JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

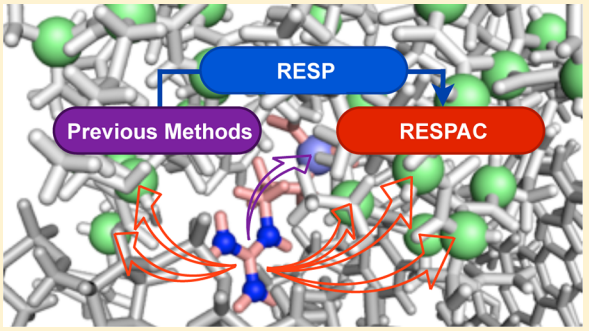
RESPAC: Method to Determine Partial Charges in Coarse-Grained Protein Model and Its Application to DNA-Binding Proteins

Tsuyoshi Terakawa and Shoji Takada*

Department of Biophysics, Graduate School of Science, Kyoto University, Kitashirakawa Owake-cho, Sakyo, Kyoto, 606-8501, Japan

Supporting Information

ABSTRACT: While coarse-grained (CG) molecular simulations for large biomolecular complexes have become popular, their electrostatic treatment is often rather simplistic. Here, for C_α-based CG models of globular proteins, we developed a method to obtain an optimal partial charge set and applied it to 17 proteins that bind to DNA. The method follows the restrained electrostatic potential (RESP) fitting method widely used for determination of atomic partial charges in all-atom (AA) molecular mechanics. The proposed method, called the RESPAC method, finds optimal partial charges on surface C_α CG beads so that these charges best approximate the electrostatic potential of the AA model under a restraint term. Comparison of the AA and CG electrostatic potentials showed that the RESPAC charges outperformed simplistic integer-valued charges. Then, the RESPAC method was applied to lac repressor binding to a nonspecific DNA sequence. We found that the CG simulations correlated well with AA molecular dynamics simulations. We also performed CG simulations of 16 other transcription factors. The differences in binding interfaces between nonspecific and specific DNAs were, on average, reduced by using the RESPAC charges. Yet, for several proteins, the nonspecific DNA binding interface was quite different from that of the specific binding interface, which is in accord with a previous report.



INTRODUCTION

Electrostatic interaction plays crucial roles in virtually all molecular processes ranging from those involving small molecules to biomolecules. Among biomolecular interactions, protein−nucleic acid interactions, for example, are dominated by electrostatic interactions because nucleic acids are highly negatively charged and their partner proteins tend to be positively charged at the interface.[1−6] In conventional molecular mechanics, electrostatic interactions are calculated by using point charges of fractional values put on atoms in the all-atom (AA) representation. Molecular dynamics (MD) simulation with the molecular mechanics has become an increasingly powerful approach in computational studies of biomolecules.[7,8]

Despite its marked progress,[9] AA MD still suffers from the huge number of atoms in biomolecules and the biologically relevant long time scale. One of the promising alternatives to speed up MD simulations is to use coarse-grained (CG) representations, which are becoming increasingly popular.[10−13] Recent development in systematic multiscale theories provides firm basis on CG models.[14] Among many different CG resolutions used for proteins, one quite popular way is to use one CG particle for each amino acid usually locating the CG particle at the C_α atom position (one-bead-per-one-amino-acid model), which has produced many promising results.[15−20]

Some studies using this type of CG simulations tried to incorporate electrostatic interactions into the CG model, though the treatment of electrostatics was often simplistic.[18,21−25] They usually put integer-valued charges (1, or −1) on each CG bead representing a charged amino acid (Asp, Glu, Lys, Arg, and His). In reality, charged amino acids have relatively long side-chains and the charges are located at the end of the side-chains. Thus, the discrepancy of charge positions between AA and C_α-based CG models is notable and can cause poor estimation of electrostatic interactions. In this study, we address how this inaccuracy can be remedied for globular proteins/domains.

When we deal with folded proteins or domains, for example in the context of protein−protein and protein−DNA dockings, many side-chains have well-defined rotameric conformations or point in specific orientations. In such cases, a charged atom at the end of a side-chain may be surrounded by certain other amino acids. Thus, within C_α-based CG modeling, putting an integer-valued charge on the C_α position of the charged amino acid is not necessarily the best approximation to the electrostatic potential around the region. Instead, distributing fractional charges to surrounding amino acids may improve the accuracy of electrostatics. We note that, for a highly fluctuating environment such as the unfolded state, amino acids that surround a charged amino acid are not fixed and thus the situation is different. Here, focusing on globular proteins/domains, we developed a method to determine an optimal set of charges on C_α atoms in CG models.

Received: August 10, 2013
Published: January 3, 2014

To do this, we borrow an idea used in determining partial charges in AA models where one popular method is to fit electrostatic potential (ESP). The ESP fitting method was first developed by Momany,[26] followed by Cox and Williams,[27] Singh and Kollman,[28] Chirlian and Francl,[29] Breneman and Wiberg,[30] and others. In these methods, partial atomic charges are derived so that the ESP calculated by numerically solving Poisson−Boltzmann equation with the derived charges reproduces the quantum-chemically calculated ESP as much as possible. In practice, it is known that the use of the ESP alone for fitting faces two problems. First, we cannot fit the points that are too close to the charge position in the fitting process because of the point-charge approximation, and thus, the fitting points need to be far from buried charges. As a consequence, charges buried inside proteins are poorly determined.[31] Second, the fitted charges are often too large in absolute value and are sensitive to molecular geometries, which is known as the overfitting problem.[32−34] This causes problems in MD simulations in particular. One of the most widely used remedies is the restrained ESP (RESP) fitting method[35] in which a penalty function is introduced to prevent charges from deviating greatly from target values during the fitting. The RESP method considerably reduces the problem, with only a minor decrease in the quality of the fit to the quantum-chemically calculated ESP.[36−43] The charge values in commonly used AA force fields such as AMBER[44] are decided by the RESP method and recognized as particularly well suited for MD simulations.

In this work, following the same philosophy as the RESP method, we developed a new RESP fitting method to derive the effective charges on surface CG beads so that the ESP of the CG model is essentially the same as that of the AA model. We call this method the RESP method from atomic charges to CG charges (RESPAC). Comparison with the atomic ESP clearly showed that the partial charge set determined by the RESPAC method outperforms the integer-valued charge set previously used in many applications. Somewhat similar, but not the same, approaches have been taken for some finer or coarser CG modeling.[45−47] In addition, a closely related, and even a more elaborating, topologically based multipolar reconstruction method has been developed.[48−50]

To test the RESPAC method, we then applied it to interactions of transcription factors with DNA, partly motivated by recent work of Marcovitz et al.[51,52] Using CG simulations with in silico designed peptides, they found that the discrepancy between the nonspecific and the specific DNA binding interfaces slows the transition from nonspecific to specific DNA binding. In addition, they calculated the binding interface discrepancy of the natural proteins and concluded that their transition from nonspecific to specific DNA binding is indeed slowed by this discrepancy. They also performed CG simulations with some of the natural proteins at low temperature and at low ion concentration, revealing that the discrepancy of the binding interfaces between the specific and the nonspecific binding complexes is well correlated with the calculated binding interface discrepancy. These were interesting observations, but that CG simulation was conducted using a CG model with integer-valued charges. Here, using the RESPAC charges, we test whether these observations were robust or altered.

Using the RESPAC charges, we simulated the binding of 17 transcription factors, especially focusing on the example of lac repressor, with a nonspecific DNA sequence at room temperature and at physiological ion concentration. Lac repressor binding is a particularly suitable system because there are NMR structures for its complexes both with a nonspecific DNA sequence and with its specific DNA sequence and because the interaction between the protein and DNA is dominated by electrostatics. The validation for lac repressor confirmed that the CG simulation with RESPAC charges reproduced the binding interface of this protein to nonspecific DNA better than that with integer-valued charges. Systematic comparison of 16 other transcription factors showed that, for some cases, we see a smaller discrepancy of binding interfaces with the RESPAC charges than with the integer-valued charges. Yet, the simulations of several proteins showed that the nonspecific binding interface is noticeably different from the specific binding interface, which is in accord with the previous observation.[51,52]

THEORY

RESPAC. Given a molecular representation, we try to approximate the molecular electrostatic interactions as accurately as possible. Within classical point-charge modeling, we often optimize fractional (i.e., noninteger) effective charges on particles using some score functions that measure the fit between the model and the reference electrostatics. One frequently used score function is the electrostatic potential (ESP),[26−30] which has been used for the derivation of atomic partial charges of AA models. In the method, we minimize

$$\chi_{\mathrm{esp}} = \int_{\Omega} d^3\vec{r}[\phi_{\mathrm{ref}} - \phi(\vec{q})]^2 \tag{1}$$

with respect to effective partial charges $\vec{q}$ (as a vector, it represents all the partial charges collectively). Here, ϕ_{ref} is the reference ESP calculated normally by the ab initio quantum chemical calculations, and $\phi(\vec{q})$ is ESP calculated by numerically solving the Poisson−Boltzmann equation with the derived charges $\vec{q}$. Ω represents the neighboring space around the molecule.

In this work, we apply the ESP score function to find partial charges on coarse-grained (CG) particles, that is, beads located at the C_α atom of every amino acid, using the reference ESP calculated by a standard AA force field. Somewhat similar approaches were proposed before at different levels of coarse-graining.[46,47]

We define ϕ_{ref} as the ESP given by the standard force field of the AA model with the continuum solvent model. Given a protein conformation, we can easily calculate ϕ_{ref} by solving the Poisson−Boltzmann equation

$$\nabla\cdot[\varepsilon(\vec{r})\nabla\cdot\phi_{\mathrm{ref}}(\vec{r})] - \varepsilon(\vec{r})\kappa\,\sinh[\phi_{ref}(\vec{r})] + \frac{4\pi\rho(\vec{r})}{kT} = 0 \tag{2}$$

where $\vec{r}$ is a position vector of each grid around the molecule, $\varepsilon(\vec{r})$ is the position-dependent dielectric constant, κ is the inverse of the Debye length, $\rho(\vec{r})$ is the atomic charge density, k is the Boltzmann constant, and T is temperature. In the current CG modeling, we use the simple Debye−Hückel model, and thus the CG ESP becomes

$$\phi = \sum_i q_i \frac{\exp(-\kappa|\vec{r} - \vec{r}_i|)}{\varepsilon|\vec{r} - \vec{r}_i|} \tag{3}$$

where $\vec{r}_i$ is a position vector of CG beads and ε is the dielectric constant of the solvent. Partial charges $\vec{q}$ in the CG representation are located on CG particles.

Although the ESP fitting method provides a promising way to obtain effective charges, it has a flaw known as the overfitting problem. To yield a minuscule decrease of χ_{esp}, some effective charges can be very sensitive to small changes in molecular structure, and so on.[31] In the context of the derivation of atomic partial charges, Bayly et al. developed a method to remedy this problem, which is called RESP (restrained electrostatic potential fitting).[35] In the RESP method, the score function,

$$\chi_{RESP} = \chi_{esp} + \chi_{rstr} \tag{4}$$

includes a restraint term defined as

$$\chi_{rstr} = \delta \sum_i (q_i - q_{0i})^2 \tag{5}$$

Here, δ is the scale factor determining the strength of the restraint. q_{0i} are target charges, to which partial charges $\vec{q}$ are restrained. In the previous work, this approach considerably reduced the overfitting problem, with only a minor decrease in the quality of fit to the quantum chemical ESP.[36−43] The RESP scheme has not been applied to one-bead-per one-amino-acid CG modeling. In this work, we applied it to find the CG effective charge. We call our method RESPAC, that is, the RESP method from atomic charges to CG charges. Specifically, we minimize the evaluation function

$$\chi_{RESPAC}(\phi_{ref}^{PB}, \vec{q}) = \int_\Omega d^3\vec{r}\left[\phi_{ref}^{PB} - \sum_i q_i \frac{\exp(-\kappa|\vec{r} - \vec{r}_i|)}{\varepsilon|\vec{r} - \vec{r}_i|}\right]^2 + \delta \sum_i (q_i - q_{0i})^2 + \lambda(q_{tot} - \sum_i q_i)^2 \tag{6}$$

with respect to the effective charges $\vec{q}$. In this equation, ϕ_{ref}^{PB} is the ESP at many grid points around a protein calculated by numerically solving the Poisson−Boltzmann equation with atomic structure, $\vec{r}$ is a position vector of each grid, $\vec{r}_i$ is a position vector of CG beads, κ is the inverse of the Debye length, and ε is the dielectric constant of the solvent. The third term of this equation is for total charge restraint. δ and λ are the scale factors determining the strength of the restraint. q_{tot} is the total charge of the atomic structure.

Charges in Coarse-Grained Protein Model. In this work, we used the CG representation in which one bead corresponds to one amino acid. Each bead is put on the corresponding C_α atom position. Of these beads, we let the beads representing a surface amino acid have nonzero charges because "buried" charges tend to be poorly determined in the RESP method and could cause unexpected destabilization of the protein structure. Therefore, i in the evaluation function (eq 6) runs over just the ID of the bead representing a surface amino acid. Here, a surface amino acid was decided in the following way. First, we prepared a 1 Å spacing 3D grid around the protein. Second, we put a probe with 4 Å radius on each grid point and defined the point as an "outside point" if the probe did not overlap with the protein. Finally, we defined the residues that were nearest to each "outside point" as "surface residues". The target charge (q_0 in eq 6) of each bead was set to 0 regardless of the residue type. This tends to distribute charges of a side chain to C_α beads around the side chain, allowing a large charge to be on the CG beads near the atomic charge position.

Charges in Coarse-Grained DNA Model. For the DNA model, each nucleotide is represented by three CG particles corresponding to sugar (S), phosphate (P), and nitrogenous base (N). As described in Methods section, we applied the RESPAC method to B-type dsDNA, as well. With the biasing term δ set to 0.0, we obtained the RESPAC charges, −0.01, −0.99, and 0.00 for S, P, and N, respectively. Thus, the RESPAC method gives nearly identical charges to simple integer charges. We can think of two reasons for this result. First, the current CG representation of DNA explicitly contains the phosphate that has −1 charge in atomic scale (This is markedly different from the case of proteins, in which charges in atomic scale are located at termini of side-chains, whereas the CG model uses only C_α). Second, B-type DNA takes a homogeneous structure. These together, perhaps, make the RESPAC procedure not crucial for the current DNA model. Thus, in this study, we used integer charges for the CG DNA model (See Methods section for more detail).

Determination of the Scale Factors δ and λ. In order to obtain the CG partial charges $\vec{q}$ which minimize χ_{RESPAC} (eq 6), we need to decide the values of the scale factors δ and λ. We can assign an arbitrarily large value to λ (we set $\lambda = 10^5$ in this work) because this parameter does not affect the resulting effective charges as long as it is sufficiently large. The root-mean-square deviation between the total RESPAC charge and the total AA charge was 0.225 and is adequately small.

The more important value is δ. The method used to determine δ in this work is based on the idea that the ESP of the effective charge derived for the native structure should not be very different from that derived for slightly perturbed structures. Based on this idea, first, we obtained 10 atomic structures by performing short molecular dynamics (MD) simulations from the native structure with an AA model together with an implicit solvent model. We call these 10 structures "decoys" hereafter. Second, for each of the decoy structures, we calculated $\vec{q}^{decoy}(\delta)$, which minimizes the evaluation function $\chi_{RESPAC}(\phi_{decoy}^{PB}, \vec{q}^{decoy}(\delta))$ (eq 6). Here, ϕ_{decoy}^{PB} is the ESP calculated by numerically solving the Poisson−Boltzmann equation using the decoy structure. We note that $\vec{q}^{decoy}$ s are dependent on δ. Third, we calculated the normalized error $\Delta(\vec{q}^{decoy}(\delta))$ for each $\vec{q}^{decoy}(\delta)$, where

$$\Delta(\vec{q}) = \frac{\int_\Omega d^3\vec{r}\left[\phi_{ref}^{PB} - \sum_i q_i \frac{\exp(-\kappa|\vec{r} - \vec{r}_i|)}{\varepsilon|\vec{r} - \vec{r}_i|}\right]^2}{\int_\Omega d^3\vec{r}[\phi_{ref}^{PB}]^2} \tag{7}$$

Note that ϕ_{ref}^{PB} is the ESP calculated by numerically solving the Poisson−Boltzmann equation using the "native" structure. Therefore, $\Delta(\vec{q}^{decoy}(\delta))$ represents the difference between the ESP of the effective charge derived using a decoy structure and that of the atomic native structure. Averaging it over 10 decoy structures, we use $\langle\Delta(\vec{q}^{decoy}(\delta))\rangle_{decoy}$ as the score function to be minimized to obtain an optimal δ. Here, $\langle\ldots\rangle_{decoy}$ represents the average over the Δ for each decoy.

METHODS

Calculation of Electrostatic Potential ϕ_{ref}^{PB} from Atomic Protein Structure. In order to obtain the ESP ϕ_{ref}^{PB} in the AA representation, we solved the Poisson−Boltzmann equation using Adoptive Poisson−Boltzmann Solver (APBS).[53] Input structure files for the 17 target proteins (the proteins listed in

Figure 1A and lac repressor DNA binding domain) were obtained using the pdb2pqr software package.[54] We selected

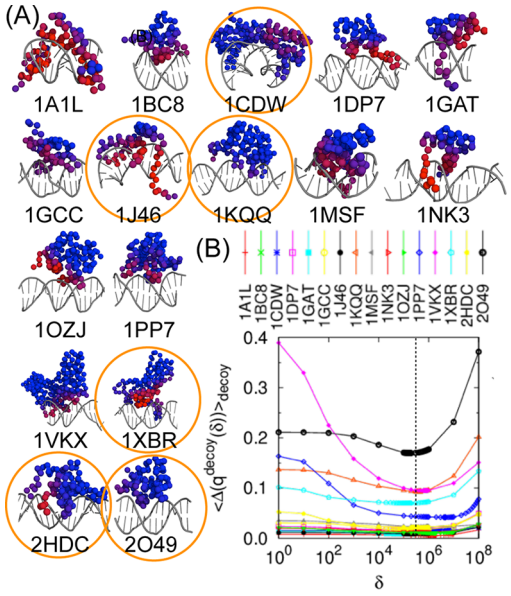


Figure 1. (A) 16 target proteins whose RESPAC charges were calculated. PDB ID is written below each structure. Balls represent C_α atoms of the protein, and gray sticks represent double-strand DNA. Color and its intensity on each bead represent the contact probability (blue represents 0.0 and red represents 1.0) calculated from the trajectory of the CG simulation with RESPAC charges. (B) Same type of plot in Figure 2A for proteins listed in part A.

these target proteins so that as many families of eukaryotic transcription factors as possible were covered. The charge set of the CHARMM27 force field[55] was used for the atomic partial charges and mass of each atom in these proteins. Based on the idea that fitting can be more accurate when more grid points used for fitting have nonzero potential value, we set $\varepsilon(\vec{r}) = \varepsilon = 1$, which is position independent. κ was set to a small but non-negligible value ($\kappa = 0.029$ Å^{-1}) on the same basis and based on the requirement that it should be as large as that used in the CG MD simulation. The same values were used to calculate the CG ESP using the Debye–Hückel model. For the numerical integration over the space Ω, we used grid points that have distances between 3 Å and 12 Å from the van der Waals surface of the protein. The protein van der Waals surface was determined using the rolling ball algorithm.[56]

Making Decoys. In order to decide the scale factor δ, we obtained the decoys by implicit solvent atomic MD simulation where the C_α atom is restrained to the initial position. In this simulation, the CHARMM27 force field with CMAP correction[55] was used for the protein, and the generalized Born/surface area (GB/SA) implicit solvent model[57] was used for the solvent. For the GBSA simulations, the Born radii were calculated using Still's algorithm.[57] The dielectric constant of water was set to 80. The surface tension for a water–oil interface is 0.0049 [kJ/mol·nm^2]. The bond lengths that include hydrogen atoms were constrained by the p-LINCS.[58] The 1-ns simulation was performed in the *NVT* ensemble at T = 300 [K] realized by Langevin dynamics using GROMACS 4.5.[59] All the simulations used a 2-fs time step. Atomic coordinates of the simulation system were saved every 1 ps, and ten decoys were randomly chosen from the saved structures. This simulation was performed for all 17 proteins (the proteins listed in Figure 1A and lac repressor DNA binding domain).

Coarse-Grained Simulation. To characterize the non-specific binding interface between protein and DNA, we performed coarse-grained (CG) simulations of 17 proteins (The proteins listed in Figure 1A and the lac repressor DNA binding domain) with 200-bp nonspecific DNA. For a CG model of the protein, we used the AICG2 model (one of the Go-like models) developed by Li et al.[20] where the atomic-interaction-based potential is imposed to stabilize the native structure. We used crystal or NMR structures of protein–DNA complexes as the reference structures. We put the effective charges derived by the RESPAC method on the protein CG beads and treated electrostatic interactions using the Debye–Hückel model.

For the DNA model, we used the 3SPN.1 model[60] developed by de Pablo's group. As described above, each nucleotide is represented by three CG particles corresponding to sugar (S), phosphate (P), and nitrogenous base (N). We applied the RESPAC for DNA in which we imposed a constraint so that the CG particles that represented the same moiety (S, P, or N) had the same charges. First, we calculated the AA ESP around the 50 bp ideal B-type DNA. Second, we obtained the RESPAC charges using eq 6 with the δ set to 0.0. We fit the ESP around the 10 bp DNA locating at the center of the 50 bp DNA to neglect the effect of the DNA termini. The obtained charges were −0.01, −0.99, and 0.00 for S, P, and N, respectively, which are essentially identical to the values in the 3SPN.1 model (0.00, −1.00, and 0.00 for S, P, N, respectively). Thus, in this study, we used the charges in the 3SPN.1 model for DNA.

Each target protein and the 200-bp dsDNA were placed in a box with dimensions of 400 × 750 × 400 Å with the DNA maintained linearly (designated as the y-axis) by pinning the two ends. Each target protein was initially placed near, but not in contact with, DNA. The production runs of the CG simulations were conducted using Langevin dynamics for 3 × 10^8 MD steps with friction coefficient $\gamma = 0.02$, temperature T = 300 K, dielectric constant $\varepsilon = 80.0$ and ionic strength $I = 0.150$. Atomic coordinates of the simulation system were saved every 1 × 10^4 MD steps. All simulations were performed with CafeMol 2.0,[61] a general-purpose CG biomolecular modeling and simulation software. Further details of the simulation can be found in Terakawa et al.[25] and Supporting Information therein.

Explicit Solvent All Atom Simulation of lac Repressor DNA Binding Domain. In order to validate a nonspecific binding interface between protein and DNA in a CG simulation, we performed explicit solvent atomic MD simulations of one of the target proteins, lac repressor DNA-binding domain. The fact that there are available NMR structures of the complex of this protein with nonspecific DNA (PDB ID: 1OSL) as well as with specific DNA (PDB ID: 1L1M), in which electrostatic interaction dominates intermolecular interaction, makes this system suitable for validation of our method. We used these structures as initial structures. In these simulations, the CHARMM27 force field with CMAP correction[55] was used for the protein and the DNA, and the TIP3P model was used for water. We simulated the system with cubic periodic boundary conditions and used the particle-mesh Ewald method[62] for electrostatic interaction. The bond lengths that include hydrogen atoms in the solute (the solvent) were constrained by p-LINCS[58] (by SETTLE[63]). Each structure was placed in a box with dimensions of 85 × 55 × 55 Å with the

DNA maintained linearly by pinning the two ends. After energy minimization of the initial structure, a 100-ps equilibration NVT MD simulation was performed at 300 K with all solute atoms positionally restrained in the NVT ensemble, followed by a 100-ps NPT MD simulation at the same temperature and at 1 atom with all solute atoms positionally restrained. The production runs were conducted using Langevin dynamics for 1-μs with GROMACS 4.5.[59] All the simulations were done by using a 2-fs time step. Atomic coordinates of the simulation system were saved every 10-ps and the trajectories after 200-ns were used for analysis.

RESULTS AND DISCUSSION

Validation of the RESPAC Method: DNA Binding Domain of lac Repressor. Here, to test the RESPAC method, we applied it to the lac repressor DNA binding domain. For this protein, the structure of its complex with a nonspecific DNA sequence has been obtained by NMR.[4] Since nonspecific protein−DNA interactions are likely to be dominated by electrostatic interactions, this is one of the ideal systems to test the RESPAC method. After obtaining the bound structure by the CG simulation with the RESPAC method, we can compare it with the NMR structure.

To do so, as described in the Theory section, we first need to decide the value of δ in eq 6 for the RESPAC. First, we obtained 10 near-native decoy structures by short MD simulation with the AA model and the implicit solvent model (see Methods section for more details). Then, using these decoys, we calculated $\langle\Delta(\vec{q}^{\,\mathrm{decoy}}(\delta))\rangle_{\mathrm{decoy}}$ and plotted it against δ (Figure 2A). In this plot, we see a single minimum at $\delta = 1.2 \times 10^6$.

Using this optimal value of δ, we calculated the effective charges which minimize eq 6 (red bars in Figure 2B). For comparison, in the figure, we also show simple integer-valued charges of amino acids, which many previous CG simulation studies used (blue bars). From this figure, we see that the absolute value of the RESPAC charge (red bars in Figure 2B) does not exceed 1.5, suggesting that a serious overfitting problem did not occur here. As a comparison, if we use $\delta = 0.0$, which turns off the restraint (green bars in Figure 2B), the absolute value of RESPAC charge exceeds 3.0 in the worst case, which caused a large value of $\langle\Delta(\vec{q}^{\,\mathrm{deco}}\ (\delta))\rangle_{\mathrm{decoy}}$ in Figure 2A.

We also see that not only the absolute values of charges but also the domain positions of the charges are different between the RESPAC charge (red bars in Figure 2B) and the integer-valued charges (blue bars in Figure 2B). Interestingly, the difference is marked only for positive charges. Figure 2B clearly indicates that the optimal CG charges are not transferable, but rather are dependent on the target protein and its local environment (similar findings were reported even for AA RESP charges[43]). Thus, they should be recomputed for every target protein.

Figure 2C illustrates why this happens. The blue beads represent the nitrogen atoms of Arg22 (indicated by an arrow in Figure 2B). In the previous integer-valued charge model, these atomic charges are put approximately at the positive unit charge on the purple bead (i.e., the C_α atom). In the case of the RESPAC charges, however, partial charges can be put on the green beads in Figure 2C. The contribution of the blue beads to the ESP around them can be better described by the partial charges on green beads than by one integer charge on the purple bead. This is because these green beads which are located far from Arg22 in the amino acid sequence are located

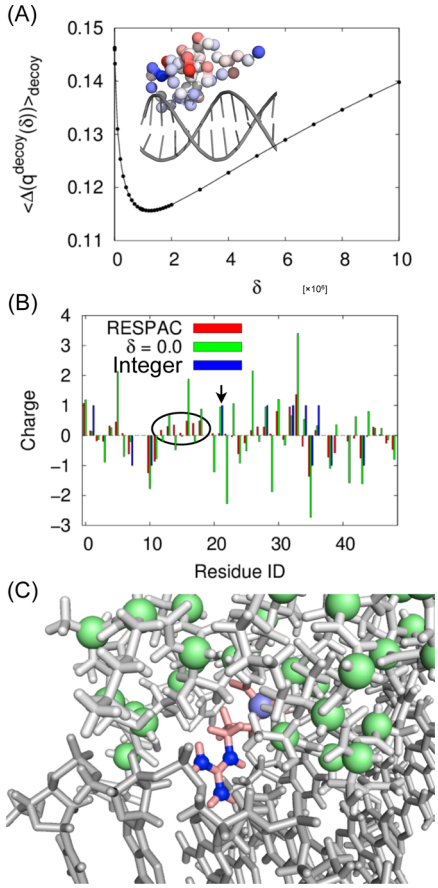


Figure 2. Derivation of RESPAC charges of lac repressor DNA binding domain. (A) $\langle\Delta(q^{\mathrm{decoy}}(\delta))\rangle_{\mathrm{decoy}}$ is plotted against δ. In the inset of part A, we show an NMR structure of the lac repressor DNA binding domain. Balls represent C_α atoms of the protein, and gray sticks represent double-strand DNA. The color and its intensity of each bead represent the RESPAC charge value on each C_α atom (blue represents +1 and red represents −1). (B) The optimal ($\delta = 1.2 \times 10^6$) RESPAC charge value ("RESPAC"; red bars) on each CG bead is plotted. For comparison, the charge value derived using RESPAC with $\delta = 0.0$ ("$\delta = 0.0$"; green bars) and charge value which is decided by amino acid type ("Integer"; blue bars) are also plotted. (C) Atomic structure of lac repressor DNA binding domain around Arg22. White and gray sticks represent the protein and a double-strand DNA, respectively. The blue beads represent charged atoms in Arg22, the purple bead represents the C_α atom of Arg22 on which the integer-valued charge is put, and the green beads represent the C_α atoms on which RESPAC charges are put.

near it in space. Indeed, the RESPAC method assigned positive partial charges on these beads (indicated by a circle in Figure 2B). Positively charged side chains of many DNA binding proteins extrude to the DNA binding interface. On the other hand, the C_α atom linked to such a side chain tends to be buried in the interior of the protein. The charge of this type of residue is distributed to several surface beads in the RESPAC method. We speculate that this is one of the reasons why the discrepancy between integer-valued charges and RESPAC charges is most apparent for positive charges.

Next, we compared electrostatic potentials (ESP) and electrostatic fields (the gradient of ESP) around the lac repressor DNA binding domain calculated by three methods: (1) The AA ESP $\phi_{\mathrm{ref}}^{\mathrm{AA}}$ obtained by numerically solving the Poisson−Boltzmann equation using the atomic structure, which

dx.doi.org/10.1021/ct4007162 | J. Chem. Theory Comput. 2014, 10, 711−721

is used for the reference, (2) the CG ESP obtained using the Debye−Hückel formula with the RESPAC charges, and (3) the CG ESP obtained using the Debye−Hückel formula with integer-valued charges on amino acids. The comparison was done in the following three ways.

First, we quantified the similarity of the ESPs to the reference by the similarity score

$$\chi = 1 - \frac{\int_\Omega d^3\vec{r}\left[\phi_{\mathrm{ref}}^{\mathrm{AA}} - \sum_i q_i \frac{\exp(-\kappa|\vec{r}-\vec{r}_i|)}{\varepsilon|\vec{r}-\vec{r}_i|}\right]^2}{\int d^3\vec{r}|\phi_{\mathrm{ref}}^{\mathrm{PB}}|} \quad (8)$$

where $\vec{r}$ is the position vector of each grid, $\vec{r}_i$ is the position vector of the i-th CG bead, κ is the inverse of the Debye length, ε is the dielectric constant of the solvent and $\vec{q}$ is the RESPAC charge or integer charge. Computing χ for the RESPAC and integer-valued charges, we obtained χ = 0.846 for RESPAC charges and χ = 0.476 for the integer charges. This indicates that the overall fit of the CG ESP to the AA ESP is significantly better with the RESPAC charges than with the integer charges.

Second, in Figure 3A, we plotted ESP along the one-dimensional axis indicated by the arrow in Figure 3A. The origin of the arrow is the geometrical center of the protein and the tip of the arrow is 14 Å away from the origin, which is close to the protein surface. The arrow points to the DNA binding interface. We calculated three ESPs around the protein in a vacuum using atomic partial charges (PB), RESPAC CG charges (RESPAC), and integer CG charges (Integer). From this plot, we see that the AA ESP is well reproduced by the RESPAC ESP for regions more than 14 Å away from the origin, whereas the integer charge ESP deviated from the AA ESP by about $5kT/e$. Both RESPAC charges and integer charges fail to reproduce the AA ESP in regions closer than 14 Å to the origin. However, this region corresponds to the protein interior and cannot be reached by other molecules (see the tip of the arrow in Figure 3A). Therefore, this result suggests that the AA intermolecular electrostatic interaction is well described by the CG RESPAC charges.

Third, in Figure 3B, we show the three-dimensional iso-surfaces of electrostatic potential (ESP) and electrostatic field (the gradient of ESP) around the lac repressor DNA binding domain. In the top row in Figure 3B, we show iso-surfaces of the ESP around the lac repressor DNA binding domain calculated using AA charges ϕ_{AA} (left), RESPAC charges ϕ_{RESPAC} (center), and integer charges ϕ_{int} (right). The shape of the iso-surface of ϕ_{RESPAC} is more similar to that of ϕ_{AA} than is that of ϕ_{int}. We also note that the shape of the iso-surface of ϕ_{RESPAC} is more complicated than that of ϕ_{int}. We think this is because the number of RESPAC charges is larger than that of the integer charges. At the bottom left of this figure, we show the iso-surface of the difference $\phi_{\mathrm{RESPAC}} - \phi_{\mathrm{AA}}$ (red) and the difference $\phi_{\mathrm{int}} - \phi_{\mathrm{AA}}$ (blue). Comparison of these iso-surfaces reveals that the interior volume of the $\phi_{\mathrm{RESPAC}} - \phi_{AA}$ iso-surface is smaller than that of $\phi_{\mathrm{int}} - \phi_{\mathrm{AA}}$, suggesting that ϕ_{AA} is more closely represented by the RESPAC charges. In the middle row of Figure 3B, we show the iso-surfaces of the norm of the electrostatic field, E, around the lac repressor DNA binding domain calculated using the AA charges E_{AA} (left), the RESPAC charges E_{RESPAC} (center), and the integer charges E_{int} (right). The shape of the iso-surface of not only E_{RESPAC} but also E_{int} is similar to the shape of the iso-surface of E_{AA}. In the center of the bottom row of this figure, we show the iso-surface of $E_{\mathrm{RESPAC}} - E_{\mathrm{AA}}$ (red) and $E_{\mathrm{int}} - E_{\mathrm{AA}}$ (blue). This panel shows

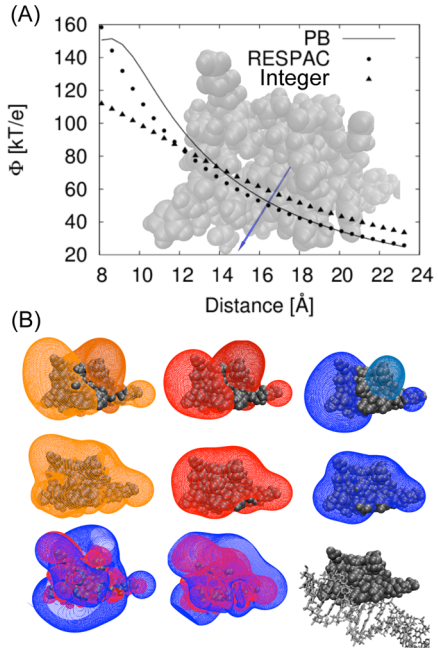


Figure 3. Comparison of the electrostatic potential for lac repressor DNA binding domain. (A) Electrostatic potential calculated with three different charge models [atomic charges (AA), RESPAC charges (RESPAC), and integer-valued charges (Integer)] along one-dimensional axis is plotted. In part A, we show the structure of the lac repressor DNA binding domain with a one-dimensional axis (blue arrow). The origin of the arrow is the geometrical center of the protein and the tip of the arrow is 14 Å away from the origin. (B) In the upper row, the $\pm 50[k_bT/e]$ iso-surface of electrostatic potential calculated using all atom charges ϕ_{AA} (left), RESPAC charges ϕ_{RESPAC} (center) and integer charges ϕ_{int} (right) are drawn around the structure of the lac repressor DNA binding domain. A negative surface is represented by darker color. In the middle row, the $50[k_bT/eA]$ iso-surfaces of the norm of the electrostatic field E calculated using all atom charges E_{AA} (left), RESPAC charges E_{RESPAC} (center), and integer charges E_{int} (right) are drawn. At the bottom left, the $\pm 20[k_bT/e]$ iso-surfaces of $\phi_{\mathrm{RESPAC}} - \phi_{\mathrm{AA}}$ (red) and $\phi_{\mathrm{int}} - \phi_{\mathrm{AA}}$ (blue) are drawn. At the bottom center, $\pm$ 1 $[k_bT/eA]$ iso-surface of $E_{\mathrm{RESPAC}} - E_{\mathrm{AA}}$ (red) and $E_{\mathrm{int}} - E_{\mathrm{AA}}$ (blue) are drawn. At the bottom right, the structure of the lac repressor DNA binding domain is shown with its specific DNA. In all the panels, the orientation of the structures is identical.

that the volume of $E_{\mathrm{RESPAC}} - E_{\mathrm{AA}}$ is smaller than that of $E_{\mathrm{int}} - E_{\mathrm{AA}}$, collectively suggesting that the absolute value of E_{AA} near the surface of the protein is represented better by RESPAC charges than by integer-valued charges, though the shape is similar.

Note that although the ESP at points near the protein surface cannot be accurately calculated using integer charges, the electrostatic field at points far from the protein surface is reasonably accurate. In MD simulations, this electrostatic field (i.e., force acting on a particle) is important (Figure 3A and the bottom−center drawing in Figure 3B). Therefore, integer-valued charges may approximately reproduce intermolecular interactions in certain cases.

Coarse-Grained Molecular Dynamics Simulation of lac Repressor. Using the RESPAC charges determined as described above, we performed CG simulation of the lac repressor DNA binding domain. From the trajectory data, we calculated and plotted the probability distribution of Q-score, which represents the ratio of the formed interbeads contacts to the natively formed contacts (Figure 4A). We consider that a

dx.doi.org/10.1021/ct4007162 | J. Chem. Theory Comput. 2014, 10, 711−721

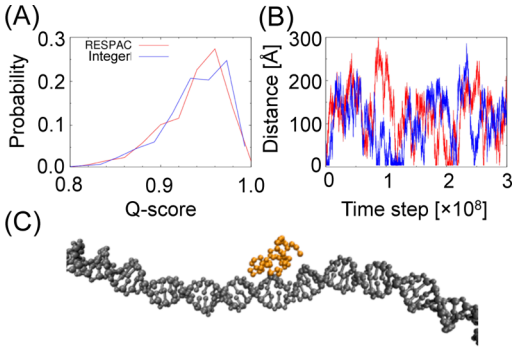


Figure 4. Coarse-grained (CG) simulations of lac repressor DNA binding domain. (A) Probability distributions of Q-score calculated from trajectories of CG simulation with RESPAC charges (red) and with integer charges (blue) are plotted. We consider that a bead pair forms a contact when $r < 1.2r_0$, where r represents the distance between two beads and r_0 represents the distance in the native structure. (B) Time trajectories of nearest distance between the protein and the DNA calculated from trajectories of CG simulation with RESPAC charges (red) and with integer charges (blue) are plotted. (C) A representative snapshot of the CG simulation trajectory. Orange and gray beads represent the protein and the double-strand DNA, respectively.

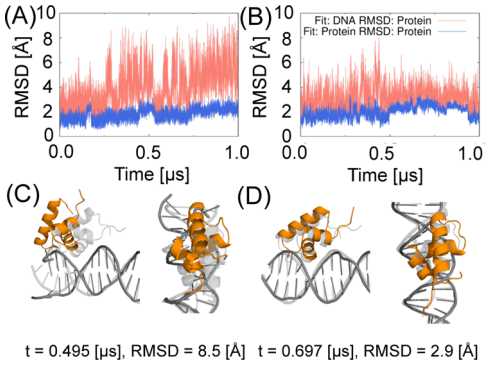


Figure 5. All atom (AA) simulations of lac repressor DNA binding domain. (A, B) Time trajectory of root-mean-square deviation (RMSD) of the protein calculated from trajectories of the AA simulation using the NMR nonspecific (A; PDB ID: 1OSL) and specific (B; PDB ID: 1L1M) complex monomer structure as an initial structure. The blue (red) line represents the RMSD of the protein with the protein (DNA) bead superimposed onto reference structure. (C, D) A representative snapshot from the nonspecific complex simulation.

bead pair forms a contact when $r < 1.2r_0$, where r represents the distance between two beads and r_0 represents the distance in the native structure. From this plot, we can see that the two probability distributions from the simulations using the RESPAC charges and the integer-valued charges, respectively, are nearly the same. This suggests that the overall protein stability is not significantly changed by the use of RESPAC charges. We also plotted the time trajectory of the nearest distance between the protein and the DNA in Figure 4B. From this plot, we see that the proteins with the integer-valued charges and with the RESPAC charges repeatedly dissociate and reassociate with similar affinity to DNA in the trajectory. Figure 4C shows a representative snapshot of the protein−DNA complex structure in the CG simulation, and reveals a similar nonspecific binding surface to that in the NMR structure, which prompted us to make a detailed comparison between these binding surfaces, as described.

All Atom Molecular Dynamics Simulation of lac Repressor. To address the validity of the nonspecific binding mode obtained using the RESPAC charges, we performed explicit solvent AA simulation using the NMR nonspecific complex monomer structure (PDB ID: 1OSL) as an initial structure. Note that, in both the NMR structures of the lac repressor DNA binding domain bound to its specific DNA and to nonspecific DNA sequences, two identical domains are covalently connected to form a homodimer to stabilize the bound structures.

In Figure 5A, we plotted the time series of the root-mean-square deviation (RMSD) measured from the initial structure. The blue curve represents the RMSD of the protein conformation, which stays below 3.0 Å, suggesting that the monomer protein structure is stable. On the other hand, the red curve is the RMSD when DNA is superimposed between the initial and the current structures, thus representing the movement of lac repressor relative to DNA, which gradually increases with time, reaching up to 10.0 Å. Thus, the binding is weak and dynamic. This result is consistent with the experimental observation that the lac repressor DNA binding domain cannot tightly bind to DNA if it does not form a dimer.[64]

In Figure 5C and D, we show representative snapshots in the simulations with a nonspecific DNA sequence. The snapshot in Figure 5C represents a transiently dissociated structure, which keeps some contacts with DNA, while losing others. Notably, this form is quite similar to one of the metastable states found in a recent atomic simulation study.[65] The close agreement of completely independent simulations with different force fields strongly suggests the robustness of this result. These fluctuations are reversible, and sometimes the protein returns to the more tightly bound structure shown in Figure 5D.

To confirm that the large fluctuation of the lac repressor structure is not due to the simulation setup or the force field, using the same setup, we performed another simulation using the specific DNA sequence (PDB ID: 1L1M as an initial structure). The results shown in Figure 5B show that the movement of lac repressor relative to DNA does not increase with time as much as those in the case of the nonspecific sequence (Figure 5A). The clear difference between the two simulations suggests that the high degree of movement in the first set is due to the nonspecific DNA sequence, which has weaker interactions with the protein.

Comparison of the Nonspecific DNA Binding Interface. Now, we compare the binding interface of the lac repressor with the nonspecific DNA obtained by the CG simulation with that obtained by the AA simulation. We plotted the contact probabilities of residues of the lac repressor DNA binding domain in Figure 6 (see Methods section in Supporting Information text for the definition of the contact probability). The contact probability in the AA simulation (gray bars in Figure 6) shows clear two peaks around residue ID 15−20 and around residue ID 30−35. On the other hand, the contact probability in the CG simulation with the integer charge set (blue points and line in Figure 6) has only one peak around residue ID 30−35, missing the first peak found in the AA simulation. In contrast, the contact probability in the CG simulation with the RESPAC charges (red points and line in Figure 6) shows two peaks at essentially the same regions as those in the AA simulation. Thus, clearly, the nonspecific binding interface of the AA simulation is much better

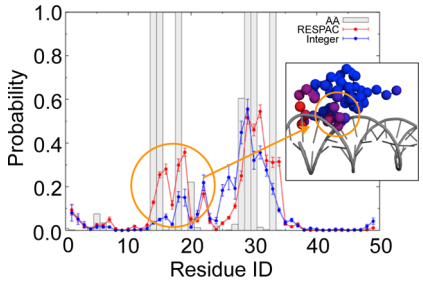


Figure 6. Comparison of nonspecific binding interface in AA simulation and CG simulation. Contact probabilities of each residue of lac repressor DNA binding domain calculated from trajectories of AA simulation (gray bars), CG simulation with RESPAC charges (red line and points), and CG simulation with integer-valued charges (blue line and points) are plotted. The average value and standard deviation (n = 10) are plotted for the results of CG simulation. In the inset, we show an NMR structure of the lac repressor DNA binding domain. Balls represent C_α atoms of the protein, and gray sticks represent double-strand DNA. Color and its intensity on each bead represent the contact probability (blue represents 0.0 and red represents 0.6) calculated from the trajectory of the CG simulation with RESPAC charges.

reproduced by the CG simulation with the RESPAC charges. This better reproduction of the binding interface is a clear qualitative improvement by the use of the RESPAC charges.

Although the position of the two peaks are nearly the same between the AA simulation and the CG simulation with the RESPAC charges, the magnitudes of the peaks are quite different. Probably this is largely due to the difference of the time scale. In a simple estimate of total length of simulation by comparing the theoretical diffusion constant of a bead of 15.0 Å (about the size of the lac repressor DNA binding domain) from the Stokes−Einstein equation with the diffusion constant of the lac repressor DNA binding domain in the CG simulation, we obtained a total CG simulation time corresponding to about 100 μs. In contrast, the AA simulation covers 1 μs.

We mapped the contact probability on the structure in the inset of Figure 6A. Red represents the high contact probability residues and blue represents the low contact probability ones. Of particular interest is the region in the orange circle, which represents the region where the high contact probability in the AA simulation is well reproduced only by the RESPAC charges. From Figure 2C, we see that the CG beads in this region have positive charges. These charges were introduced to reproduce the nearby ESP dominated by the charged side-chain atoms of Arg22 (blue small beads in Figure 2C). This illustrates that the RESPAC charges can fine-tune the electrostatic-driven nonspecific binding interface. Thus, proteins bind to nonspecific DNA with proper binding interfaces in CG simulations with RESPAC charges.

Application of RESPAC to 16 Target Proteins. Encouraged by the successful reproduction of the nonspecific binding interface of the lac repressor DNA binding domain using the CG simulation with the RESPAC charges, we applied the RESPAC charges to 16 other proteins listed in Figure 1A. To determine δ in eq 6 for each target protein, we prepared decoys by implicit solvent AA simulation. Using these decoys, we calculated $\langle \Delta(\vec{q}^{\,\mathrm{decoy}}(\delta)) \rangle_{\mathrm{decoy}}$ and plotted it against δ and decided δ for each target protein (Figure 1B). From this plot, we can see that the $\langle \Delta(\vec{q}^{\,\mathrm{decoy}}(\delta)) \rangle_{\mathrm{decoy}}$ values of most target proteins (e.g., 1A1L) do not strongly depend on δ. We can also see that the $\langle \Delta(\vec{q}^{\,\mathrm{decoy}}(\delta)) \rangle_{\mathrm{decoy}}$ values of some target proteins (e.g., 1J46, 1CDW, 1PP7) have a broad basin in the δ range from 10^5 to 10^6. Therefore, we speculate that nonoptimal fixed value $\delta = 5 \times 10^5$ can be used for proteins having various sizes and charge distributions without compromising the quality of fit in future applications, though the optimal value is used in this study. Using the fixed value of δ eliminates the need for making decoys and should facilitate the use of the RESPAC charges in future studies.

In order to quantitatively compare the nonspecific DNA binding interfaces of the 16 proteins obtained by the CG simulations using RESPAC charges with those obtained using integer-valued charges, in Figure 7A, we plot the structural

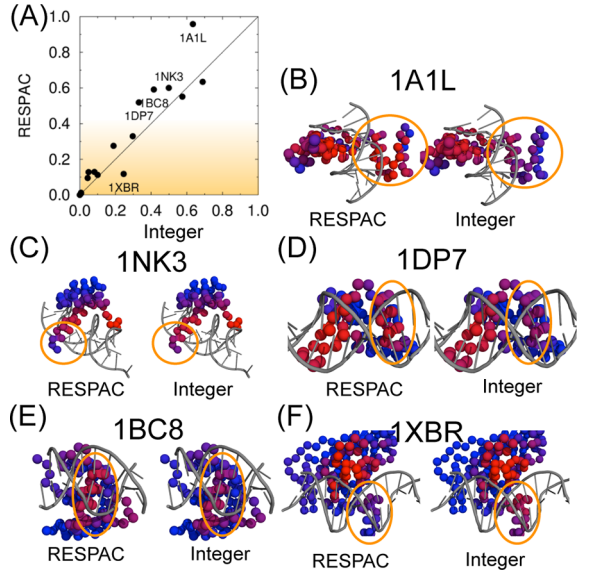


Figure 7. Comparison of nonspecific DNA binding interfaces obtained from CG simulation with RESPAC charges and with integer charges. (A) Structural similarity between the nonspecific DNA binding interface and the specific one was quantified (As similarity becomes high, the score approaches 1.0) and plotted. The similarity score from CG simulation with RESPAC charges is plotted against that from CG simulation with integer-valued charges. (B−F) We show structures of 1A1L, 1NK3, 1DP7, 1BC8, and 1XBR. Balls represent C_α atoms of the proteins, and gray sticks represent double-strand DNA. Color and its intensity on each bead represent the contact probability (Blue represents 0.0 and red represents 1.0) calculated from the trajectory of the CG simulation with RESPAC charges (left) and with integer charges (right).

similarity score obtained by using the RESPAC charges (on the y-axis) and that obtained by using the integer charges (on the x-axis). Briefly, as similarity between specific and nonspecific DNA binding interface becomes high, the score approaches 1.0 (see Methods section in Supporting Information text for the definition of the similarity score). Of the 16 cases, 4 proteins (PDB ID: 1A1L, 1NK3, 1BC8, 1DP7) showed significantly higher similarity by using the RESPAC charges, whereas the other 11 cases showed almost equivalent similarity scores between the two charge sets. Only in one case (PDB ID: 1XBR), the similarity obtained by using the RESPAC charges was lower than that obtained by using the integer-valued charges.

In order to clarify the difference between the nonspecific binding interface obtained using RESPAC charges and that obtained using integer charges, we mapped the contact probability of each residue bead to DNA on the crystal

structure (Figure 7B−F). As for 1A1L (Figure 7B) and 1DP7 (Figure 7D), there are broadly two binding interfaces to DNA, one of which recognizes the major groove and the other the minor groove. In both proteins, the interface that recognizes the major groove was found in CG simulations on the nonspecific DNA sequence with both types of charge sets. On the other hand, for both proteins, only the RESPAC charge set led to the contact at the interface that recognizes the minor groove (Encircled region in Figure 7BD). For 1NK3 (Figure 7C), the RESPAC charges make the region which does not specifically bind to DNA (Encircled region in Figure 7C) unbound in CG simulation. As for 1BC8 (Figure 7E), the RESPAC charges make the major groove-recognizing interface (Encircled region in Figure 7E) bind more tightly to nonspecific DNA than that obtained with the integer charges. As for 1XBR (Figure 7F), the nonspecific binding interface of this protein is dissimilar to the specific binding interface regardless of which charges are used. In addition, the groove-recognizing helix (Encircled region in Figure 7E) weakly binds to DNA when the RESPAC charge set is used, which makes the similarity score lower than integer charges (Figure 7A). Apart from the degree and direction of the change, Figure 7A shows that the RESPAC charge set changes the nonspecific binding interface of several proteins compared to that obtained with integer charges. It is noteworthy that there were several proteins whose similarity score was significantly low (Yellow colored region in Figure 7A) even with the RESPAC charges. 1XBR can be regarded as a representative of such proteins. Marcovitz et al. recently showed that the discrepancy between the specific and the nonspecific DNA binding interfaces slows the transition from nonspecific to specific DNA binding using in silico designed artificial peptides.[51,52] They also suggested that the degree of structural difference between the specific binding complex and the nonspecific binding complex of the natural protein is well correlated with the calculated binding interface discrepancy using CG simulation and static structural analysis with integer-valued charges. In the present work, using a more sophisticated charge assignment method, we showed that the difference between the specific binding and the nonspecific binding interface is significant in several proteins (similarity score is less than 0.2; encircled proteins in Figure 1A), supporting the earlier observation. The degree of the discrepancy could have been somewhat overestimated or underestimated in the previous study depending on the protein.

CONCLUDING REMARKS

In previous studies, the treatment of electrostatics in the one bead per one amino acid CG model was not very sophisticated. Therefore, one of the keys to extending the applicable range of CG simulation is improvement of the charge arrangement and consequently greater sophistication of the electrostatics treatment. In this work, following the same philosophy as that underlying the RESP, which is widely used to derive atomic partial charges, we developed a new ESP fitting method to derive the effective charges on CG beads so that the ESP of the CG model is fitted with that of the AA model, which we termed the RESPAC method. Comparison of calculated ESPs around the lac repressor DNA binding domain suggests that the intermolecular electrostatic interaction is well described by the RESPAC charges and the accuracy is comparable to that of the atomic charge distribution. Our findings suggested that integer-value charges could provide modest precision that approximately reproduced intermolecular interactions, though accurate reproduction of a binding interface required the more sophisticated RESPAC charges. We performed AA simulations and CG simulations of the lac repressor DNA binding domain, and compared their binding interfaces. The results suggest that the nonspecific binding interface of the AA simulation of this protein is well reproduced by the CG simulation with RESPAC charges but not by the CG simulation with integer charges.

Encouraged by the successful reproduction of the nonspecific binding interface of the lac repressor DNA binding domain using the CG simulation with RESPAC charges, we calculated the RESPAC charges of 16 other target proteins. In this process, we noticed that nonoptimal fixed value $\delta = 5 \times 10^5$ (scale factor in eq 6) can be used for proteins having various size and charge distribution without compromising the quality of fit. This fixed value of δ should facilitate the use of RESPAC charges in future studies. We also performed CG simulations of these proteins with RESPAC charges at room temperature and at physiological ion concentration, resulting in a smaller binding interface discrepancy than that in the simulation with the integer-valued charges. In addition, we predicted that the nonspecific DNA binding interface of several proteins is quite different from the specific DNA binding interface, indicating, in accord with a recent study,[51,52] that the transition of nonspecific to specific DNA binding of some natural proteins would actually be slowed as a consequence.

In this work, we improved the way of putting charges on each CG bead. We think that the main basis for this improvement is that we distribute charges of the side chain in the atomic structure onto the C_α beads around it. Therefore, we speculated that, if we used the CG model with side chain beads and put integer-valued or RESPAC charges on them, we could fit the electrostatic field in more detail. However, it is also valuable to fit the electrostatic field using a model as simple as possible, such as the one used in this work because such a model can be used to simulate large systems without extensive computational cost. Further improvement of the electrostatics treatment in the CG model is thought likely to require a more sophisticated model than the Debye−Hückel model, which is most widely used in this field, because this model lacks the treatment of self-energy and a dielectric constant depending on the molecular environment. These flaws make it difficult to perform a CG simulation of intrinsically disordered regions that frequently and markedly change their conformation. Some of these regions even fold only when binding to DNA. Considering that about 30% of all proteins and about 90% of transcription factors possess extended regions of intrinsic disorder,[66] it is highly desirable to develop a way to treat the electrostatics of such regions in CG simulations in order to elucidate biologically relevant processes such as the searching mechanism of transcription factors. There are several CG electrostatic interaction models that allow the structure to deform, best exemplified by the topologically based multipolar reconstruction (TBMR) model proposed previously.[48−50] However, their accurate reproduction of the electrostatic field around a molecule strongly depends on sophisticated modeling of dipole moments. Therefore, expansion of this scheme to nucleic acids has not been achieved so far and should be approached in future work.

Although there is plenty of room for improvement, as a new standard method of deciding the charge in the one bead per one amino acid model, RESPAC is applicable to a broad range of studies where a protein does not change its conformation so markedly and electrostatic interaction plays an important role.

Furthermore, charges derived with this method can be used even in the more sophisticated electrostatics treatments to be developed in the future.

■ ASSOCIATED CONTENT

Supporting Information

Additional methods. 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. Chang Lee for providing the van dar Waals surface calculation program. The calculations in this work were partly performed by using the supercomputer of ACCMS, Kyoto University. T.T. acknowledges the support of a JSPS fellowship. This work was supported by Grant-in-Aid for JSPS Fellows, by Grant-in-Aid for Scientific Research on Innovative Areas, by the Global COE Program "Formation of a Strategic Base for Biodiversity and Evolutionary Research: from Genome to Ecosystem" of the Ministry of Education Culture, Sports, Science, and Technology (MEXT) and by the Excellent Graduate Schools Program "Biodiversity and Evolution from Genome to Ecosystem" of the MEXT.

■ REFERENCES

(1) Winkler, F. K.; Banner, D. W.; Oefner, C.; Tsernoglou, D.; Brown, R.; Heathman, S.; Bryan, R.; Martin, P.; Petratos, K.; Wilson, K. *EMBO J.* **1993**, *12*, 1781−1795.

(2) Misra, V. K.; Hecht, J. L.; Yang, A.-S.; Honig, B. *Biophys. J.* **1998**, *75*, 2262−2273.

(3) Viadiu, H.; Aggarwal, A. K. *Mol. Cell* **2000**, *5*, 889−895.

(4) Kalodimos, C. G.; Biris, N.; Bonvin, A. M.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. *Science* **2004**, *305*, 386−389.

(5) Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2006**, *128*, 14506−14518.

(6) Dahirel, V.; Paillusson, F.; Jardat, M.; Barbi, M.; Victor, J. M. *Phys. Rev. Lett.* **2009**, *102*, 228101−228101.

(7) Dror, R. O.; Dirks, R. M.; Grossman, J.; Xu, H.; Shaw, D. E. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.

(8) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2012**, 58−65.

(9) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915−5920.

(10) Kamerlin, S. C.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41−64.

(11) Hyeon, C.; Thirumalai, D. *Nat. Commun.* **2011**, *2*, 487.

(12) Takada, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137.

(13) Vuzman, D.; Levy, Y. *Mol. BioSyst.* **2012**, *8*, 47−57.

(14) Saunders, M. G.; Voth, G. A. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.

(15) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937−953.

(16) Karanicolas, J.; Brooks, C. L., III *J. Mol. Biol.* **2003**, *334*, 309−325.

(17) Prieto, L.; de Sancho, D.; Rey, A. *J. Chem. Phys.* **2005**, *123*, 154903.

(18) Azia, A.; Levy, Y. *J. Mol. Biol.* **2009**, *393*, 527−542.

(19) Li, W.; Wolynes, P. G.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3504−3509.

(20) Li, W.; Terakawa, T.; Wang, W.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17789−17794.

(21) Kim, Y. C.; Hummer, G. *J. Mol. Biol.* **2008**, *375*, 1416−1433.

(22) Zarrine−Afsar, A.; Zhang, Z.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. *Proteins* **2012**, *80*, 858−870.

(23) Okazaki, K.-i.; Sato, T.; Takano, M. *J. Am. Chem. Soc.* **2012**, *134*, 8918−8925.

(24) Chu, X.; Wang, Y.; Gan, L.; Bai, Y.; Han, W.; Wang, E.; Wang, J. *PLoS Comput Biol* **2012**, *8*, e1002608.

(25) Terakawa, T.; Kenzaki, H.; Takada, S. *J. Am. Chem. Soc.* **2012**, *134*, 14555−14562.

(26) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592−601.

(27) Cox, S.; Williams, D. *J. Comput. Chem.* **1981**, *2*, 304−323.

(28) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.

(29) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905.

(30) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.

(31) Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367−383.

(32) Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith, V. H. *J. Comput. Chem.* **1990**, *11*, 297−310.

(33) Stouch, T.; Williams, D. E. *J. Comput. Chem.* **1992**, *13*, 622−632.

(34) Merz, K. M. *J. Comput. Chem.* **1992**, *13*, 749−767.

(35) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(36) Woods, R.; Chappelle, R. *J. Mol. Struct.* **2000**, *527*, 149−156.

(37) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048−1057.

(38) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 7300−7307.

(39) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacK-erell, A. D. *J. Chem. Theory Comput.* **2005**, *1*, 153−168.

(40) Okiyama, Y.; Watanabe, H.; Fukuzawa, K.; Nakano, T.; Mochizuki, Y.; Ishikawa, T.; Tanaka, S.; Ebina, K. *Chem. Phys. Lett.* **2007**, *449*, 329−335.

(41) Chen, D.−L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 10225−10233.

(42) Zeng, J.; Duan, L.; Zhang, J. Z.; Mei, Y. *J. Comput. Chem.* **2012**, *34*, 847−853.

(43) Chang, L.; Ishikawa, T.; Kuwata, K.; Takada, S. *J. Comput. Chem.* **2013**, *34*, 1251−1257.

(44) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461−470.

(45) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868−3878.

(46) Zhang, Q.; Beard, D. A.; Schlick, T. *J. Comput. Chem.* **2003**, *24*, 2063−2074.

(47) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B* **2007**, *111*, 9390−9399.

(48) Cascella, M.; Neri, M. A.; Carloni, P.; Peraro, M. D. *J. Chem. Theory Comput.* **2008**, *4*, 1378−1385.

(49) Alemani, D.; Collu, F.; Cascella, M.; Dal Peraro, M. *J. Chem. Theory Comput.* **2009**, *6*, 315−324.

(50) Spiga, E.; Alemani, D.; Degiacomi, M. T.; Cascella, M.; Dal Peraro, M. *J. Chem. Theory Comput.* **2013**, *9*, 3515−3526.

(51) Marcovitz, A.; Levy, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17957−17962.

(52) Marcovitz, A.; Levy, Y. *J. Phys. Chem. B* **2013**, *117*, 13005−13014.

(53) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(54) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, 522−525.

(55) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(56) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351−371.

(57) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(58) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(59) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845−854.
(60) Sambriski, E.; Schwartz, D.; De Pablo, J. *Biophys. J.* **2009**, *96*, 1675−1690.
(61) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K.-i.; Yao, X.-Q.; Takada, S. *J. Chem. Theory Comput.* **2011**, *7*, 1979−1989.
(62) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(63) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(64) Kalodimos, C. G.; Folkers, G. E.; Boelens, R.; Kaptein, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6039−6044.
(65) Yonetani, Y.; Kono, H. *J. Phys. Chem. B* **2013**, *117*, 7535−7545.
(66) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2006**, *45*, 6873−6888.