molecules were explicitly included in the systems using the TIP3P model and ionic parameters developed by Åqvist.<sup>70</sup> For modified nucleic acids, we used Aduri et al.'s parameters.<sup>71</sup> When an original PDB entry had divalent cations, we retained them. Then, we added monovalent ions, Na<sup>+</sup> and Cl<sup>-</sup>, so that the system was neutralized and the ion concentration corresponded to 150 mM NaCl. Generally speaking, RNA molecules are less stable than proteins so that a more cautious equilibration process is needed. We followed a equilibration protocol suggested by Auffinger et al.<sup>72</sup> Simulations were performed under the NPT ensemble in which the temperature and the pressure were controlled by a Berendsen thermostat to maintain 300 K and 1.0 atm, with a coupling constant of 0.2 and 0.5 ps, respectively. Finally 20 ns production runs were performed.

The CG MD simulations were performed with the CafeMol package.<sup>36</sup> Langevin dynamics were employed with the temperature set as 300 K.

**2.7. Fluctuation Matching.** First, from AA-MD trajectories, we extracted sets of coordinates of CG particles, which is followed by the calculation of the mean square fluctuation (MSF) values for CG variables, such as virtual bonds, virtual bond angles, and so on.

$$\text{MSF}_{\chi}^{\text{AA}} = \langle X_{\chi}^2 - \langle X_{\chi} \rangle_{\text{AA}}^2 \rangle_{\text{AA}} \quad (14)$$

where  $\chi$  represents a particular CG interaction term,  $X_{\chi}$  is a instantaneous value of the interaction term, and  $\langle \dots \rangle_{\text{AA}}$  means the average over 20 ns AA-MD snapshots.

At each ( $I$ th) generation of the parameter learning, with a previously obtained set of CG parameters, we performed CG-MD. Given AA-MD based MSF,  $\text{MSF}_{\chi}^{\text{AA}}$ , and  $I$ th generation CG-MD based MSF,  $\text{MSF}_{\chi}^{\text{CG},(I)}$ , we update the corresponding parameter,  $k_{\chi}$  for  $\chi$  parameters as

$$k_{\chi}^{(I+1)} = k_{\chi}^{(I)} \times \left[ 1 + \left( 1 - \frac{\sum_{\chi} \text{MSF}_{\chi}^{\text{AA}}}{\sum_{\chi} \text{MSF}_{\chi}^{\text{CG},(I)}} \right) \times f \right] \quad (15)$$

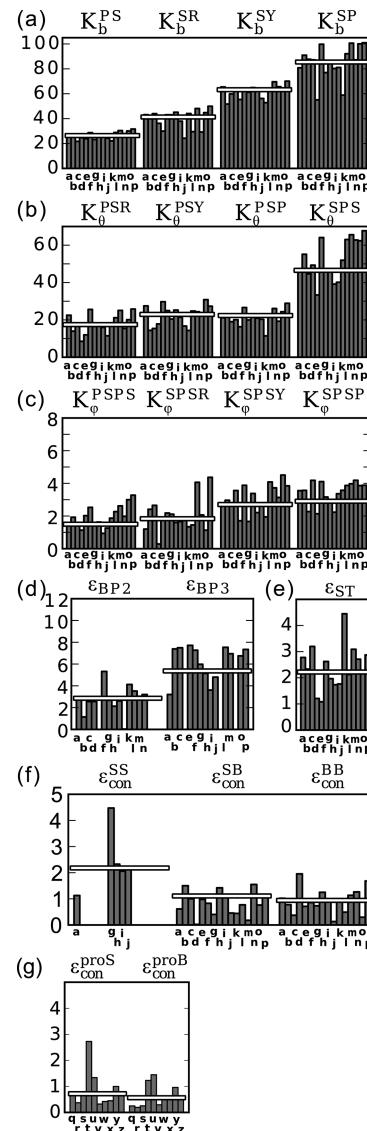
$$k_{\chi} = K_b^{\text{PS}}, K_b^{\text{SP}}, \dots, \varepsilon_{\text{BP}2}, \varepsilon_{\text{BP}3}, \varepsilon_{\text{st}} \quad (16)$$

where  $f$  is the damping factor set as 0.2 for the virtual bond length and the virtual bond angle, 0.15 for the dihedral angle, 0.09 for the base pair and the base stacking, and 0.1 for other nonlocal contacts. In the learning procedure,  $2 \times 10^7$  step Langevin dynamics CG simulations were performed for each iteration step, and a total of 100 iterations were completed.

### 3. RESULTS AND DISCUSSIONS

**3.1. Multiscale Parameterization.** To perform the parameterization, we first chose 16 RNA molecules and 10 RNA–protein complexes as a training set. For each of them, after equilibration, we performed 20 ns all-atom (AA) MD simulations (see Model and Methods for the MD details). From these trajectories, we calculated mean-square-fluctuation values for each term of the CG potential energy function. Next, the fluctuation matching was performed in two stages. First, only RNA-related parameters (all coefficients excepting  $\varepsilon_{\text{con}}^{\text{pro-S}}$  and  $\varepsilon_{\text{con}}^{\text{pro-B}}$ ) were fitted using the 16 RNA targets, and then

RNA–protein interactions ( $\varepsilon_{\text{con}}^{\text{pro-S}}$  and  $\varepsilon_{\text{con}}^{\text{pro-B}}$ ) were optimized using 10 complexes with fixed RNA-related parameters. For each target molecule, 100 iterative CG simulations were enough to get converged parameter values. The obtained parameters for model A are shown in Figure 2 (for models B and C, see



**Figure 2.** Obtained parameters in the case of model A. Parameters obtained by the fluctuation matching are shown for (a) virtual bond length, (b) virtual bond angle, (c) dihedral angle, (d) base-pairing, (e) base-stack, (f) intra-RNA native contact, and (g) native contact between the protein and RNA. Horizontal white bars indicate weight-averaged values which are used as the general parameter set. Alphabets along horizontal axes are the target IDs listed in Tables S1 and S2 in the Supporting Information. The unit of energy is kcal/mol.

Figure S3 in the Supporting Information). In some training molecules, some potential terms, such as  $\varepsilon_{\text{BP}2}$  of molecules  $e$ ,  $j$ ,  $o$ , and  $p$ , did not appear, and thus these parameters were not fitted. In some other cases, some potential terms appeared very few times, and the fitted values of these parameters are less robust.

With these in mind, we see in Figure 2 that the variability of each parameter among different molecules in the training set was overall relatively small, suggesting a robust nature of the

Table 1. Parameters Obtained by the Fluctuation Matching

		model		(kcal/mol)	model		
		A	B, C		A	B	C
bond length	$K_b^{PS}$	26.5	26.5	stacking	$\epsilon_{ST}$	2.18	2.06
	$K_b^{SR}$	40.3	40.3		$\epsilon_{BP2}$	2.83	2.94
	$K_b^{SY}$	62.9	62.9	base pair	$\epsilon_{BP3}^{S-S}$	5.33	5.37
	$K_b^{SP}$	84.1	84.1		$\epsilon_{con}^{S-B}$	2.13	1.48
bond angle	$K_\theta^{PSR}$	17.8	18.0	contact	$\epsilon_{con}^{B-B}$	1.06	0.98
	$K_\theta^{PSY}$	22.6	22.8		$\epsilon_{con}^{pro-S}$	0.97	0.93
	$K_\theta^{PSP}$	21.5	22.1		$\epsilon_{con}^{pro-B}$	0.72	0.73
	$K_\theta^{SPS}$	47.4	47.8			0.58	0.52
dihedral angle	$K_\phi^{SPS}$	1.53	1.64				0.74
	$K_\phi^{SPSR}$	1.87	1.88				0.62
	$K_\phi^{SPSY}$	2.78	2.82				
	$K_\phi^{SPSP}$	2.96	2.98				

current fluctuation matching scheme. More specifically, parameters in the local interaction terms, i.e., the virtual bond length, the virtual bond angle, and the dihedral angle terms, were very similar for all of the molecules. For the nonlocal interaction terms, the fitted values were somewhat more diverse among the molecules. To obtain one generic set of parameters that can be used for any other RNA–protein complexes, we averaged the fitted parameter values over all molecules in the training set with the weight factors proportional to the number of corresponding terms in each molecule.

The parameter values averaged over the training molecules are presented in Table 1 for all three models. Comparing among the three models, we see that most fitted parameters are quite similar to each other. For example, the bond length parameters are identical. This indicates that the existence of the electrostatic term does not affect most local and nonlocal parameters in these models. Only the contact interaction parameters between sugar particles,  $\epsilon_{con}^{S-S}$ , are significantly different; it is a larger value in model A than in the other models. Model A contains the electrostatic interaction explicitly. Since, in model A, the explicit electrostatics makes two segments of RNA repulsive, this effect is probably augmented by introducing stronger sugar–sugar attraction in the structure-based potential.

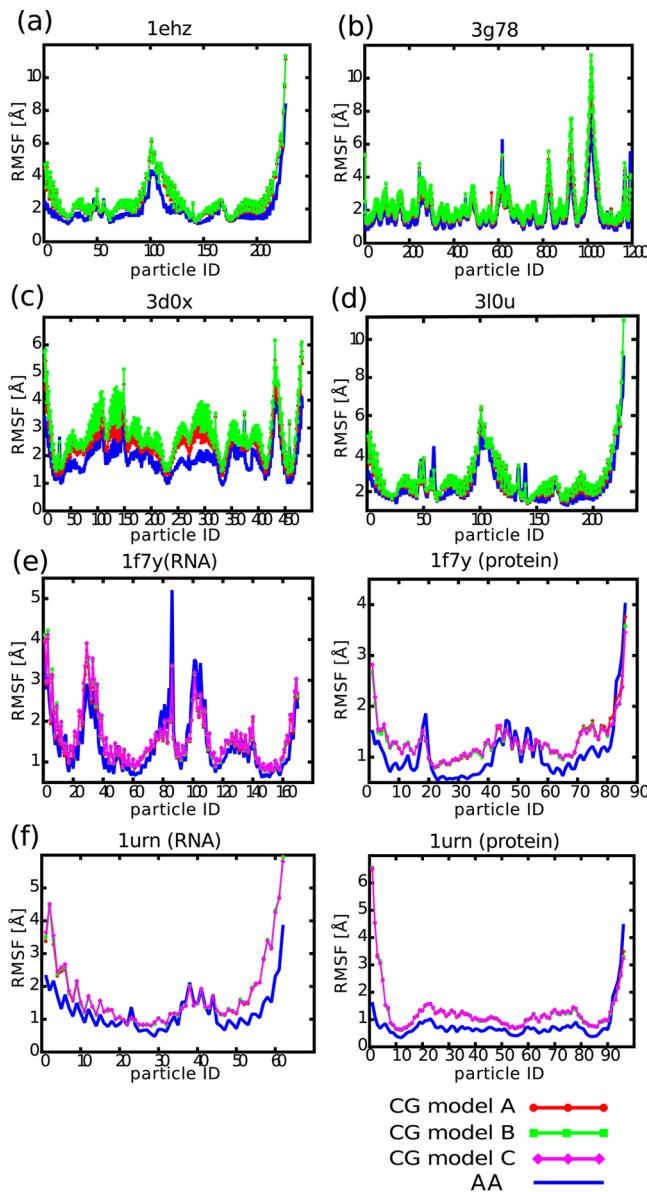
In general, one may think of the base stacking as providing more stabilization than the base pairing.<sup>73</sup> Then, one might think it strange that the base-pairing coefficients ( $\epsilon_{BP2}$  and  $\epsilon_{BP3}$ ) have larger values than the stacking ( $\epsilon_{ST}$ ) in the model parameters (Table 1). In the CG model, however, base-stacking interactions are present separately within each of the RNA strands. This means that, in a double helical region of RNA, one base pair formation contributes one base paring and two base stackings, since both of the strands have stacking interactions. With this in mind, we can compare those energy contributions as follows. In model A, for instance,  $\epsilon_{BP2}$  is 2.83 kcal/mol (for the A–U pair),  $\epsilon_{BP3}$  is 5.33 kcal/mol (for the G–C pair), and the stacking energy  $2 \times \epsilon_{ST}$  is 4.36 kcal/mol. Then, the stacking energy is comparable to the contribution from the base pairing, at least larger than the base pair energy for the A–U pair.

**3.2. Validation of the Parameters by RMSF.** To validate the model, we first compared fluctuations around native structures calculated from CG-model simulations and those from AA simulations with an AMBER force field. Specifically, we compared the root-mean-square fluctuation (RMSF) calculated from CG and AA simulations for some molecules.

When comparing the RMSF of a molecule in the training set, we used a jackknife test-like protocol. Namely, for each of the targets, we reaveraged each parameter value over the training set molecules excluding the one being compared. Figure 3a–c show the comparison for RNA included in the training set. The results show that the RMSF values of models A and B are almost identical to each other, and also they are in good agreement with the RMSF from the AA simulation. In the case of 3d0x, the CG results show a slightly larger fluctuation than that of AA simulations. In Figure 3d, we compare the RMSFs for tRNA<sup>Phe</sup>, which did not belong to the training set. The figure shows also good agreement with the RMSF of the AA simulation. In Figure 3e and f, the RMSFs for the two RNA–protein complexes are shown. In both complexes, RMSF values of both RNA and protein are nearly identical among the three models. These results indicate that the parameters are successfully determined by the fluctuation matching procedure so that the fluctuations are reproduced well regardless of whether the electrostatic interaction is included explicitly or not.

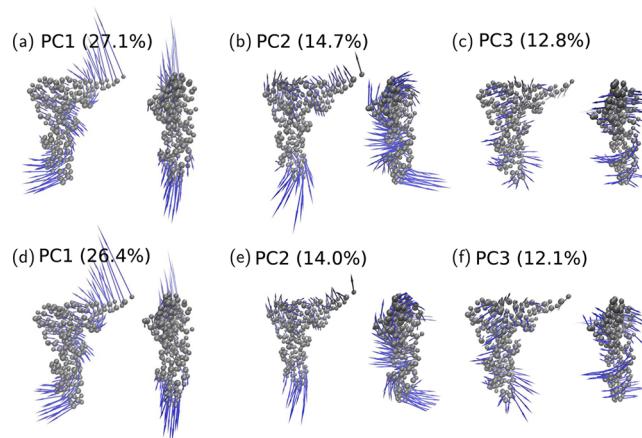
**3.3. tRNA and Ribosome Motion in Fluctuation.** In order to evaluate the further applicability of the model, we next looked into collective motions in the native-state dynamics of tRNA<sup>Phe</sup>. Using an ensemble obtained from trajectories of 300 K CG-MD simulations, we performed the principal component analysis to obtain the information of dominant modes of fluctuations in the native state. In Figure 4, the three largest modes are shown for both models A and B. We see that the first three modes in both models show very similar behaviors to each other, suggesting that the difference in the treatment of electrostatics does not affect so much the collective modes of fluctuations. Moreover, these three modes are similar to those obtained in previous theoretical studies.<sup>11,74,75</sup>

tRNA is known to be a highly dynamic molecule that takes markedly different conformations upon binding to its partners, most notably ribosome. Indeed, crystal structures of tRNA in various environment have been determined. First, structures without bound partners, i.e., the apo form, are known (e.g., tRNA<sup>Phe</sup>; PDB 1ehz<sup>53</sup>). Second, structures bound to A, P, and E sites of the ribosome in the canonical form are known.<sup>68</sup> Third, a tRNA conformation that corresponds to the hybrid state of translocation was reported.<sup>76</sup> In the hybrid intermediate state of translocation from the P site to the E site, a tRNA interacts with the small subunit of ribosome at the P site and with the large subunit at the E site (thus denoted as P/E state). Last, in yet another conformation of tRNA, it interacts with the

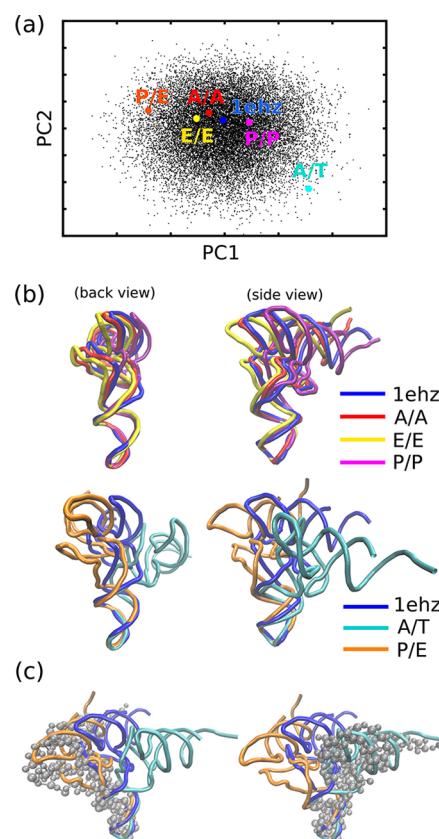


**Figure 3.** RMSF calculated by CG and AA simulations. RMSF values calculated by models A (red, circle), B (green, square), and C (magenta, diamond) and AA simulation (blue line) are plotted for phenylalanine tRNA (a; 1ehz), group II intron (b; 3g78), lysine riboswitch (c; 3d0x), unmodified phenylalanine tRNA (d; 3l0u), the complex of ribosomal protein S15 and a fragment of 16S rRNA (e; 1f7y), and the U1A/RNA complex (f; 1urn). All horizontal axes are the particle ID in the CG model, which is identical to the residue ID in proteins but not to the nucleotide ID in RNA because the CG model is three-beads-per-nucleotide.

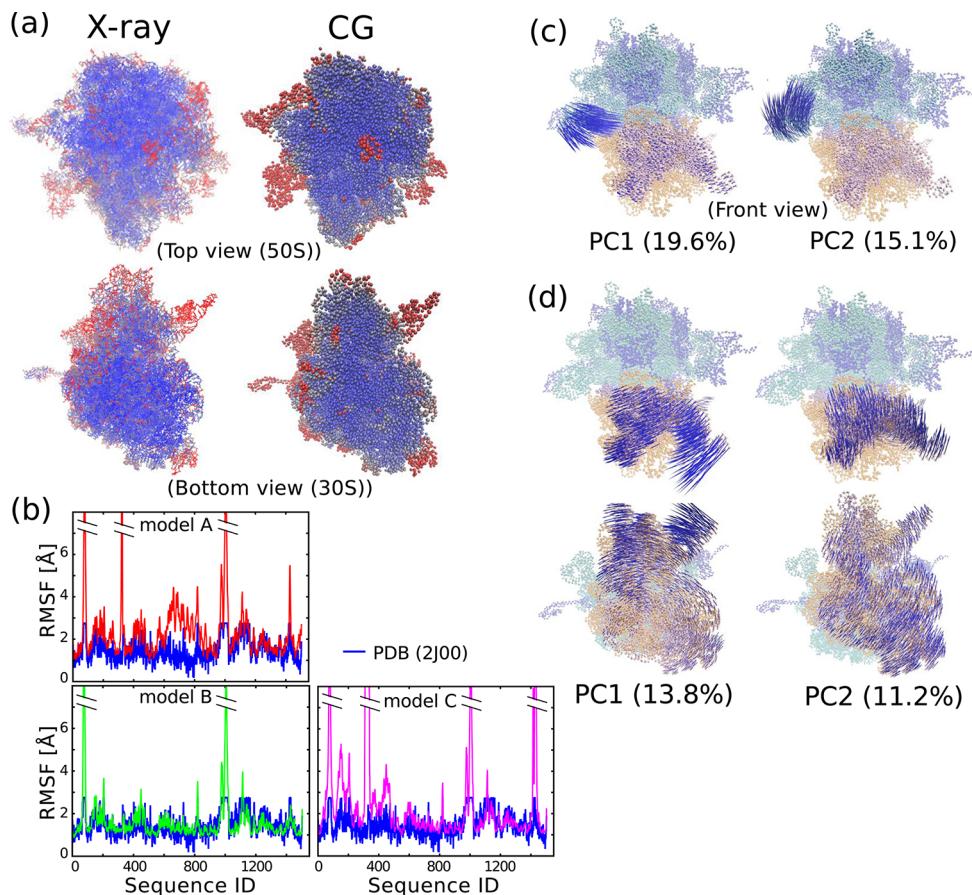
small subunit at the A site and with EF-Tu on the outside of the ribosome (denoted as A/T).<sup>77,78</sup> When they are superimposed at the anticodon stem region, the apo form and the canonical forms bound at A/A, P/P, and E/E quite well (Figure 5b, upper panel). On the other hand, tRNA in the hybrid form P/E and the A/T form bound to EF-Tu are largely swung to left and right, respectively (Figure 5b, lower panel). We now investigate how native fluctuations in the apo form (the reference structure being 1ehz) can explain/accommodate these diverse conformations. Using the first and second principal components, Figure 5a plots snapshots sampled from the current 300 K CG-MD as well as several crystal structures. The reference structure



**Figure 4.** Principal component modes of tRNA. Three slowest modes detected by principal component analysis are shown with percentages of contribution for model A ((a) first, (b) second, (c) third) and model B ((d) first, (e) second, (f) third).



**Figure 5.** Comparison between CG-simulation structures and some crystal structures. (a) Sampled structures in 300 K CG simulation of tRNA<sup>Phe</sup> (PDB: 1ehz) scattered on the two principal component axes. Positions of the reference and some functional structures are also indicated. (b) Superpositions of some functional structures. Canonical conformations (upper; A/A, P/P, E/E) and bent conformations (lower; A/T, P/E) are superimposed on the reference structure (blue; 1ehz) at the anticodon stem region. All of these structures were taken from the Protein Data Bank. (A/A, 3i8h; P/P, 3r8o; E/E, 3i8h; A/T, 2xqd; P/E, 3r8n) (c) Two representative structures obtained in the sampling simulation (gray sphere). A/T (cyan), P/E (orange), and the reference structure (blue) are also shown for comparison.



**Figure 6.** CG simulations of ribosome. (a) B-factors calculated from 300 K CG simulation with model B and obtained from experimental data (PDB: 2j00) are shown by colors. (b) RMSF values from 300 K CG simulations with all of the models comparing with the RMSF value calculated from the experimental B factor. Note that the experimental RMSF estimated from the B factor is bound to 2.76. (c,d) The first and second principal component modes of the entire ribosome (c) and only the small subunit motion (d) calculated from the 300 K CG simulation trajectory.

used in the CG-MD, 1ehz, is mapped near the origin. All of the canonical structures, A/A, P/P, and E/E, were mapped near the densely sampled central area. On the other hand, two hybrid state structures were mapped near the edge of sampled regions. The P/E hybrid state corresponds to the boundary of dense and sparse regions, while the A/T structure corresponds to a sparse data area. The result suggests that, although rare, the native dynamics of tRNA include large-amplitude motions up to hybrid states. That the A/T state is found in a sparse region implies that the conformation stores more internal energy. Valle et al. suggested that the deformation energy of the A/T state could serve to promote a following accommodation.<sup>77</sup> Our result qualitatively supports it. Notably, we fit the mean square fluctuation of the CG model with that of the AA model, and thus the absolute scale of fluctuation is of reasonable range. With the AA model, we cannot easily simulate these rare and large-amplitude motions, but the AA-based CG model can realize these rare events.

Finally, we briefly tested the models for the ribosome, one of the largest RNA–protein complexes. In the same manner as tRNA, we performed a 300 K CG-MD simulation and analyzed the absolute value of fluctuation and direction using principal component analysis. In Figure 6a, we compare fluctuations calculated from CG simulations with model B with those obtained from X-ray crystallography.<sup>67</sup> Clearly, the CG simulation reproduces well the characteristic distribution of fluctuation expressed in crystallographic B factor. Conversely,

models A and C, in which the electrostatics are included explicitly, overestimated molecular fluctuations in some peripheral regions (Figure 6b). It may suggest that the simple Debye–Hückel theory does not work well for large-amplitude fluctuations of highly charged molecules and that the structure-based term cannot compensate for the overestimated electrostatics. More sophisticated treatment of the ion environment may be necessary. Otherwise, employing only the structure-based potential (i.e., model B) is a better way to simulate fluctuations around the vicinity of the native state. In Figure 6c and d, we show the principal components for the entire structure (c) and only for the small subunit motion (d) from the model B simulation. The most fluctuated part globally is the L1 stalk region, which is consistent with other experimental and theoretical studies.<sup>7,8,10,79,80</sup> In the small subunit, a ratchet-like rotation is detected as the principal mode. This is also consistent with other studies.<sup>7,8,10,79,76</sup> Both the L1 stalk and the small subunit motion are thought to be important for ribosomal function at least in the translational translocation. Our CG simulation model can reproduce well such functionally important motions.

#### 4. CONCLUSIONS

We developed a CG model for RNA and RNA–protein complexes, in which RNA and the protein are represented by three-beads-per-nucleotide and one-bead-per-amino-acid resolutions, respectively.

The potential function is primarily structure-based, and parameters were determined by the fluctuation matching method so that the CG model well reproduces the fluctuation of the AA model around the native state. Using the data of 20 ns AA simulations of 26 RNAs and RNA–protein complexes, we successfully obtained a set of generic CG parameters. The model accuracy is examined by comparing RMSF values and also principal modes analyzed by the principal component analysis for tRNA and ribosome.

In the development, we tested three different treatments of the electrostatic effect. We found that the explicitness of electrostatic potential does not affect the results for tRNA. On the other hand, in the ribosome case, models A and C, which explicitly include electrostatic potential, seem to overestimate the electrostatic interaction. A recent study proposed an explicit-ion implicit-water CG model for DNA and showed that such a treatment can well reproduce the molecular behavior in an ionic environment.<sup>81</sup>

Our CG model and parameters can generally be applied to RNA–protein complexes which have unique structures, and it should be useful to study RNA-related biomolecular functions. As a future work, with the use of several reference structures corresponding to distinct functional states, the structure-based model must be able to realize large-scale conformational changes in the same manner as the switching potential and/or the multibasin potential used for protein.<sup>28,27</sup>

## ASSOCIATED CONTENT

### Supporting Information

List of target RNA molecules and RNA–protein complexes used in the fluctuation matching procedure; figures of the statistical survey of PDB structures; obtained parameters for models B and C; statistics of the electrostatic interactions in the ribosome. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: takada@biophys.kyoto-u.ac.jp.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Dr. Wenfei Li for useful discussions on the fluctuation matching method and Dr. Hiroo Kenzaki for leading the development of our simulation software, CafeMol. The calculations in this work were partly performed by using the RIKEN Integrated Cluster of Clusters (RICC) facility. N.H. acknowledges the support of a JSPS fellowship. This work was supported by a Grant-in-Aid for JSPS Fellows, KAKENHI 10J01723, partly by Research and Development of the Next-Generation Integrated Simulation of Living Matter and partly by a Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

## REFERENCES

- Bleichert, F.; Baserga, S. J. Ribonucleoprotein multimers and their functions. *Crit. Rev. Biochem. Mol.* **2010**, *45*, 331–350.
- Dethoff, E. A.; Chugh, J.; Mustoe, A. M.; Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* **2012**, *482*, 322–330.
- Elliott, D.; Ladomery, M. *Molecular Biology of RNA*, 1st ed.; Oxford University Press: New York, 2011.

(4) Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.

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

(6) Malhotra, A.; Tan, R. K.; Harvey, S. C. Modeling large RNAs and ribonucleoprotein particles using molecular mechanics techniques. *Biophys. J.* **1994**, *66*, 1777–1795.

(7) 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*, 9319–9323.

(8) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* **2004**, *147*, 302–314.

(9) Hyeon, C.; Thirumalai, D. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6789–6794.

(10) Trylska, J.; Tozzini, V.; McCammon, J. A. Exploring global motions and correlations in the ribosome. *Biophys. J.* **2005**, *89*, 1455–1463.

(11) Wang, Y.; Jernigan, R. L. Comparison of tRNA motions in the free and ribosomal bound structures. *Biophys. J.* **2005**, *89*, 3399–3409.

(12) Cui, Q.; Tan, R. K.-Z.; Harvey, S. C.; Case, D. A. Low-resolution molecular dynamics simulations of the 30S ribosomal subunit. *Multiscale Model. Simul.* **2006**, *5*, 1248–1263.

(13) Hyeon, C.; Dima, R. I.; Thirumalai, D. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* **2006**, *14*, 1633–1645.

(14) Das, R.; Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14664–14669.

(15) Kurkcuoglu, O.; Doruker, P.; Sen, T. Z.; Kloczkowski, A.; Jernigan, R. L. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. *Phys. Biol.* **2008**, *5*, 046005.

(16) Voltz, K.; Trylska, J.; Tozzini, V.; Kurkal-Siebert, V.; Langowski, J.; Smith, J. Coarse-grained force field for the nucleosome from self-consistent multiscaling. *J. Comput. Chem.* **2008**, *29*, 1429–1439.

(17) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* **2008**, *14*, 1164–1173.

(18) Jonikas, M. A.; Radmer, R. J.; Laederach, A.; Das, R.; Pearlman, S.; Herschlag, D.; Altman, R. B. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* **2009**, *15*, 189–199.

(19) Xia, Z.; Gardner, D. P.; Gutell, R. R.; Ren, P. Coarse-grained model for simulation of RNA three-dimensional structures. *J. Phys. Chem. B* **2010**, *114*, 13497–13506.

(20) Pasquali, S.; Derreumaux, P. HiRE-RNA: a high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B* **2010**, *114*, 11957–11966.

(21) Trylska, J. Coarse-grained models to study dynamics of nanoscale biomolecules and their applications to the ribosome. *J. Phys.: Condens. Matter* **2010**, *22*, 453101.

(22) Paliy, M.; Melnik, R.; Shapiro, B. A. Coarse-graining RNA nanostructures for molecular dynamics simulations. *Phys. Biol.* **2010**, *7*, 036001.

(23) Bernauer, J.; Huang, X.; Sim, a. Y. L.; Levitt, M. Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA* **2011**, *17*, 1066–1075.

(24) Go, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183–210.

(25) Bryngelson, J. D.; Onuchic, J. N.; Soccia, N. D.; Wolynes, P. G. Funnel, pathways and the energy landscape of protein folding: a synthesis. *Proteins* **1995**, *21*, 167–195.

(26) 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*, 937–953.
- (27) 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*, 5367–5372.
- (28) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844–11849.
- (29) Whitford, P. C.; Schug, A.; Saunders, J.; Hennelly, S. P.; Onuchic, J. N.; Sanbonmatsu, K. Y. Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function. *Biophys. J.* **2009**, *96*, L7–L9.
- (30) Whitford, P. C.; Ahmed, A.; Yu, Y.; Hennelly, S. P.; Tama, F.; Spahn, C. M. T.; Onuchic, J. N.; Sanbonmatsu, K. Y. Excited states of ribosome translocation revealed through integrative molecular modeling. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 18943–18948.
- (31) Tozzini, V. Multiscale modeling of proteins. *Acc. Chem. Res.* **2010**, *43*, 220–230.
- (32) Chu, J.; Izveko, S.; Voth, G. A. The multiscale challenge for biomolecular systems: coarse-grained modeling. *Mol. Simul.* **2006**, *32*, 211–218.
- (33) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, 244114.
- (34) Malhotra, A.; Harvey, S. C. A quantitative model of the Escherichia coli 16 S RNA in the 30 S ribosomal subunit. *J. Mol. Biol.* **1994**, *240*, 308–340.
- (35) Li, W.; Yoshii, H.; Hori, N.; Kameda, T.; Takada, S. Multiscale methods for protein folding simulations. *Methods* **2010**, *52*, 106–114.
- (36) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K.; Yao, X.-Q.; Takada, S. CafeMol: a coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* **2011**, *7*, 1979–1989.
- (37) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (38) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. A coarse grain model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901.
- (39) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E. L. I.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **2008**, *14*, 465–481.
- (40) Lejeune, D.; Delsaux, N.; Charlotteau, B.; Thomas, A.; Brasseur, R. Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. *Proteins* **2005**, *61*, 258–271.
- (41) Gan, H. H.; Schlick, T. Chromatin ionic atmosphere analyzed by a mesoscale electrostatic approach. *Biophys. J.* **2010**, *99*, 2587–2596.
- (42) Oosawa, F. *Polyelectrolytes*; Marcel Dekker: New York, 1971.
- (43) Draper, D.; Grilley, D.; Soto, A. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 221–243.
- (44) Macke, T.; Case, D. A. In *Molecular Modeling of Nucleic Acids*; Leontis, N. B., Santa-Lucia, J., Eds.; American Chemical Society: Washington, DC, 1998; pp 379–393.
- (45) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; et al. AMBER 11; University of California: San Francisco, CA, 2011. <http://ambermd.org/> (accessed July 2012).
- (46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (47) Allain, F. H.; Varani, G. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **1995**, *250*, 333–353.
- (48) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **1996**, *264*, 968–980.
- (49) Chang, K.-Y.; Tinoco, I. The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* **1997**, *269*, 52–66.
- (50) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. Crystallographic structure of an RNA helix: [U(UA)<sub>6</sub>A]<sub>2</sub>. *J. Mol. Biol.* **1989**, *209*, 459–474.
- (51) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science* **1996**, *274*, 2065–2069.
- (52) Su, L.; Chen, L.; Egli, M.; Berger, J. M.; Rich, A. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Mol. Biol.* **1999**, *6*, 285–292.
- (53) Shi, H.; Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA* **2000**, *6*, 1091–1105.
- (54) Timsit, Y.; Bombard, S. The 1.3 Å resolution structure of the RNA tridecamer r(GCGUUUGAACGC): Metal ion binding correlates with base unstacking and groove contraction. *RNA* **2007**, *13*, 2098–2107.
- (55) Garst, A. D.; Héroux, A.; Rambo, R. P.; Batey, R. T. Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.* **2008**, *283*, 22347–22351.
- (56) Wang, J. Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr., Sect. D* **2010**, *66*, 988–1000.
- (57) Sekine, S.; Nureki, O.; Dubois, D. Y.; Bernier, S.; Chenevert, R.; Lapointe, J.; Vassylyev, D. G.; Yokoyama, S. ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J.* **2003**, *22*, 676–688.
- (58) Hauenstein, S.; Zhang, C.-M.; Hou, Y.-M.; Perona, J. J. Shape-selective RNA recognition by cysteinyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1134–1141.
- (59) Conn, G. L.; Draper, D. E.; Lattman, E. E.; Gittis, A. G. Crystal structure of a conserved ribosomal protein-RNA complex. *Science* **1999**, *284*, 1171–1174.
- (60) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaia, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. The crystal structure of UUCG tetraloop. *J. Mol. Biol.* **2000**, *304*, 35–42.
- (61) Fedorov, R.; Meshcheryakov, V.; Gongadze, G.; Fomenkova, N.; Nevskaia, N.; Selmer, M.; Laurberg, M.; Kristensen, O.; Al-Karadaghi, S.; Liljas, A.; Garber, M.; Nikonov, S. Structure of ribosomal protein L15 complexed with RNA provides new insights into the CTC family of stress proteins. *Acta Crystallogr., Sect. D* **2001**, *57*, 968–976.
- (62) Tishchenko, S.; Nikonova, E.; Nikulin, A.; Nevskaia, N.; Volchkov, S.; Piendl, W.; Garber, M.; Nikonov, S. Structure of the ribosomal protein L1-mRNA complex at 2.1 Å resolution: common features of crystal packing of L1-RNA complexes. *Acta Crystallogr., Sect. D* **2006**, *62*, 1545–1554.
- (63) Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C.-H.; Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **1994**, *372*, 432–438.
- (64) Wang, X.; Hall, T. M. T. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **2001**, *8*, 141–145.
- (65) Wild, K.; Sinning, I.; Cusack, S. Crystal structure of an early protein-RNA assembly complex of the signal recognition particle. *Science* **2001**, *294*, 598–601.
- (66) Hainzl, T.; Huang, S.; Sauer-Eriksson, A. E. Structure of the SRP19-RNA complex and implications for signal recognition particle assembly. *Nature* **2002**, *417*, 767–771.
- (67) Selmer, M.; Dunham, C.; Murphy, F., IV; Weixlbaumer, A.; Petry, S.; Kelley, A.; Weir, J.; Ramakrishnan, V. Structure of the 70S

ribosome complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935–1942.

(68) Jenner, L.; Demeshkina, N.; Yusupova, G.; Yusupov, M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **2010**, *17*, 555–560.

(69) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. a.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

(70) Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **1990**, *94*, 8021–8024.

(71) Aduri, R.; Pisciuk, B. T.; Saro, P.; Taniga, H.; Schlegel, H. B.; SantaLucia, J. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theory Comput.* **2007**, *3*, 1464–1475.

(72) Auffinger, P.; Louise-May, S.; Westhof, E. Multiple molecular dynamics simulations of the anticodon Loop of tRNA<sup>Asp</sup> in aqueous solution with counterions. *J. Am. Chem. Soc.* **1995**, *117*, 6720–6726.

(73) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; Chapter 6, pp 116–158.

(74) Nakamura, S.; Doi, J. Dynamics of transfer RNAs analyzed by normal mode calculation. *Nucleic Acids Res.* **1994**, *22*, 514–521.

(75) Caulfield, T. R.; Devkota, B.; Rollins, G. C. Examinations of tRNA range of motion using simulations of cryo-EM microscopy and X-ray data. *J. Biophys.* **2011**, *2011*, 219515.

(76) Dunkle, J. A.; Wang, L.; Feldman, M. B.; Pulk, A.; Chen, V. B.; Kapral, G. J.; Noeske, J.; Richardson, J. S.; Blanchard, S. C.; Cate, J. H. D. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **2011**, *332*, 981–984.

(77) Valle, M.; Zavialov, A.; Li, W.; Stagg, S. M.; Sengupta, J.; Nielsen, R. C.; Nissen, P.; Harvey, S. C.; Ehrenberg, M.; Frank, J. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat. Struct. Biol.* **2003**, *10*, 899–906.

(78) Voorhees, R.; Schmeing, T.; Kelley, A.; Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **2010**, *330*, 835–838.

(79) Valle, M.; Zavialov, A. V.; Sengupta, J.; Rawat, U.; Frank, J. Locking and unlocking of ribosomal motions. *Cell* **2003**, *114*, 123–134.

(80) Cornish, P. V.; Ermolenko, D. N.; Staple, D. W.; Hoang, L.; Hickerson, R. P.; Noller, H. F.; Ha, T. Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2571–2576.

(81) Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J. A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. *J. Chem. Phys.* **2011**, *135*, 165104.