Explicit ions/implicit water generalized Born model for nucleic acids

Igor S. Tolokh,[1] Dennis G. Thomas,[2] and Alexey V. Onufriev[3]
[1]*Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061, USA*
[2]*Computational Biology, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, USA*
[3]*Departments of Computer Science and Physics, Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*

(Received 28 February 2018; accepted 23 April 2018; published online 18 May 2018)

The ion atmosphere around highly charged nucleic acid molecules plays a significant role in their dynamics, structure, and interactions. Here we utilized the implicit solvent framework to develop a model for the explicit treatment of ions interacting with nucleic acid molecules. The proposed explicit ions/implicit water model is based on a significantly modified generalized Born (GB) model and utilizes a non-standard approach to define the solute/solvent dielectric boundary. Specifically, the model includes modifications to the GB interaction terms for the case of multiple interacting solutes—disconnected dielectric boundary around the solute-ion or ion-ion pairs. A fully analytical description of all energy components for charge-charge interactions is provided. The effectiveness of the approach is demonstrated by calculating the potential of mean force for Na^+–Cl^- ion pair and by carrying out a set of Monte Carlo (MC) simulations of mono- and trivalent ions interacting with DNA and RNA duplexes. The monovalent (Na^+) and trivalent ($CoHex^{3+}$) counterion distributions predicted by the model are in close quantitative agreement with all-atom explicit water molecular dynamics simulations used as reference. Expressed in the units of energy, the maximum deviations of local ion concentrations from the reference are within k_BT. The proposed explicit ions/implicit water GB model is able to resolve subtle features and differences of CoHex distributions around DNA and RNA duplexes. These features include preferential CoHex binding inside the major groove of the RNA duplex, in contrast to CoHex biding at the "external" surface of the sugar-phosphate backbone of the DNA duplex; these differences in the counterion binding patters were earlier shown to be responsible for the observed drastic differences in condensation propensities between short DNA and RNA duplexes. MC simulations of CoHex ions interacting with the homopolymeric poly(dA·dT) DNA duplex with modified (de-methylated) and native thymine bases are used to explore the physics behind CoHex-thymine interactions. The simulations suggest that the ion desolvation penalty due to proximity to the low dielectric volume of the methyl group can contribute significantly to CoHex-thymine interactions. Compared to the steric repulsion between the ion and the methyl group, the desolvation penalty interaction has a longer range and may be important to consider in the context of methylation effects on DNA condensation. *Published by AIP Publishing.* https://doi.org/10.1063/1.5027260

I. INTRODUCTION

Ions play essential roles in governing the structure, dynamics, and function of nucleic acids.[1–7] For example, they stabilize the conformations of nucleic acids and modulate their interactions with ligands and proteins. Multivalent cations are able to induce the inter-helix attraction in DNA molecules resulting in DNA aggregation.[8–10] Differences in ion binding patterns between DNA and RNA can lead to profound differences in how their mechanical properties respond to changes in the ion concentration.[11] Subtle differences in how tetravalent spermine polyions bind to unmethylated vs. methylated DNA may help understand how these epigenetic markers affect chromatin compaction.[12] Therefore, an accurate description of interactions of ions with DNA and RNA molecules and between themselves is of fundamental importance in our understanding of the behavior of nucleic acids and their complexes.

All-atom explicit solvent molecular dynamics (MD) simulations allow one to quantify thermodynamic properties and structural details of ion-nucleic acid interactions at atomic resolution. These simulations can reveal mechanisms underlying many biologically important, yet completely unintuitive phenomena, often not accessible experimentally. However, on microsecond time scales that may be needed for adequate sampling of large nucleic acid systems in the presence of complex multivalent ions with strong binding affinities, these simulations can be very demanding.[13] For much larger systems such as the nucleosome, time scales beyond 1 μs[14] are currently outside the reach of traditional explicit solvent atomistic simulations. Simulation of chromatin components such as nucleosome arrays is even more computationally demanding, while conformational preferences in these systems are very sensitive[15] to the presence of multi-valent ions such as Mg^{2+} or $CoHex^{3+}$. Multi-scale modeling[16–18] of chromatin that can retain the realism of the

explicit treatment of multi-valent ions remains a significant challenge.

At the same time, simulations based on the implicit solvent framework[19–28] can extend the effective simulation times by one to two orders of magnitude due to enhanced conformational sampling,[29] with further speed-ups possible through algorithmic advances.[30,31] Within this framework, the solvent is treated as a dielectric continuum interacting with the solutes through electrostatic and dispersion forces. A widely used method of estimating the solvent electrostatic contribution to the system free energy is by solving the Poisson-Boltzmann (PB) equation. Advanced numerical algorithms for solving this equation have been developed,[32–36] and proof-of-concept MD simulations based on the numerical PB had been reported.[37–40] Specifically, an explicit ions/implicit water model based on the numerical PB approach was proposed[41] and was used to simulate short DNA and RNA fragments interacting with monovalent ions. The model, however, was not tested with multivalent ions, which is our focus here. In general, it is believed that the algorithmic complexity of the numerical PB approach, conceptual difficulties associated with incorporating PB-based algorithms into Molecular Dynamics,[42] and significant computational expense[43] preclude its wide adoption in dynamics.

Instead, the most widely used practical implicit solvent model in Molecular Dynamics is the generalized Born (GB) approximation, which represents a compromise between the computational efficiency and accuracy. Some of the GB models approach the accuracy of the numerical PB,[43] including ligand binding calculations,[44,45] which is directly relevant to our goals here. The GB can be rigorously derived from the PB for a special case of a perfectly spherical single solute;[28,46] for an arbitrary molecular shape, the GB can be considered an approximation to the PB. As such, the GB model inherits many of the approximations to reality already present in the PB model and adds several more. The effect of the associated inaccuracies on the outcomes of practical GB-based simulations is system specific: below, we review the ones most critical to highly charged nucleic acids interacting with multivalent ions.

Both classical PB and GB models suffer from inaccuracies in treating ion atmosphere around, and estimating solvation free energies of highly charged solutes.[41,47–49] This is because both models utilize a mean-field description of the ions and, therefore, do not account for ion-ion correlations or discreteness of ions near the charged solute surface. While in the case of monovalent ions, the correlation effects are small and can often be neglected, the correlations between multivalent ions can introduce significant corrections to ion distributions and electrostatic potentials around solutes.[50–54] Moreover, both classical PB and GB models, when treating ions implicitly, neglect, or at best only crudely approximate, ion desolvation effects which appear when ions approach the low dielectric solute region. We note that an approach to approximating ion desolvation penalty directly within the PB formalism has been recently proposed.[55] Nevertheless, without explicitly considering multivalent ions tightly bound to the macromolecule,[56–58] one can not expect from the "classical" PB/GB approach an accurate quantitative description of the electrostatic potentials and ion distributions around the solutes, critical for related condensation phenomena. To sum up, while the mean-field atmosphere of monovalent ions around DNA can be fairly well described by the classical PB [with an appropriate choice of the dielectric boundary (DB)], the model fails to accurately describe many important aspects of the distribution of multivalent ions around charged solutes.

Another deficiency of both the PB or GB models originates from their consideration of only the dipole polarization of the solvent. In the case of water, the higher electric multipole moments of water molecules lead to multiple effects, e.g., charge hydration asymmetry (CHA). The neglect of these effects by the GB/PB can result in a significant deviation from reality of the hydration free energies of ions,[59] and even neutral polar molecules,[60] unless the definition of the dielectric boundary (DB) is adjusted to account for these effects.[61,62]

Finally, we have discovered in the course of this work that the "canonical" GB model[63] has a specific flaw absent at the PB level, at least in principle. Namely, the canonical GB model is not designed to handle disconnected (discontinuous) dielectric boundaries, that is, solutes which are not topologically equivalent to a sphere. For example, in the limit of infinitely high solvent dielectric, the numerical PB correctly predicts the zero interaction energy between two ions separated by a distance larger than the diameter of the water molecule, while the GB incorrectly predicts a non-zero interaction. This limitation is of little consequence in many practical simulations, e.g., of a single solute, but becomes a critical flaw in the case of multiple highly charged ions interacting, but not permanently associating, with a solute; see Fig. 1.

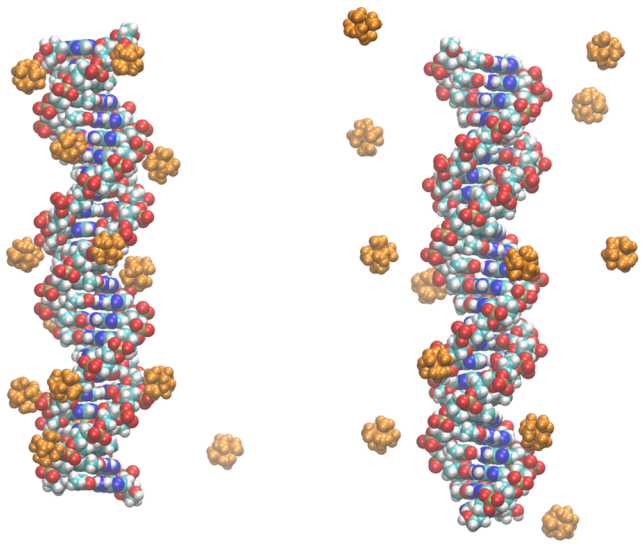

FIG. 1. Qualitatively different distributions of multivalent $CoHex^{3+}$ ions around the DNA duplex in the explicit (left) and implicit (right) water. Shown are the representative configurations from an all-atom MD simulation of explicit CoHex ions around a 25 base pair long DNA duplex. Ion distribution resulting from the standard explicit water treatment (TIP3P,[64] left) is compared to the one based on one of the latest GB models[65] available in AMBER[66] (right).

To address the above deficiencies of the PB/GB conceptual approach while retaining the advantage of the GB in dynamics applications, here we develop and test an explicit ions/implicit water GB-based model that is able to accurately describe the ion atmosphere around charged nucleic acid molecules. The model considers ions as explicit particles and directly estimates all the interactions between them and nucleic acid molecules. At the same time, the solvent effects are treated implicitly using a modified GB model for the electrostatic interactions and the effective surface tension model for the non-polar effects. Since the results of the PB and GB models are rather sensitive to the choice of the solute/solvent dielectric boundary, especially in the case of highly charged solutes, we develop here a non-standard approach to define the DB around ions and nucleic acids which preserve the geometry of the Lee-Richards[67] solvent accessible surface (SAS) within physically justified limits based on the experimentally determined atomic and ionic radii. The "disconnected boundary" defect of the GB model described above is addressed.

II. THEORY

A. Canonical generalized Born (GB) model: A brief introduction

The key component of the interaction potential between solutes in the presence of high dielectric solvent (e.g., water) is the change of the solvation free energy of the complex. The solvation energy ΔG_{solv} can be represented as a sum of electrostatic ΔG_{el} and non-polar $\Delta G_{\text{nonpolar}}$ contributions,[20]

$$\Delta G_{\text{solv}} = \Delta G_{\text{el}} + \Delta G_{\text{nonpolar}}, \tag{1}$$

which are estimated independently.

The generalized Born (GB) model provides an approximate way [relative to the exact solution the Poisson equation (PE)] to calculate the electrostatic part, ΔG_{el}, of the solvation free energy of any singly connected solute or tightly bound complex.[28] The canonical GB approximation is based on the equation proposed by Still *et al.*,[63]

$$\Delta G_{\text{el}} = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{d_{ij}^2 + R_i R_j \exp\left(-d_{ij}^2/(4R_iR_j)\right)}}, \tag{2}$$

where ϵ_{in} and ϵ_{out} are the dielectric constants of the solute and the solvent, respectively, d_{ij} is the distance between solute atoms i and j, and q_i are the atomic charges. The main parameters of the equation are *the effective Born radii* R_i, with R_i^{-1} characterizing the average degree of solvent exposure of atom i.

Among the variety of methods and models to calculate R_i (see, e.g., Ref. 28 for review), we will use the so-called "R6" GB model,[68] which is theoretically well-grounded[68,69] and, at the same time, among the most accurate[70] GB models, including the calculation of binding energies.[45] The R6 model evaluates R_i through the integration of r^{-6} function centered on atom i over the volume V_{ext}, outside the solute dielectric boundary (DB),

$$R_i^{-1} = \left(\frac{3}{4\pi}\int\limits_{V_{ext}}\frac{dV}{|\mathbf{r}-\mathbf{r}_i|^6}\right)^{\frac{1}{3}}, \tag{3}$$

where $\mathbf{r}_i$ is the radius-vector of atom i. This approach assumes that DB is a surface separating low dielectric region of a solute from the high dielectric domain of a solvent.

Equation (3) can be transformed to a more computationally convenient form,

$$R_i^{-3} = \rho_i^{-3} - \frac{3}{4\pi}\int\limits_{V_{int}}\frac{dV}{|\mathbf{r}-\mathbf{r}_i|^6}, \tag{4}$$

where the region of integration V_{int} is now inside the solute DB but outside the sphere of radius ρ_i centered on atom i, i.e., $|\mathbf{r}-\mathbf{r}_i| \geq \rho_i$ and $\mathbf{r} \notin V_{ext}$. The values of ρ_i are usually set equal to the atomic (van der Waals) radii of solute atoms, which are typically used to set the dielectric surface of a solute.

In the case of two non-bonded solute atoms i and j in a solvent (see Fig. 2), the integration volume V_{int} for calculating R_i consists of two parts: the sphere around atom j determined by the atomic dielectric radius ρ_j, and the region between the two atoms inaccessible for a solvent probe of radius ρ_w (so called the "Neck" region). The formation of the "Neck" region at small atom-atom separations reflects a discrete molecular structure of the solvent and causes a partial desolvation of the two atoms (ions). This desolvation leads to a substantial deviation of the electrostatic interactions between the two atomic size charges in a solvent from the Coulomb law. This effect can be approximated within the implicit solvent approach in several distinct ways.[34,73,74] Here we follow a geometric approach introduced in Ref. 74 in the context of the GB model.

The results of the r^{-6} integration over the two parts of V_{int} in Eq. (4) are called "Sphere," $I_{ij}^{(S)}$, and "Neck," $I_{ij}^{(N)}$, integrals, respectively, and Eq. (4) can be rewritten as

$$R_i^{-3} = \rho_i^{-3} - \frac{3}{4\pi}\left(I_{ij}^{(N)} + I_{ij}^{(S)}\right). \tag{5}$$

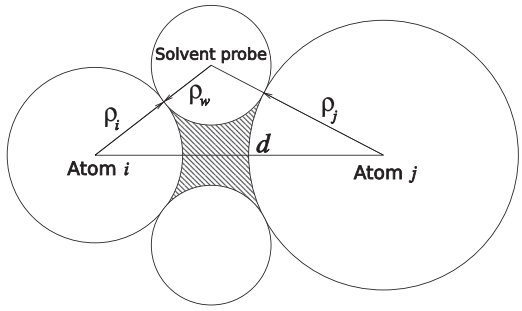

FIG. 2. Sketch of the dielectric boundary (DB) around a pair of non-bonded solute atoms (ions) in a solvent, separated by a distance d. The DB is defined as a surface around the volume inaccessible to a solvent probe of radius ρ_w [Richards-Connolly[71,72] molecular surface (MS)]. This volume consists of two atomic spheres determined by atomic dielectric radii ρ_i and ρ_j and a region between these spheres (depicted as dashed area) inaccessible to the solvent probe and called the "Neck" region.

The "Sphere" integral in Eq. (5) can be estimated analytically as[75–77]

$$I_{ij}^{(S)}(d) = \frac{4\pi\rho_j^3}{3(d^2-\rho_j^2)^3}, \quad d > \rho_i + \rho_j$$

$$I_{ij}^{(S)}(d) = \frac{\pi}{12d}\left(\frac{d+3\rho_j}{(d+\rho_j)^3} + \frac{3(\rho_j^2-\rho_i^2-(d-\rho_i)^2)+2d\rho_i}{\rho_i^4}\right), \quad |\rho_j-\rho_i| < d < \rho_i+\rho_j\,. \tag{6}$$

The "Neck" integral in Eq. (5) can be approximated analytically using the following equations[77]

$$\begin{aligned} &I_{ij}^{(N)}(d) = 0\,,\ d < \alpha_{ij},\\ &I_{ij}^{(N)}(d) = A_{ij}(d-\alpha_{ij})^4(\beta_{ij}-d)^4\,,\ \alpha_{ij} < d < \beta_{ij}\,,\\ &I_{ij}^{(N)}(d) = 0\,,\ d > \beta_{ij}\,. \end{aligned} \tag{7}$$

Here, parameter A_{ij} sets the value of the maximum of $I_{ij}^{(N)}(d)$ at $d = d_{ij}^{max}$ (both quantities can be tabulated by using the corresponding numerical integration over the "Neck" region for a given pair of solute atoms), $\beta_{ij} = \rho_i + \rho_j + 2\rho_w$ is the smallest atom-atom distance at which the solvent probe can fit between the two dielectric spheres, and $\alpha_{ij} = d_{ij}^{max} - (\beta_{ij} - d_{ij}^{max})$ is the separation at which the symmetrical analytical function $I_{ij}^{(N)}(d)$ vanishes.

The non-polar term $\Delta G_{\text{nonpolar}}$ in Eq. (1) can be regarded as the free energy cost to create a cavity in a solvent to accommodate the solute and van der Waals interactions of the solute with water. We follow the widely used approach[20,78–82] assuming that $\Delta G_{\text{nonpolar}}$ is linear with the area of the Lee-Richards solvent accessible surface (SASA) of a solute,

$$\Delta G_{\text{nonpolar}} = \gamma \text{SASA}, \tag{8}$$

where γ is an effective microscopic surface tension parameter. In the case of an ion-ion or ion-atom pair, the SASA of each ion can change from its value $\text{SASA}_i = 4\pi(\rho_i+\rho_w)^2$ for the fully solvated isolated ion i (when $d > \rho_i + \rho_j + 2\rho_w$) to a reduced value for the partially desolvated ion, when the SASs of the two ions (ion and atom) overlap,

$$\text{SASA}_i(d) = 2\pi(\rho_i+\rho_w)^2\left(1 + \frac{d^2+(\rho_i+\rho_w)^2-(\rho_j+\rho_w)^2}{2d(\rho_i+\rho_w)}\right), \quad d \le \rho_i+\rho_j+2\rho_w\,. \tag{9}$$

B. Effective charge-charge interactions in implicit solvent

The ability to estimate ΔG_{solv} using the implicit solvent approach, in particular the GB model, gives us a way to estimate the interactions between solutes in a solvent. Below we present standard equations for calculating the interaction free energy between two charged atomic particles in a solvent. The equations can be easily generalized for more than two particles, i.e., for the interactions between a many-atom molecular solute and ions. The approach described below is widely used to estimate a free energy of ligand biding.[70] The assumptions involved are described in Ref. 22.

The interaction free energy $W_{ij}(d)$ of two atomic charges i and j in an implicit solvent separated by a distance d can be approximated as a sum of their interaction energy in "vacuum," $E^{(ij)}$, and the change of their solvation free energy, $\Delta G_{\text{solv}}^{(ij)}$, relative to the solvation free energy at infinite separation, $\Delta G_{\text{solv}}^{(i)} + \Delta G_{\text{solv}}^{(j)}$,

$$W_{ij}(d) = E^{(ij)}(d) + \Delta G_{\text{solv}}^{(ij)}(d) - (\Delta G_{\text{solv}}^{(i)} + \Delta G_{\text{solv}}^{(j)}). \tag{10}$$

Generalized to the many-atom solute, this is the equation that governs the energetics of ion-DNA interactions in our MC simulations below.

Excluding the electronic polarization component, "vacuum" interaction energy $E^{(ij)}(d)$ in Eq. (10) can be presented as a sum of Coulombic and van der Waals interactions,

$$E^{(ij)}(d) = E_{el}^{(ij)}(d) + E_{vdW}^{(ij)}(d) = \frac{q_i q_j}{\epsilon_{in} d} + \varepsilon_{ij}\left(\left(\frac{\tilde{R}_{ij}}{d}\right)^{12} - 2\left(\frac{\tilde{R}_{ij}}{d}\right)^{6}\right), \tag{11}$$

where ε_{ij} and $\tilde{R}_{ij}$ are parameters of van der Waals interaction in the form of the Lenard-Jones (LJ) potential.

III. RESULTS: MAJOR MODIFICATIONS OF THE GB MODEL IN THE PROPOSED EXPLICIT IONS/IMPLICIT WATER APPROACH

A. Solute-solvent dielectric boundary (DB)

One of the key steps in an implicit solvent approach, including the GB and PE/PB methods, is a construction of the dielectric boundary (DB)—the region that separates the low dielectric solute domain from the high dielectric (in the case of water) domain of the bulk solvent.[21,28,32,34,83] This transition region is not an observable entity on the microscopic level and depends on the dielectric model. For example, the region can be described by a smooth dielectric function $\epsilon(\mathbf{r})$ derived from a Gaussian-based approach,[34,84] or it can be represented by a sharp boundary (the solute/solvent dielectric surface), which separates the solute domain with a dielectric constant ϵ_{in} from the solvent domain with ϵ_{out}. Here we use the sharp DB as the basis for our model.

In the GB method, the DB is crucial in determining the effective Born radii, Eq. (3), the main parameters in the work of Still *et al.* Eq. (2).[43,63,69,70] Outcomes of implicit solvent calculations are extremely sensitive to details of the DB.[83,85] An optimal DB geometry absorbs many approximations made by the specific continuum solvent model.

Two of the most widely used definitions of the solute/solvent DB in the GB approximation are the Richards-Connolly solvent excluded surface (SES)[71,72] [also called the molecular surface (MS)] and the van der Waals surface (VDWS).[85,86] To specify the geometry of DB, both definitions (SES and VDWS) commonly use atomic radii sets. In the case of water, the standard value 1.4 Å of a water molecule radius is used for as a water probe radius for SES. One of the earliest and widely used sets is the *bondi* van der Waals radii set determined from the experiments.[87] Its modified version, *mbondi*,[88] is currently standard in AMBER. Other radii sets have the values of atomic radii optimized for specific types of solutes.

For many solutes, SES and VDWS based DB surfaces produce reasonable values of the solvation free energies.[43,83,86,89]

But in the case of ions or salt bridges, the substantial deviations in these energies from their experimental values were observed with the realistic sizes of the atomic and ionic radii.[47,90,91] To improve the agreement of PB based ion-ion interactions with the reference interaction site model (RISM)[92] or the MD[93] results, the cationic radii in Refs. 90 and 91 had to be substantially increased compared to their values derived from the experiment.[94–96] Simultaneously, a water probe radius was decreased from its standard value 1.4 Å to 0.8 Å.[90] These optimized values of ionic and solvent probe radii are very different from their experimentally determined values and are difficult for physical interpretation.[62] They also distort the geometry of SAS needed for estimation of the nonpolar part of the solvation energy, e.g., Eq. (8). It was suggested that the origin of this significant difference is attributed to higher order electric multipole moments of water molecules[59] and is beyond the dipole polarization approximation in the implicit solvent model.

To avoid the potential difficulties in physical interpretation of the radii parameters used to determine the DB and SAS, we propose to distinguish between the two types of the solute/solvent interfaces, Fig. 3: (1) The interface described by the SES determined with the standard sets of experimental van der Waals atomic (ionic) radii a_i and standard (1.4 Å) water probe size r_w. This interface can be interpreted as the boundary between the solute and solvent electron densities and is used to build the SAS. (2) The DB which, in the case of a "two-dielectric" model, is treated as the surface of the localization of induced solvent polarization charges. The proposed DB is neither VDWS nor SES with the standard water probe of 1.4 Å but is the surface derived from the standard SAS geometry using an effective water probe dielectric radius ρ_w. By construction, the DB can be built as an effective SES using effective atomic (ionic) and water probe dielectric radii, ρ_i and ρ_w, which maintain the geometry of the standard SAS.

We are going to show that by taking into account the charge distribution in water at the solute/solvent interface one can construct the DB which, being used in the standard GB model, is capable of reproducing the electrostatic part of ΔG_{solv} for a cation-anion pair on the level of accuracy comparable with the explicit water simulation results.

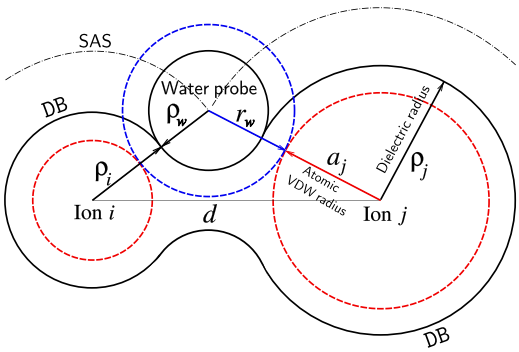

FIG. 3. The geometry of the proposed dielectric boundary (DB) for a pair of ions (atoms) separated by a distance $d < a_i + 2r_w + a_j$. Ions and water probe van der Waals surfaces, related to the corresponding van der Waals radii a_i, a_j, and r_w, are depicted by red and blue dashed lines, respectively. The DB of the ion pair is defined by the effective ionic (atomic) dielectric radii ρ_i and ρ_j, Eq. (12), and by the effective water probe dielectric radius ρ_w which is specified by the distance from the SAS to the DB. For the solvent separated configurations ($d > a_i + 2r_w + a_j$), the ion pair DB consists of two spheres of radii ρ_i and ρ_j.

In a water molecule, the partially positively charged hydrogen atoms are located at the distance of about 1 Å from the oxygen atom center.[97] Most of explicit water models reflect this fact by having two positive partial charges placed at about the same distance from the oxygen center. A major part of the dielectric response of liquid water to a solute electric field is due to the orientational polarization of water, which changes the thermally averaged microscopic charge distribution in water layers adjacent to the solute. The most significant changes in the first hydration shell around the solute begin at about 1 Å below the SAS, in the layer of the closest approach of water hydrogen nuclei to the SES. Therefore, we suggest to place the solute/solvent DB at $\rho_w = 1.0$ Å below the SAS, at the layer where the most significant changes of the averaged water charge distribution due to orientational polarization of water are expected.

Since the SAS near the solute atom is determined by the sum of atomic and water probe van der Waals radii, $(a_i + r_w)$, and is an experimentally observable quantity,[94,95,98] the geometry of the proposed DB can be specified by the effective dielectric radii ρ_i of atoms (ions),

$$\rho_i = (a_i + r_w) - \rho_w, \tag{12}$$

and the radius of the sphere ρ_w, corresponding to the proposed distance from the SAS to DB. The quantity ρ_w can be considered as an effective dielectric radius of the water probe, $\rho_w < r_w$, which together with the atomic (ionic) effective dielectric radii ρ_i, Eq. (12), fully determine the solute/solvent DB in the form of the Richards-Connolly surface; see Fig. 3.

In many GB models, the effective dielectric probe radius ρ_w and the atomic dielectric radii ρ_i are somewhat arbitrary parameters. Contrary to this, some physical restrains can be applied to the atomic (a_i) and water probe (r_w) van der Waals radii. For example, the sums $a_i + r_w$ determine the position of the first hydration shell around the solute, which can be accessed experimentally.[94,95,98] Defining atomic (ionic) dielectric radii ρ_i through Eq. (12) retains these restrains. The proposed approach still requires the knowledge of a_i and r_w or their sum but needs no optimization of the effective dielectric radii ρ_i defined via the position of SAS and the only additional parameter ρ_w, which we set to 1.0 Å.

In what follows, we use the proposed DB to calculate the effective Born radii, considering three distinct cases: (1) for the isolated DNA, which is kept fixed during our simulations, we use the DB around the DNA to numerically calculate the "ion-free" effective Born radii of all atoms in the DNA; (2) for the ion-ion interactions, we approximate the corrections to each ion's effective Born radius (starting from ρ_j for an isolated ion) by considering the DB for each possible ion-ion pair (as depicted in Fig. 3) and then analytically calculating the corresponding "Neck" and "Sphere" integral corrections, Eq. (5); (3) for the DNA-ion interactions, all the corrections to the "ion-free" effective Born radii of the DNA atoms and the ions are approximated by considering the DB for each DNA atom-ion pair separately and analytically calculating the corresponding "Neck" and "Sphere" integrals.

B. Interaction in cation-anion pair: Water probe size and Na^+–Cl^- potential of mean force (PMF)

In this section, we explore and refine the proposed above definition of the solute/solvent dielectric boundary. We apply the new methodology of defining the DB to calculate the potential of mean force $W(d)$ for an isolated Na^+–Cl^- ion pair in water and find the adjusted value of the water probe radius r_w. To evaluate PMF $W(d)$ [Eq. (10)], we use the unmodified GB model [Eqs. (2) and (5)–(7)] for clarity and simplicity (see a note at the end of Sec. III C). The critical modifications needed to account for the presence of the DNA molecule and much larger trivalent CoHex ions will be discussed in Sec. III C.

The proposed DB around the ion pair is specified by the effective water probe dielectric radius ρ_w = 1.0 Å and the dielectric radii of ions ρ_i. The latter can be determined from Eq. (12) using the intrinsic ionic radii a_i and the realistic water probe radius r_w.

The standard water probe size r_w = 1.4 Å is typically considered complementary to the ionic radii from Refs. 94–96 used here. However, at small ion-ion separations, the electric field between the cation and anion is greater than the field around the single ion, resulting in a greater water polarization and stronger ion-water interaction. These factors lead to reduced ion-water ($a_i + r_w$) distances for water molecules between the two ions.

The ion-water distances are related to the position of the second minimum in the cation-anion PMF determined from the MD simulations[99,100] (see Fig. 4). This minimum reflects the beginning of a partial dehydration of the cation-anion pair due to steric displacement of water molecules from the region between the two ions and roughly corresponds to the $a_i + 2r_w + a_j$ distance. The sum of the ionic radii, 1.02 Å for Na^+ and 1.81 Å for Cl^-,[96] and the two standard water probe size values r_w = 1.4 Å is clearly larger than the positions of the second Na^+–Cl^- PMF minimum (5.0–5.2 Å) shown in Fig. 4.

Keeping the ionic radii a_i unchanged, the reduction of the ion-water distances can be taken into account by decreasing the r_w compared to its standard 1.4 Å value. This reduction is important for the proper description of the shape and position of the dehydration energy barrier. With our choice of ρ_w = 1.0 Å (distance from the SAS to the DB, an effective dielectric water probe radius), which determines the effective dielectric ionic radii ρ_i [see Eq. (12)], we found that the value r_w = 1.22 Å satisfies the above requirements for the position of the second PMF minimum and reasonably reproduces the explicit solvent MD PMFs using the proposed model for the DB within the GB approach. This new value of r_w results in the gap between the SES determined by a_i and r_w and the DB specified by ρ_i and ρ_w which is estimated as $(r_w - \rho_w) = 0.22$ Å (see Fig. 3). We will use this r_w = 1.22 Å value (along with ρ_w = 1.0 Å) to determine all other effective dielectric atomic (ionic) radii for ions and solutes (DNA and RNA duplexes) described in this study.

The GB derived Na^+–Cl^- PMF in Fig. 4 is calculated using the same ε_{ij} and $\tilde{R}_{ij}$ CHARMM27 LJ parameters[101] [Eq. (11)] which were used for the MD derived PMF[99] and the parameter γ = 10 cal/(mol Å^2) for the nonpolar part $\Delta G_{\text{nonpolar}}$ [Eq. (8)] of the ion hydration free energy.

As an additional confirmation for the proposed implicit water approach that uses unmodified, experimental ion radii (e.g., a_i = 1.02 Å for Na^+), we have calculated distributions of Na^+ ions around the DNA duplex within the standard non-linear PB approach, Fig. 5. As expected, the non-linear PB theory is capable of describing monovalent ion distributions around the DNA (Fig. 5), but only when the cation (Na^+) radius is 1.02 Å, chosen based on the physical considerations presented above. The accuracy of non-linear PB theory for predicting the number of bound ions (specific or non-specific binding) to a biomolecule depends on the Stern layer thickness because its value is added to the radii of the solute atoms to determine the ion accessible regions. Typically, ionic radii values of 2 Å have been used to define the Stern layer thickness for hydrated Na^+.[102,103] Our PB calculations of Na^+ distribution around the DNA with this "standard" 2.0 Å Na^+ radius

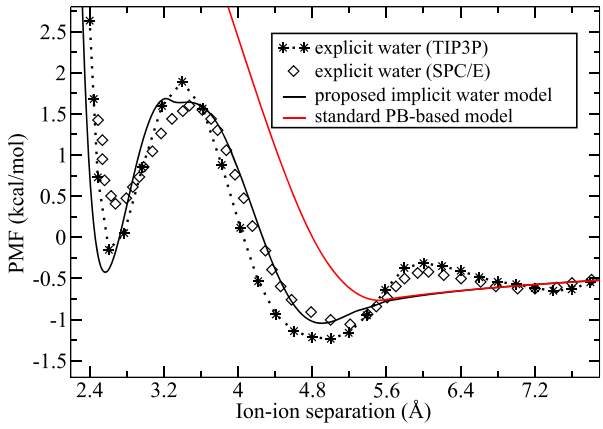

FIG. 4. Na^+–Cl^- ion pair PMF $W(d)$ [Eq. (10)] calculated using the GB approach with the analytical estimation of the "Sphere" [Eq. (6)] and "Neck" [Eq. (7)] integrals and the new dielectric boundary (DB) definition (see Fig. 3) (black solid line), numerical solution of PB equation with the "standard" solvent excluded surface (SES) based definition of the DB and water probe radius r_w = 1.4 Å (solid red line), and the results from the MD simulations with explicit TIP3P (stars)[99] and SPC/E (diamonds)[100] water models.

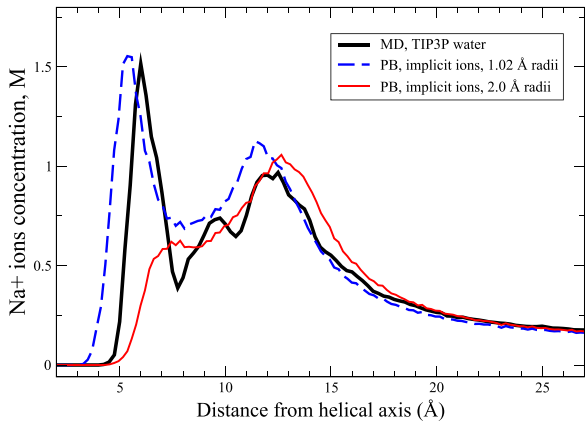

FIG. 5. Na^+ ion cylindrical distributions around the 25 bp poly(dA·dT) DNA duplex derived from the explicit water MD simulations with the TIP3P water model (thick solid black line) and from the non-linear PB calculation using experimental Na^+ value 1.02 Å for monovalent ion radii (dashed blue line) and "standard" 2 Å Na^+ radius (thin solid red line).

did not result in a good agreement with the MD-derived distributions, Fig. 5. Similar findings were reported for predicting the number of bound ions to a lipid bilayer, where the radius of 1.4 Å gave a better agreement with MD results[104] than the radius of 2.0 Å. Our calculations suggest that a non-hydrated sodium radius of 1.02 Å is the appropriate value to use for calculating Na^+ interactions with strongly charged molecules such as DNA.

C. Modification to the GB model to account for non-connected geometries

It is known that the "canonical" GB model[63] becomes essentially exact (compared to PB) in the limiting case of a spherical geometry of the solute in water. For this case, the GB Green function can be derived (in the limit $\epsilon_{out} \to \infty$) from the exact PE solution on a sphere.[28] If the solute shape is substantially non-spherical, the GB estimations of pairwise interactions between individual atomic charges of the solute may deviate substantially from the PB estimates,[105] although they are still correct qualitatively. However, the very physics of the canonical GB approximation can break down for solutes composed of multiple disconnected parts in close proximity to each other.

Considering explicit ions in an implicit solvent, we often encounter this non-connected geometry when the charges are completely separated by the solvent. For example, for the two ions in the bulk or for the ion near the solvent exposed solute atom, when the separation distance d_{ij} is greater than the sum $\rho_i + 2\rho_w + \rho_j$, the "Neck" region shown in Fig. 2 (dashed area) completely disappears. In this case, all the electrostatic interactions between the corresponding charges occur through the solvent, the case where the GB model may drastically deviate from the PE; see Table I. In the limiting case $\epsilon_{out} \to \infty$, both the PE and the Coulomb law predict vanishing of the ion-ion interactions, whereas the GB model does not. This deficiency of the canonical GB model in the case of disconnected solute geometries leads to the overestimation of the repulsive contributions between counterions around the charged solute resulting in wrong counterion distribution. The example of the consequences of these incorrect ion-ion interactions in the canonical GB approach has been presented in Fig. 1 for the case of CoHex counterions around the DNA molecule.

To approximate the Coulomb like behavior for the disconnected (solvent separated) configurations of two atomic charges (see Fig. 6), predicted by the PE, we modify Eqs. (5) and (2). Namely, when the neck integral $I_{ij}^{(N)}(d)$ in Eq. (5) vanishes at the solvent separated distances, the sphere integral corrections $I_{ij}^{(S)}$ is set to zero, reducing the effective Born radius R_i in Eq. (5) to its value calculated as if there were no ion j. In addition, to cancel the electrostatic "vacuum" parts of Eqs. (2) and (11), the exponent in Eq. (2) for the cross terms ($i \neq j$) is set to zero too.

TABLE I. The breakdown of the GB model for disconnected solute geometries. Comparison of the electrostatic interaction energy (kcal/mol) of two CoHex ions at the separation distance $d = 9.8$ Å ($d > a_i + 2r_w + a_i$, $a_i = 3.63$ Å, $r_w = 1.22$ Å), using the GB model, numerical solution of the PE, and the Coulomb law. Solvent dielectric constant $\epsilon_{out} = 80$.

Evaluation model	Generalized Born	Poisson equation	Coulomb law
Interaction energy	9.9	4.3	3.8

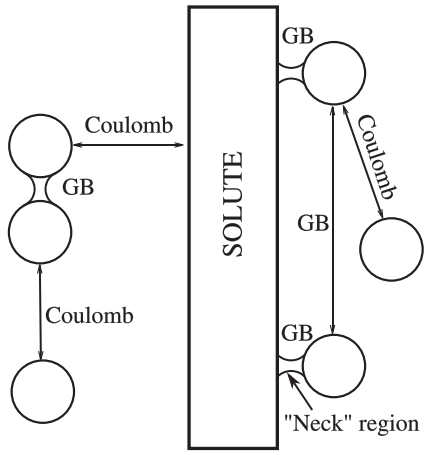

FIG. 6. Solute-ion and ion-ion interaction modes in the proposed explicit ions/implicit water model. Interactions between ions separated by a layer of the solvent (no ion-ion "Neck" formed) are estimated via the modified GB equations resulting in the Coulomb law behavior. For smaller separations ("Necks" formed), ion-ion interactions are estimated via the canonical GB equation. Interactions between two ions, both forming "Necks" with the solute, are estimated via the canonical GB equation at any ion-ion distance. Interactions between the solute atoms and ions being within ion-solute "Neck" distance are estimated via the canonical GB equation. At larger ion-solute distances, the interactions are estimated via the modified GB equations approximating the Coulomb law behavior. All the interactions within the (singly connected) solute are estimated via the canonical GB.

When the two ions are solvent separated ($I_{ij}^{(N)} = 0$), but both of them are bound to the solute (create "Necks" with the nearest solute atoms), part of the charge-charge interactions between these ions can occur through the low dielectric solute region. In these cases, both ions are considered as parts of the solute, and unmodified Eqs. (2) and (5) are applied to describe the interactions between these ions (see Fig. 6).

We have found that the deviation of the new formalism from the canonical GB for a single Na^+–Cl^- pair (relatively small monovalent ions) is quite small and has negligible effect on the computed ion-ion PMF; however, the deviation becomes large and very important once the DNA is added or large trivalent CoHex ions are involved. The use of the new formalism is therefore recommended in general for disconnected solute geometries.

D. Implementation details. Effective potentials for ion-DNA(RNA) and ion-ion interactions in implicit water

In Secs. III A–III C, we have formulated a new approach to define the dielectric boundary around charged solutes (see Fig. 3) and proposed modifications to the canonical GB model to account for the disconnected geometries of ion-ion and solute-ion configurations in the explicit ions/implicit water GB model. We showed that the proposed DB and the GB based analytical interaction potentials between charged particles can reasonably reproduce the cation-anion PMF calculated using explicit water models.

In this section, we extend the proposed explicit ions/implicit water GB approach to many-atomic solutes and multiple explicit ions and formulate analytical equations for corrections to the effective Born radii of solute atoms and ions.

In the Sec. IV, we will apply the derived equations to simulate ion distributions around constrained (all atom positions are fixed) short DNA and RNA duplexes.

1. Effective Born radii of solute atoms interacting with explicit ions

The effective Born radii R_k of constrained solute (DNA) atoms for a given fixed solute configuration are pre-calculated without ions in the implicit solvent using the GBNRS6 GB flavor[77] and the proposed new DB definition. In the presence of explicit ions near the solute, the magnitudes of R_k can deviate from their initial "ion-free" values R_k^0. Similar to Eq. (5), the corrections to R_k due to surrounding ions are calculated as sums of "Sphere" ($I_{kj}^{(S)}$) and "Neck" ($I_{kj}^{(N)}$) integrals over the corresponding dielectric volumes of neighboring ions (index j). These integrals are modified as described in Refs. 74, 77, and 106.

(1) We follow Mongan *et al.*[74] and scale the "Neck" contributions $I_{kj}^{(N)}$ to R_k by factors n_j,

$$I_k^{(N)} = \sum_j n_j I_{kj}^{(N)}. \tag{13}$$

Factors n_j depend on the ion size and charge and, therefore, are ion type specific. They are subject to optimization; see below.

(2) To implement the connectivity rules discussed in Sec. III C, we scale the "Sphere" contributions $I_{kj}^{(S)}$ to R_k by factors s_j,

$$I_k^{(S)} = \sum_j s_j I_{kj}^{(S)}. \tag{14}$$

$s_j = 1$ ("Sphere" integral $I_{kj}^{(S)}$ contributes to $I_k^{(S)}$) if ion j is within the "Neck" distance with at least one solute atom k' ($d_{k'j} < a_{k'} + 2r_w + a_j$, $I_{k'j}^{(N)} \neq 0$). Otherwise, the contribution of $I_{kj}^{(S)}$ should be neglected by setting $s_j = 0$. That is, an ion separated from the solute by a continuous solvent layer does not affect Born radii of the solute atoms.

The cumulative quantities $I_k^{(N)}$ and $I_k^{(S)}$ contribute to R_k in a manner similar to Eq. (5) as follows. For reasons described in Ref. 77, the maximum value of the corrected effective Born radius R_k is capped by the electrostatic size A of the solute molecule[77,107] (in the case of the DNA duplex $A = 18$ Å),[77,107]

$$(R_k)^{-3} = (R_k^0)^{-3}(1 - c_k \tanh(F_k)), \tag{15}$$

where

$$F_k = \frac{3}{4\pi}(I_k^{(N)} + I_k^{(S)})/\left(c_k (R_k^0)^{-3}\right) \tag{16}$$

and

$$c_k = 1 - \frac{(R_k^0)^3}{A^3}. \tag{17}$$

2. Effective Born radii of explicit ions interacting with solute atoms and other explicit ions

In the presence of low dielectric volumes of the solute (DNA) and other ions, the effective Born radius R_j of an ion j may change. Similar to the case of solute atoms, the correction to R_j^0 of an isolated ion is due to the contributions of the "Sphere" ($I_j^{(S)}$) and "Neck" ($I_j^{(N)}$) integral sums.

The "Sphere" integral correction to R_j^0 can be split into the ion-solute ($I_j^{(S1)}$) and ion-ion ($I_j^{(S2)}$) parts,

$$I_j^{(S)} = I_j^{(S1)} + I_j^{(S2)}. \tag{18}$$

The ion-solute part can be evaluated as a sum of pairwise ion-atom contributions $I_{jk}^{(S)}$ [Eq. (6)],

$$I_j^{(S1)} = s_j \sum_k I_{jk}^{(S)}, \tag{19}$$

where factor s_j accounts for the connectivity rules in the same way as in Eq. (14). That is, $s_j = 1$ when ion j is within the "Neck" distance to at least one of the solute atoms k, and $s_j = 0$ otherwise.

Equation (19) replaces the $1/r^6$ integration over the solute dielectric volume [Eq. (4)] by a sum of the integrals over the individual atom dielectric volumes. This procedure over-counts the solute volume due to overlapping of the spherical dielectric volumes of neighboring solute atoms.[108] To reduce this over-counting, we follow the practice of AMBER GB models[109] and scale the solute dielectric atomic radii ρ_k entering $I_{jk}^{(S)}$ [Eq. (6)] by correction coefficients fs_k (see Table II). Factors fs_k depend on the solute atom type: they are optimized to reproduce $I_j^{(S1)}$ calculated by numerical integration over the dielectric volume of DNA. Thus, they are specific for nucleic acids and the proposed strategy to build the DB around charged solutes. The optimization is done by minimizing RMSD between the analytical [Eqs. (19) and (6)] and numerical (GBNSR6 code[77]) evaluations of $I_j^{(S1)}$ over a large set (75) of random bound ion positions ($s_j = 1$) around the DNA molecule. Initially, the original GBNSR6 code was used to build the DB surface around the solute without the explicit ions. Then, the code was modified to use the pre-build DB and to do the $1/r^6$ integration over the volume defined by the DB, external to the ion positions. The optimization of the radii scaling factors fs_k in the analytical evaluation of $I_j^{(S1)}$ was done using the MATLAB function "fminsearch." The optimized values of fs_k specific for DNA (RNA) molecules are presented in Table II.

The ion-ion ($I_j^{(S2)}$) part of the "Sphere" integral correction to R_j^0 can be written as

TABLE II. Volume over-counting radii correction coefficients fs_k and the unscaled effective dielectric atomic radii ρ_k (in Å) [Eq. (12)] for different nucleic acid atom types.

Solute atom type	H	C	N	O	P
Dielectric radii ρ_k	1.44	1.92	1.77	1.72	2.07
Coefficients fs_k	0.992	0.000	1.002	0.994	0.937

$$I_j^{(S2)} = \sum_{j'} s_{jj'} I_{jj'}^{(S)}. \quad (20)$$

Here, $I_{jj'}^{(S)}$ is the $1/r^6$ "Sphere" integral centered on ion j over the dielectric volume of ion j' [Eq. (6)], $s_{jj'}$ the connectivity switching factors. In agreement with the connectivity rules introduced in Sec. III C, factor $s_{jj'} = 1$ if ions j and j' are within the ion-ion "Neck" distance ($d_{jj'} < a_j + 2r_w + a_{j'}$). Otherwise, $s_{jj'} = s_j s_{j'}$ where factors s_j were introduced in Eq. (14) and correspond to the ion-solute connectivity rules. Thus, the integral $I_{jj'}^{(S)}$ contributes to the $I_j^{(S2)}$ if both ions j and j' are within the "Neck" distance to the solute (DNA) atoms or the distance between these ions is within the ion-ion "Neck" distance.

Similar to the "Sphere" integral correction [Eq. (18)], the "Neck" integral correction to the effective Born radius of ion j can be slit into two parts, ion-solute ($I_j^{(N1)}$) and ion-ion ($I_j^{(N2)}$) "Neck" integral contributions,

$$I_j^{(N)} = I_j^{(N1)} + I_j^{(N2)}. \quad (21)$$

As in Eq. (13), we evaluate the $I_j^{(N1)}$ as a scaled sum of the ion-atom "Neck" integrals $I_{jk}^{(N)}$,

$$I_j^{(N1)} = c_j^V n_j \sum_k I_{jk}^{(N)}. \quad (22)$$

Neck scaling factors n_j are the same as in Eq. (13) for the solute-ion "Neck" integrals. They are ion specific and are optimized to reproduce ion distributions around DNA molecules, see Secs. IV A and IV B. Additional variable scaling factor c_j^V in Eq. (22) is applied to counteract the overestimation of the sum of "Neck" integrals in Eq. (22) due to possible overlaps of multiple ion-atom "Neck" volumes when an ion is close to multiple groups of solute atoms; for example, Na^+ in the minor groove of DNA. Factor $c_j^V = c_j^V(\hat{V}_j)$ depends on the relative ion j position via the weighted relative volume $\hat{V}_j$ of the solute atoms within a predefined spherical volume (radius R_s) around the ion.[77] Similar to the "measure of the volume" introduced by the FACTS model of solvation,[110] the $\hat{V}_j$ is defined as

$$\hat{V}_j = \frac{\sum_k a_k^3 \Theta_{kj}}{R_s^3}, \quad (23)$$

where a_k are solute atomic radii and

$$\Theta_{kj} = \begin{cases} \left(1 - \left(d_{kj}/R_s\right)^2\right)^2 & d_{kj} \le R_s, \\ 0 & d_{kj} > R_s. \end{cases} \quad (24)$$

The parameter R_s is set to 10 Å, which is the same value used in the FACTS method[110] and in Ref. 77. The increase of the weighted relative volume $\hat{V}_j$ decreases the scaling factor $c_j^V(\hat{V}_j)$ as

$$c_j^V(\hat{V}_j) = 1 - f_c \tanh(f_v \hat{V}_j). \quad (25)$$

Parameters $f_c = 0.68$ and $f_v = 8.0$ have been optimized to reproduce the dependence of the solvation free energy of the 12 base pair DNA duplex (PDB ID: 2BNA) and one Na^+ ion bound to the DNA minor groove on the ion position in the groove. We used 90 DNA-ion configurations with the equidistant ion positions 0.1 Å apart, along the midsection of the minor groove. The reference energies of the DNA-ion complex were calculated using the APBS PB solver.[32] Depending on the volume $\hat{V}_j$, the scaling factor c_j^V varies from 1 to ~0.3.

The ion-ion "Neck" integral term ($I_j^{(N2)}$) is represented as an unscaled sum of all ion-ion "Neck" integrals $I_{jj'}^{(N)}$,

$$I_j^{(N2)} = \sum_{j'} I_{jj'}^{(N)}. \quad (26)$$

With the "Neck" ($I_j^{(N)}$) and "Sphere" ($I_j^{S)}$) integral corrections, the effective Born radius R_j of ion j can be evaluated in the same manner as the solute atom radius R_k, using Eqs. (15)–(17), where atom index k is replaced by ion index j.

All the remaining parameters of the model not shown in Table II are given in the supplementary material. A collection of in-house Fortran programs to reproduce the GB-MC results of this work is available at http://people.cs.vt.edu/~onufriev/software.php.

IV. MONTE CARLO SIMULATION RESULTS USING PROPOSED EXPLICIT IONS/IMPLICIT WATER GB MODEL

We apply the proposed explicit ions/implicit water model to evaluate the interactions of ions with highly charged DNA and RNA molecules in water. To validate the model, we calculate ion distributions around the nucleic acid duplexes using ensembles of equilibrium ion configurations. To generate the configurations, we use an in-house implementation of the canonical Metropolis Monte Carlo (MC) method[111] with the proposed interaction potentials described in Secs. II and III. All the simulation details are described in Sec. V.

A. Distributions of monovalent Na^+ ions around DNA duplex

We use the proposed GB based interaction potentials between a highly charged solute and explicit ions in implicit water to calculate monovalent Na^+ ion distributions around the homopolymeric 25 base pair (bp) poly(dA·dT) DNA duplex.

The "Neck" integrals scaling parameter n_j [Eq. (13)] for the proposed GB approach is optimized by comparing the evaluated ion distributions with the results from explicit water MD simulations using TIP3P[64] and TIP4P-Ew[112] water models. All other parameters are described in Sec. V or presented in the Appendix.

The Na^+ ion distributions derived from the explicit ions/implicit water GB-MC simulations with the optimized "Neck" scaling parameters ($n_j = 0.4$) and from the MD simulations using two different explicit water models (TIP3P[64] and TIP4P-Ew[112]) are presented in Fig. 7. One can see that the GB derived ion distribution reproduces the MD results with a reasonable accuracy, comparable with the accuracy of explicit water simulations. The latter one can be estimated as the ratio of the Na^+ ion concentrations at the distribution peak at 12-13 Å (the ions residing in the vicinity of the DNA phosphate groups) derived from the TIP4P-Ew vs.

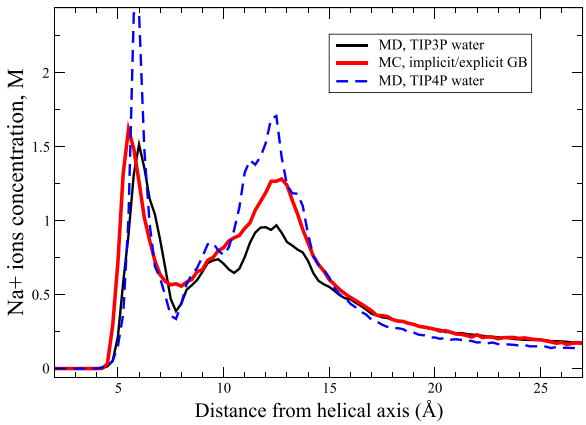

FIG. 7. Na^+ ion cylindrical distributions around the 25 bp poly(dA·dT) DNA duplex derived from the MC simulations using the proposed explicit ions/implicit water GB model (thick solid red line) and from the explicit water MD simulations with the TIP3P water model (solid black line) and with the TIP4-Ew water model (dashed blue line). The GB based result is shown for the optimized "Neck" integrals scaling parameter $n_j = 0.4$.

TIP3P water based simulations. Almost double the concentration of Na^+ ions at this peak in TIP4P-Ew water compared to TIP3P suggests the difference in the effective Na^+-DNA interaction potentials for these two explicit water models of the order of only k_BT. Since the GB derived distribution is roughly within the accuracy gap between the two explicit water distributions, the accuracy (k_BT) of the effective Na^+-DNA interactions of these two explicit water models can serve as a measure of the accuracy of the proposed GB model interaction potentials in the case of monovalent ions. We note that our model reproduces the double-peak structure of Na^+ distributions in Fig. 7 better than the previously published explicit ions/implicit water model[41] that used the near "standard" 1.8 Å Na^+ radius. This observation further emphasizes the importance of the dielectric boundary definition proposed in this work.

B. Distributions of trivalent CoHex ions around DNA and RNA duplexes

To test the proposed explicit ions/implicit water GB approach in the case of multivalent ions where the ion-ion correlations play a significant role in the effective ion-solute interactions, we apply the modified GB model to evaluate interactions of trivalent $CoHex^{3+}$ counterions with DNA and RNA duplexes. A proper reproduction of the multivalent ion distributions around nucleic acid duplexes is especially interesting since some subtle features of these distributions are responsible for the aggregation propensity of the nucleic acids.[113]

We use the same Monte Carlo code with the proposed GB interaction potentials as in the case of monovalent ions and calculate the distributions of explicit trivalent CoHex ions around the 25 bp homopolymeric poly(dA·dT) DNA duplex, mixed sequence DNA duplex, and homopolymeric poly(rA·rU) RNA duplex, assuming fixed configurations of nucleic acid molecules. Simulation details are described in Sec. V. The only parameter which needed to be re-optimized compared to the small monovalent ion case is the "Neck" scaling factor n_j which is ion type sensitive. For this optimization, we compare the results of the GB-MC simulations of the poly(dA·dT) DNA-CoHex system with the results of all-atom MD simulation of the same DNA duplex and CoHex counterions using the explicit TIP3P water model.[113] We find that the best agreement of the implicit water results to the explicit water reference corresponds to $n_j = 0.9$. Only the homopolymeric DNA duplex results were used for the re-optimization of this parameter. The other two test systems, the mixed sequence DNA and homopolymeric RNA, are simulated without any additional re-optimization.

CoHex distributions around the 25 bp poly(dA·dT) DNA duplex derived from the explicit ions/implicit water GB-MC simulation with the optimized parameter n_j and from the all-atom MD simulation using the explicit TIP3P water model are presented in Fig. 8. The distribution derived from the non-linear PB calculation with implicit CoHex ions is presented in Fig. 8 as well. As expected, the mean-field PB approach does not reproduce the key qualitative feature of CoHex ion distribution around the DNA—a sharp peak at 13 Å. Instead, the PB-derived CoHex distribution is broad and diffuse. This is because ion-ion correlations important for proper condensation of trivalent ions at the DNA surface are missing in the implicit ion mean-field description. As a result, the ion-ion repulsion is overestimated, leading to a more diffuse ion distribution compared to the explicit ions/explicit solvent. When applied to explicit ions, the canonical GB also can not reproduce the correct CoHex distribution, Fig. 1. As we have discovered, charge-charge repulsion between the CoHex ions trying to condense onto the DNA surface is also overestimated in this case, resulting in a diffuse counterion cloud. However, the physical origin of the overestimation is different from that in the PB model. Specifically, the canonical GB model overestimates the repulsive desolvation contributions to the interaction between charges when applied to the discontinuous dielectric boundary[114] around the solute-solute, solute-ion, or ion-ion pairs. Some of these overestimations cancel each other to an

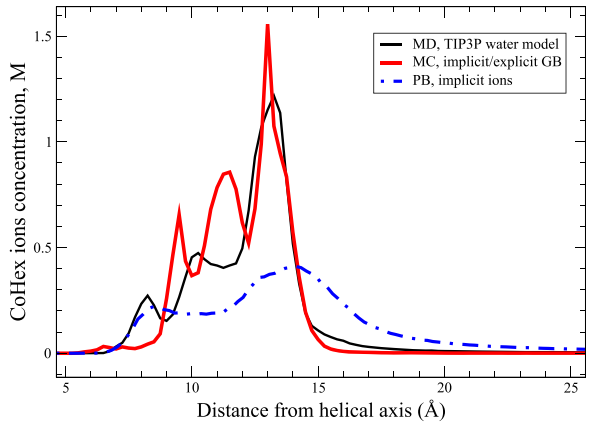

FIG. 8. CoHex ion distributions around the 25 bp homopolymeric poly(dA·dT) DNA duplex derived from the MC simulations using the proposed explicit ions/implicit water GB model (thick solid red line) from the explicit water MD simulations using the TIP3P water model (thin solid black line) and from the non-linear PB calculation with explicit CoHex ions (dot-dashed blue line). The GB derived results are shown for the optimized "Neck" integrals scaling parameter $n_j = 0.9$.

extent between GB self- and cross terms in the case of opposite charge interactions but not for interactions between like charges.

By contrast, the CoHex distribution estimated based on the proposed implicit/explicit model, Fig. 8, reasonably reproduces the position, the height, and the width of the major peak of the CoHex density at 13 Å from the helical axis, which represents most (~2/3) of the DNA bound CoHex ions. The ions under this peak are bound at the "external" surface of the DNA sugar-phosphate backbone and strongly interact with the DNA phosphate groups. Two smaller peaks in the ion distribution at 9.5 and 11.5 Å may correspond to the lesser peaks seen in the explicit water at 8 and 10 Å, respectively. These represent CoHex ions in the major groove but still bound to the phosphates. Even with the presence of the two "spurious" peaks, the corresponding difference in the ion concentrations seen in the implicit vs. the explicit solvent distributions (at the peak positions) is still small—expressed as the effective CoHex-DNA interaction energies; the interactions in the implicit water differ by less than k_BT from the explicit water results, which may be acceptable. A more detailed analysis will be needed to pinpoint the origin of the shift and enhancement of these two peaks relative to the explicit solvent results and to fine tune the interaction potentials of our model to obtain better agreement with the reference.

To further test the proposed explicit ions/implicit water GB model without any additional parameter re-optimization, we carried out the GB-MC simulations of the 25 bp mixed sequence DNA and homopolymeric poly(rA·rU) RNA duplexes. Representative configurations of CoHex ions bound to DNA and RNA duplexes from the MD and GB-MC simulations are shown in Fig. 9 for all three simulated systems. In all three cases, the GB-MC simulations (right panels in Fig. 9) reproduce the major features of CoHex binding to DNA and RNA molecules (left panels in Fig. 9). In the case of DNA duplexes, CoHex ions bind preferentially to the surface of the phosphate-sugar backbone exposed to the bulk, in the "external" ion biding shells of the duplexes (at 12–16 Å from the helical axis).[114] GB-MC simulation of the mixed sequence DNA duplex reproduces the sequence specific binding of CoHex ions to the nucleotide bases at Guanine-phosphate-Cytosine (GpC) steps in the major groove of DNA,[113] Fig. 9 (green ions). In the case of the A-form RNA duplex, CoHex ions bind mostly inside the RNA major groove, in the "internal" ion binding shell (within 12 Å from the helical axis), still strongly interacting with the phosphates group oxygens exposed into the RNA major groove.

It has been suggested[113,115] that the fraction of neutralizing CoHex counterions bound in the "external" ion binding shell of nucleic acid duplexes is responsible for the condensation propensity of these duplexes. The larger this fraction, which is duplex structure and sequence dependent, the higher the condensation propensity of the nucleic acid molecules. It has been suggested that the preferential "internal" shell binding of CoHex counterions and, therefore, small fraction of CoHex counterions bound in the "external" shell of RNA duplexes are responsible for a very weak attraction between the duplexes preventing RNA condensation in solutions.

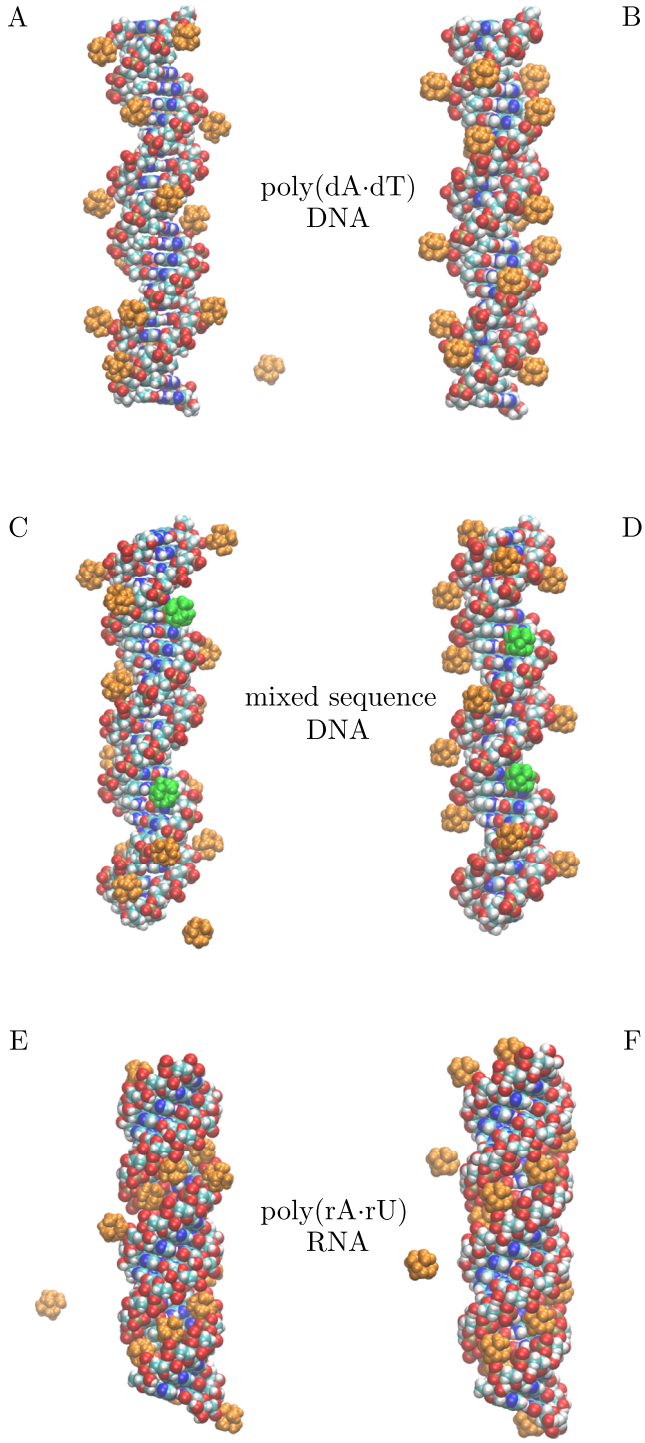

FIG. 9. Representative snapshots of CoHex ions (orange) binding to 25 bp DNA and RNA duplexes from explicit (TIP3P) water MD simulations (left) and from the proposed explicit ions/implicit water GB model used in MC simulations (right). CoHex ions bound to the GpC steps in the major grooves of the mixed sequence DNA duplexes are shown in green.

All-atom MD simulations correctly reproduce the relative order of the values of this fraction for different DNA and RNA duplexes,[113] which correspond to the relative order of their condensation propensities. Thus, a proper reproduction of the "internal" and "external" CoHex ion binding modes for different types of nucleic acid duplexes would be an important characteristic of the proposed explicit ions/implicit water GB model.

The estimates of the average quantities of bound CoHex ions in the "external" and "internal" ion binding shells for the three DNA and RNA duplexes in the GB-MC and explicit water MD simulations are summarized in Table III. The number of ions bound in the "internal" ion binding shells also includes ions referred to as deeply buried ions in Ref. 113, which are bound to the nucleotide bases deep inside the major groove of a nucleic acid duplex, Fig. 9 (green ions).

The results presented in Table III show a reasonable agreement between the proposed explicit ions/implicit water GB-MC approach and the fully explicit solvent reference, suggesting that the proposed model captures the main features of trivalent CoHex distributions around nucleic acid duplexes. The proposed GB model reproduces the major difference in ion binding to DNA and RNA duplexes, i.e., the much smaller number of bound CoHex ions in the "external" shell of the RNA duplex. This reduced quantity of the "externally" bound CoHex ions of RNA duplexes was suggested as the main reason of a low condensation propensity of RNA molecules.[113]

However, the presented results suggest that the accuracy of our GB based ion-DNA interaction potentials is not sufficient to reproduce a more subtle difference in CoHex binding in the "external" ion binding shells of the homopolymeric and mixed sequence DNA duplexes. While the MD simulations with the explicit water show about 0.8 bound ion difference, the explicit ions/implicit water GB-MC approach shows none. This can be attributed to the excessive binding of CoHex ions in the major groove of poly(dA·dT) DNA at a distance of 10-12 Å from the helical axis to the phosphates in the GB-MC simulations. On the other hand, our GB-MC approach reproduces the specific binding of two CoHex ions to the GpC steps inside the major groove of the mixed sequence DNA duplex, Fig. 9 (green ions). This specific binding was suggested[113] to be one of the reasons for a reduction of the fraction of "externally" bound CoHex ions and consequently the reduced condensation propensity of the mixed sequence DNA duplexes compared to the homopolymeric DNA.

The explicit ions/implicit solvent GB-MC approach has a great advantage in the speed of generating equilibrium configurations of ions over the all-atom explicit water MD simulations. As presented above, the GB-MC CoHex ion distribution was calculated from a set of ion configurations generated during 1.6×10^6 MC ion moves, which took 17 h using a single CPU core. A good convergence of the distribution was reached after 0.8×10^6 MC ion moves. On the other hand, in the case of explicit water MD simulations, it was necessary to generate at least 100–300 ns trajectory to calculate the equivalent converged CoHex distribution. Using the same single CPU core with the observed simulation rate 0.8 ns/day would take 125 to 375 days to generate the required trajectory. Thus, the proposed GB-MC method has an about 200 fold advantage in speed.

TABLE III. Average numbers of bound CoHex ions in the "external" (12–16 Å from the helix axis) and "internal" (0–12 Å) ion binding shells of 25 bp DNA and RNA duplexes derived from the explicit (TIP3P) water MD and explicit ions/implicit water GB-MC simulations.

	"External" shell ions		"Internal" shells ions	
		Explicit solvent		Explicit solvent
	GB-MC	MD	GB-MC	MD
Poly(dA·dT) DNA	8.7	9.4	7.0	5.2
Mixed sequence DNA	8.7	8.6	7.2	5.9
Poly(rA·rU) RNA	1.6	1.3	11.5	12.8

C. Effect of demethylation of thymine residue in poly(dA·dT) DNA duplex

Analyzing the effect of DNA sequence on the tetravalent spermine induced interaction between DNA duplexes, Yoo *et al.*[12] have shown that the methylation of Cytosines in GC-rich DNA molecules results in increased counterion-mediated attraction between DNA molecules. Comparing the interactions between AT-rich and GC-rich DNA duplexes, the authors also suggested that methyl groups of thymine act as steric blocks forcing out spermine ions from the major grooves of AT-rich duplexes to the "external" shell regions, which increases the effective DNA-DNA attraction.

To test the effect of the thymine methyl groups on CoHex binding to the poly(dA·dT) DNA within our explicit ions/implicit water GB-MC approach, we "de-methylated" all of the thymine groups. Namely, in each thymine base of the homopolymeric DNA duplex, we replaced the methyl group by a hydrogen atom with a partial charge equal to the sum of all partial charges of the replaced methyl group. The resulting CoHex distribution derived from extended (1.6×10^6 steps) MC simulation of the modified DNA has shown a slight decrease (0.4 ion) of the number of CoHex ions in the "external" ion binding shell and a corresponding increase (0.6 ion) of the number of ions in the "internal" shells. The changes are consistent with the idea[12] that the presence of methyl groups can shift the ion distribution toward the external shell, conducive of DNA-DNA attraction. However, in the case of CoHex ions and poly(dA·dT) DNA, the effect is rather small.

To check to what extent this small effect may be explained by the steric (van der Waals) repulsion or by a partial CoHex desolvation due to the proximity of the methyl group representing a low dielectric volume, we "methylated" the "de-methylated" DNA in a way that mimics only the desolvation, but not the steric, effect of the methyl group. Namely, we increased the size of the hydrogen atoms that replaced the methyl groups in the modified DNA duplex from 1.2 to 2.0 Å, mimicking the dielectric volume of the methyl group but kept the original van der Waals interaction parameters of the hydrogen atom. The resulting numbers of bound CoHex ions observed in a GB-MC simulation in both shells have returned approximately to their values in the original "control" duplex, constituting 8.9 and 6.8 ions, respectively.

The above results suggest that it is ion desolvation penalty due to the low dielectric volume of the thymine methyl group that is mostly responsible for the ion binding difference in the case of highly charged CoHex counterions. A simple analytical estimation of the effect of the low dielectric sphere of the methyl group size on the approaching trivalent CoHex

shows that the effective range of the CoHex ion desolvation extends beyond the methyl group steric repulsion distance by roughly the size of the water molecule, which is ~3 Å. As a result, the effective volume of the methyl group unavailable for CoHex due to the desolvation induced repulsion becomes about 10 times larger than excluded volume due to the steric repulsion alone. We conclude that desolvation effects should be taken into account in the analysis/explanation of the effects of methylation on binding of multi-valent counterions.

V. SIMULATION PROTOCOLS

A. Poisson-Boltzmann calculations

We performed nonlinear Poisson-Boltzmann calculations using APBS (v. 1.3) to compute the cation concentration profiles around the 25 bp poly(dA·dT) DNA duplex in NaCl or $CoHexCl_3$ salt solutions. Calculations were performed on a 200 Å cube box with 1.041 66 Å grid resolution (193 × 193 × 193 grid dimensions). The bulk salt concentration was set to 128 mM and 5 mM for the DNA-NaCl and DNA-$CoHexCl_3$ systems, respectively. The solvent probe radius and temperature T were 1.4 Å and 298 K, respectively. The ion accessible region was defined using the same radius for the cation and the anion: 1.02 Å or 2.0 Å for the DNA-NaCl system and 3.63 Å for the DNA-$CoHexCl_3$ system. The solvent and solute dielectric constants were 80 and 4, respectively. The molecular surface definition with 9-point harmonic averaging was used to define and smooth the dielectric and ion-accessibility coefficients (*srfm* keyword value was *smol*). Cubic B-spline discretization was applied to map the solute point charges onto the grid (*chgm* keyword value was *spl2*). The individual ion concentrations at each grid point were finally calculated using the formula

$$c_{ion} = c_{bulk} \times Acc \times \exp(-q\phi), \tag{27}$$

where, c_{bulk} is the bulk ion concentration in M units, Acc is the ion accessibility map ($0 \le Acc \le 1$), q is the ion charge in e units, and ϕ is the electrostatic potential in k_BT/e units. To compute the 1D (cylindrical) concentration profile, the grid concentration values were interpolated onto a finer grid with a resolution of 0.104 166 Å using cubic spline interpolation and then averaged within cylindrical shells of $2 \times 0.104\,166$ Å width at each radial distance from the DNA helical axis.

B. DNA–NaCl system. Explicit ions/implicit water GB-MC simulations

25 bp poly(dA·dT) DNA duplex was constructed in the canonical B-form using Nucleic Acid Builder (NAB).[116] For Monte Carlo simulations with the proposed explicit ions/implicit solvent GB approach, the DNA duplex and 72 Na^+ and 24 Cl^- ions are placed into a cylinder of 34 Å radius and 114 Å height with the harmonic reflective boundary potential [3 kcal/(mol Å^2) force constant]. The dielectric constants of the solutes and the solvent were chosen as $\epsilon_{in} = 4$ and $\epsilon_{out} = 80$, respectively. The effective surface tension parameter in the nonpolar solvation term is $\gamma = 5$ cal/(mol Å^2). The DNA duplex atomic partial charges and the parameters for van der Waals interactions between DNA, Na^+, and Cl^- ions are taken from the AMBER *ff99bsc0* force-field.[66] Ionic radii for Na^+ and Cl^- are taken from Ref. 96 and are 1.02 Å and 1.81 Å, respectively. The DNA duplex was fixed in its original canonical B-form. 960 000 MC moves of ions were generated at temperature $T = 300$ K. The optimized "Neck" scaling factor n_j was 0.4.

For the estimation of the free energies of the DNA-ion configurations in the Monte Carlo procedure, the additive constants $\Delta G^{(i)}_{\text{solv}}$ and $\Delta G^{(j)}_{\text{solv}}$ in Eq. (10)—the hydration free energies of isolated ions and solute—are not calculated.

C. DNA–NaCl system. Explicit water MD simulations

All-atom MD simulations of the monovalent ion distributions around the same 25 bp poly(dA·dT) DNA duplex in canonical B-form using an explicit TIP3P water model[64] were carried out using the AMBER12[66] package and AMBER *ff99bsc0* force field.[117,118] The simulated system contained 72 Na^+ and 24 Cl^- ions and 16 885 TIP3P water molecules. The DNA atoms were harmonically restrained with 50 kcal/(mol Å^2) force constant. Canonical NVT ensemble, periodic boundary conditions, and the particle mesh Ewald (PME) method were used to generate 320 ns trajectory utilizing 2 fs time step at T = 300 K temperature. The latter was maintained using the Langevin thermostat with the collision frequency of 1 ps^{-1}. The force filed parameter for Na^+ and Cl^- ions are default AMBER *parm99* force field ion parameters, adapted from Refs. 119 and 120, respectively. First 20 ns of the trajectory were excluded from calculating Na^+ distributions. The distributions calculated using next 100 ns and larger 300 ns portions of the trajectory are practically indistinguishable, suggesting good convergence of the system with monovalent ions. Additional simulation of the same system was carried out using the TIP4P-Ew[112] explicit water model and Joung and Cheatham[121] parameters for Na^+ and Cl^- ions. The rest of the simulation details are as described above.

D. DNA(RNA)–CoHex system. Explicit ions/implicit water GB-MC simulations

The construction of the 25 bp homopolymeric poly(dA·dT) DNA duplex is described above. The 25 bp mixed sequence[113] DNA duplex and the 25 bp homopolymeric poly(rA·rU) RNA duplex were constructed in canonical B- and A-form, respectively, using NAB.[116] Each duplex (charge −48 e) was neutralized by 16 CoHex ions. No other ions were included in the simulations. All other parameters and the force-fields were the same as in the case of the DNA–NaCl system described above. For the GB based interactions, $CoHex^{3+}$ ions are considered as spheres of 3.63 Å radius: the sum of 2.43 Å cobalt-hydrogen distance and 1.2 Å *bondi* hydrogen atom radius. For the van der Waals interactions, all 25 atoms of CoHex ion have been considered explicitly. 1 600 000 MC moves of ions were generated at temperature T = 300 K. The optimized "Neck" scaling factor n_j was 0.9.

In summary, to reproduce the results shown in Fig. 8 and Table III, one needs the set of parameters presented in the supplementary material.

E. DNA(RNA)–CoHex system. Explicit water MD simulations

The details of the all-atom MD simulations of DNA(RNA)–CoHex systems using the explicit TIP3P water model[64] are described elsewhere[113] and are similar to the all-atom MD simulation of the DNA–NaCl system described above. For the homopolymeric and mixed sequence DNA duplexes in the canonical B-form, we use the results of MD simulations described in Ref. 113. Additional MD simulation has been performed for the homopolymeric RNA duplex in the canonical A-form using the TIP3P water model. All MD simulations were carried out using the AMBER12[66] package and AMBER *ff99bsc0* force field[117,118] with CoHex parameters described in Ref. 122.

VI. CONCLUSIONS

Accurate evaluation of ionic atmosphere around hydrated nucleic acids is of great importance for understanding their structure and dynamics, but the use of traditional explicit water models for such computations may become prohibitively expensive. Implicit solvation methods eliminate solvent degrees of freedom allowing substantial reduction of computational efforts. Equally important, these methods facilitate the understanding of complex phenomena by dissecting the underlying energetics into distinct physical components. However, atomistic simulations of multi-valent ion distributions around charged polymers with an accuracy comparable to that expected from the explicit solvent MD simulations remains problematic.

Here, we have developed and tested a modification of the analytically generalized Born approach which allows one to explicitly consider ions around charged polymers and, at the same time, to treat water implicitly. The major components of the proposed model are an approach for positioning the dielectric boundary (DB) around highly charged solutes and explicit ions and a modification of the GB electrostatic interaction terms to account for the case of disconnected DB around a solute-ion or ion-ion pair. Specifically, we propose that the DB should be placed at the distance of the water O–H bond (about 1 Å) below the solvent accessible surface (SAS) around the solute. The GB interactions are modified according to solute-ion and ion-ion connectivities: low dielectric regions within a single continuous DB interact according to the canonical GB, while regions belonging to disjoint, separate pieces of the DB interact via a scaled Coulomb potential. Key parameters of the model have been optimized against the explicit solvent potential of mean force for a Na^+–Cl^- ion pair, the numerical PB-derived solvation energy of explicit Na^+ ion in the minor groove of the DNA duplex, and the ion distribution around homopolymeric poly(dA·dT) DNA obtained from the explicit water MD simulations.

We have tested the proposed explicit ions/implicit water GB model by carrying out Monte Carlo (GB-MC) simulations of monovalent (Na^+ and Cl^-) and trivalent [cobalt(III) hexammine ($CoHex^{3+}$)] ions around DNA and RNA nucleic acid duplexes of various sequences. The results of these GB-MC simulations are compared with traditional explicit water all-atom MD simulations. The counterion distributions derived from the explicit ions/implicit water GB-MC simulations are found to be in reasonable agreement with the explicit water MD results: The GB-MC derived ion distributions suggest that the effective ion interaction potentials based on the modified GB are within $\sim k_BT$ agreement with the explicit solvent interactions. The proposed approach confidently reproduces all major features of CoHex distributions around DNA and RNA duplexes, most relevant to counterion-induced condensation of nucleic acids. These include the preferential CoHex binding to the outer surface of the sugar-phosphate backbone of the B-form DNA duplex and primary CoHex binding inside the major groove of the A-form RNA duplex. Moreover, the proposed modified GB approach is able to distinguish between the specific CoHex binding inside the major groove of the mixed sequence DNA duplex and the lack of such biding in the case of homopolymeric poly(dA·dT) DNA, which is a relatively fine distinction. However, the model is not sensitive enough to resolve an even finer difference in CoHex biding to the external surface of these duplexes seen in the explicit water MD simulations. Further fine tuning of the effective interaction potentials is needed to overcome these discrepancies. When constructing the current model, we opted for the simplest functional forms consistent with the general principle; there is definite room for improvement. The use of additional characteristics of the counterion distribution as target quantities for model parameter optimization should also be considered. For example, pair correlation functions of CoHex distributions can be sensitive to the details of the interaction potential.

We have applied the new model to investigate the effect of the thymine methyl group on subtle features of CoHex distribution around DNA, following a recent suggestion that the steric effects are responsible for the difference in the ion distribution patters between methylated and unmethylated DNA, leading to differences in counterion-induced aggregation. Within our implicit water approach, we can easily separate the purely steric (methyl group–ion) repulsion from an effective repulsion induced by desolvation of the counterion coming in contact with the methyl group. We find that the desolvation induced repulsion of CoHex due to the low dielectric volume of the methyl group has a longer range and becomes appreciably sooner than the pure steric repulsion upon CoHex binding to the DNA at the thymine methyl group. A more detailed future study can easily extend the sequence space to be explored owing to the computational effectiveness of the proposed explicit/implicit model. Such a study can utilize a combination approach in which the most interesting predictions from the explicit ions/implicit water model are further refined within fully explicit, traditional simulations. DNA methylation is a class of post-translational modifications extremely important for the epigenetic control of cell function, but atomistic mechanisms behind this control are only beginning to emerge.

The proposed model is fully analytical, which should facilitate its eventual adaptation for MD simulations. An adaptation to Grand Canonical Monte Carlo simulations will be more straightforward. Incorporating more complex types of polyions, including flexible ones such as spermine, is

likely within easy reach. We also envision incorporation of the model into multi-resolution approaches to the simulation of chromatin components, where the atomistic treatment of multi-valent counterions may be critical.

SUPPLEMENTARY MATERIAL

See supplementary material for the force field and solvation parameters used in our modified generalized Born model.

ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health (NIH) No. R01 GM099450 and the National Science Foundation (NSF) No. MCB-1715207.

We thank Nathan A. Baker for helpful suggestions.

[1]C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules," Proc. Natl. Acad. Sci. U. S. A. **94**, 6185–6190 (1997).
[2]G. S. Manning, "The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides," Q. Rev. Biophys. **11**, 179–246 (1978).
[3]J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, "Salt dependence of the elasticity and overstretching transition of single DNA molecules," Biophys. J. **82**, 3160–3169 (2002).
[4]Z. J. Tan and S. J. Chen, "Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length," Biophys. J. **90**, 1175–1190 (2006).
[5]A. K. Mazur, "Local elasticity of strained DNA studied by all-atom simulations," Phys. Rev. E **84**, 021903 (2011).
[6]B. Luan and A. Aksimentiev, "Strain softening in stretched DNA," Phys. Rev. Lett. **101**, 118101 (2008).
[7]L. Bao, X. Zhang, L. Jin, and Z. J. Tan, "Flexibility of nucleic acids: From DNA to RNA," Chin. Phys. B **25**, 18703 (2016).
[8]V. A. Bloomfield, "Condensation of DNA by multivalent cations: Considerations on mechanism," Biopolymers **31**, 1471–1481 (1991).
[9]V. A. Bloomfield, "DNA condensation by multivalent cations," Biopolymers **44**, 269–282 (1997).
[10]G. C. L. Wong and L. Pollack, "Electrostatics of strongly charged biological polymers: Ion-mediated interactions and self-organization in nucleic acids and proteins," Annu. Rev. Phys. Chem. **61**, 171–189 (2010).
[11]A. V. Drozdetski, I. S. Tolokh, L. Pollack, N. Baker, and A. V. Onufriev, "Opposing effects of multivalent ions on the flexibility of DNA and RNA," Phys. Rev. Lett. **117**, 028101 (2016).
[12]J. Yoo, H. Kim, A. Aksimentiev, and T. Ha, "Direct evidence for sequence-dependent attraction between double-stranded DNA controlled by methylation," Nat. Commun. **7**, 11045 (2016).
[13]A. M. Katz *et al.*, "Spermine condenses DNA, but not RNA duplexes," Biophys. J. **112**, 22–30 (2017).
[14]A. K. Shaytan *et al.*, "Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: Atomistic insights into nucleosome functions," J. Mol. Biol. **428**, 221–237 (2016).
[15]N. V. Berezhnoy *et al.*, "The influence of ionic environment and histone tails on columnar order of nucleosome core particles," Biophys. J. **110**, 1720–1731 (2016).
[16]G. Arya and T. Schlick, "Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model," Proc. Natl. Acad. Sci. U. S. A. **103**, 16236–16241 (2006).
[17]R. Collepardo-Guevara *et al.*, "Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study," J. Am. Chem. Soc. **137**, 10205–10215 (2015).
[18]N. Korolev, L. Nordenskiöld, and A. P. Lyubartsev, "Multiscale coarse-grained modelling of chromatin components: DNA and the nucleosome," Adv. Colloid Interface Sci. **232**, 36–48 (2016).
[19]C. J. Cramer and D. G. Truhlar, in *Solvent Effects and Chemical Reactivity*, edited by O. Tapia and J. Bertran (Kluwer Academic Publishers, Netherlands, 1995), pp. 1–80.
[20]B. Roux and T. Simonson, "Implicit solvent models," Biophys. Chem. **78**, 1–20 (1999).
[21]B. Honig and A. Nicholls, "Classical electrostatics in biology and chemistry," Science **268**, 1144–1149 (1995).
[22]J. D. Madura *et al.*, "Biological applications of electrostatic calculations and Brownian dynamics simulations," Rev. Comput. Chem. **5**, 229–267 (1994).
[23]M. K. Gilson, "Theory of electrostatic interactions in macromolecules," Curr. Opin. Struct. Biol. **5**, 216–223 (1995).
[24]M. Scarsi, J. Apostolakis, and A. Caflisch, "Continuum electrostatic energies of macromolecules in aqueous solutions," J. Phys. Chem. A **101**, 8098–8106 (1997).
[25]T. Simonson, "Electrostatics and dynamics of proteins," Rep. Prog. Phys. **66**, 737–787 (2003).
[26]N. Baker, D. Bashford, and D. Case, *Implicit Solvent Electrostatics in Biomolecular Simulation*, Lecture Notes in Computational Science and Engineering (Springer, 2006), Vol. 49, pp. 263–295.
[27]J. P. Bardhan, "Biomolecular electrostatics—I want your solvation (model)," Comput. Sci. Discovery **5**, 013001 (2012).
[28]A. Onufriev, in *Modeling Solvent Environments*, edited by M. Feig (Wiley, USA, 2010), pp. 127–165.
[29]R. Anandakrishnan, A. Drozdetski, R. C. Walker, and A. V. Onufriev, "Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations," Biophys. J. **108**, 1153–1164 (2015).
[30]S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Implicit solvent simulation of million-atom structures: Insights into the organization of 30-nm chromatin fiber," J. Chem. Theory Comput. **12**, 5946–5959 (2016).
[31]Z. Xu, X. Cheng, and H. Yang, "Treecode-based generalized Born method," J. Chem. Phys. **134**, 064107 (2011).
[32]N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome," Proc. Natl. Acad. Sci. U. S. A. **98**, 10037–10041 (2001).
[33]D. Bashford, in *Scientific Computing in Object-Oriented Parallel Environments*, Lecture Notes in Computer Science (Springer Berlin, Heidelberg, 1997), Vol. 1343, pp. 233–240.
[34]L. Li, C. Li, Z. Zhang, and E. Alexov, "On the dielectric 'constant' of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi," J. Chem. Theory Comput. **9**, 2126–2136 (2013).
[35]D. Chen, Z. Chen, C. Chen, W. Geng, and G. W. Wei, "MIBPB: A software package for electrostatic analysis," J. Comput. Chem. **32**, 756–770 (2011).
[36]J. Wang, C. Tan, E. Chanco, and R. Luo, "Quantitative analysis of Poisson-Boltzmann implicit solvent in molecular dynamics," Phys. Chem. Chem. Phys. **12**, 1194–1202 (2010).
[37]M. Totrov and R. Abagyan, "Rapid boundary element solvation electrostatics calculations in folding simulations: Successful folding of a 23-residue peptide," Biopolymers **60**, 124–133 (2001).
[38]J. A. Grant, B. T. Pickup, and A. Nicholls, "A smooth permittivity function for Poisson-Boltzmann solvation methods," J. Comput. Chem. **22**, 608–640 (2001).
[39]N. V. Prabhu, P. Zhu, and K. A. Sharp, "Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method," J. Comput. Chem. **25**, 2049–2064 (2004).
[40]R. Luo, L. David, and M. K. Gilson, "Accelerated Poisson-Boltzmann calculations for static and dynamic systems," J. Comput. Chem. **23**, 1244–1253 (2002).
[41]N. V. Prabhu, M. Panda, Q. Yang, and K. A. Sharp, "Explicit ion, implicit water solvation for molecular dynamics of nucleic acids and highly charged molecules," J. Comput. Chem. **29**, 1113–1130 (2008).
[42]L. Xiao, C. Wang, and R. Luo, "Recent progress in adapting Poisson-Boltzmann methods to molecular simulations," J. Theor. Comput. Chem. **13**, 1430001 (2014).
[43]M. Feig *et al.*, "Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures," J. Comput. Chem. **25**, 265–284 (2004).
[44]N. Forouzesh, S. Izadi, and A. V. Onufriev, "Grid-based surface generalized Born model for calculation of electrostatic binding free energies," J. Chem. Inf. Model. **57**, 2505–2513 (2017).
[45]S. Izadi, R. C. Harris, M. O. Fenley, and A. V. Onufriev, "Accuracy comparison of generalized Born models in the calculation of electrostatic binding free energies," J. Chem. Theory Comput. **14**, 1656–1670 (2018).

[46]G. Sigalov, P. Scheffel, and A. Onufriev, "Incorporating variable dielectric environments into the generalized Born model," J. Chem. Phys. **122**, 094511 (2005).

[47]Z. Yu, M. P. Jacobson, J. Josovitz, C. S. Rapp, and R. A. Friesner, "First-shell solvation of ion pairs: Correction of systematic errors in implicit solvent models," J. Phys. Chem. B **108**, 6643–6654 (2004).

[48]C. Tan, L. Yang, and R. Luo, "How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis," J. Phys. Chem. B **110**, 18680–18687 (2006).

[49]G. M. Giambaşu, T. Luchko, D. Herschlag, D. M. York, and D. A. Case, "Ion counting from explicit-solvent simulations and 3D-RISM," Biophys. J. **106**, 883–894 (2014).

[50]S. Gavryushov and P. Zielenkiewicz, "Electrostatic potential of B-DNA: Effect of interionic correlations," Biophys. J. **75**, 2732–2742 (1998).

[51]Z. J. Tan and S. J. Chen, "Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte," J. Chem. Phys. **122**, 044903 (2005).

[52]R. Kjellander, "Intricate coupling between ion–ion and ion–surface correlations in double layers as illustrated by charge inversion–combined effects of strong Coulomb correlations and excluded volume," J. Phys.: Condens. Matter **21**, 424101 (2009).

[53]M. S. Loth and B. I. Shklovskii, "Non-mean-field screening by multivalent counterions," J. Phys.: Condens. Matter **21**, 424104 (2009).

[54]M. L. Sushko *et al.*, "The role of correlation and solvation in ion interactions with B-DNA," Biophys. J. **110**, 315–326 (2016).

[55]Z. Jia, L. Li, A. Chakravorty, and E. Alexov, "Treating ion distribution with Gaussian-based smooth dielectric function in DelPhi," J. Comput. Chem. **38**, 1974–1979 (2017).

[56]M. Petukh *et al.*, "Predicting nonspecific ion binding using DelPhi," Biophys. J. **102**, 2885–2893 (2012).

[57]M. Petukh, M. Zhang, and E. Alexov, "Statistical investigation of surface bound ions and further development of BION server to include pH and salt dependence," J. Comput. Chem. **36**, 2381–2393 (2015).

[58]A. Chakravorty, Z. Jia, L. Li, and E. Alexov, "A new DelPhi feature for modeling electrostatic potential around proteins: Role of bound ions and implications for zeta-potential," Langmuir **33**, 2283–2295 (2017).

[59]A. Mukhopadhyay, A. T. Fenley, I. S. Tolokh, and A. V. Onufriev, "Charge hydration asymmetry: The basic principle and how to use it to test and improve water models," J. Phys. Chem. B **116**, 9776–9783 (2012).

[60]D. L. Mobley, A. E. Ii, C. J. Fennell, and K. A. Dill, "Charge asymmetries in hydration of polar solutes," J. Phys. Chem. B **112**, 2405–2414 (2008).

[61]W. M. Latimer, K. S. Pitzer, and C. M. Slansky, "The free energy of hydration of gaseous ions, and the absolute potential of the normal calomel electrode," J. Chem. Phys. **7**, 108–111 (1939).

[62]A. A. Rashin and B. Honig, "Reevaluation of the Born model of ion hydration," J. Phys. Chem. **89**, 5588–5593 (1985).

[63]W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechnics and dynamics," J. Am. Chem. Soc. **112**, 6127–6129 (1990).

[64]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).

[65]H. Nguyen, A. Pérez, S. Bermeo, and C. Simmerling, "Refinement of generalized Born implicit solvation parameters for nucleic acids and their complexes with proteins," J. Chem. Theory Comput. **11**, 3714–3728 (2015).

[66]D. A. Case *et al.*, "The Amber biomolecular simulation programs," J. Comput. Chem. **26**, 1668–1688 (2005).

[67]B. Lee and F. Richards, "Interpretation of protein structures: Estimation of static accessibility," J. Mol. Biol. **55**, 379–400 (1971).

[68]T. Grycuk, "Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation," J. Chem. Phys. **119**, 4817–4826 (2003).

[69]J. Mongan, W. A. Svrcek-Seiler, and A. Onufriev, "Analysis of integral expressions for effective Born radii," J. Chem. Phys. **127**, 185101 (2007).

[70]S. Izadi, B. Aguilar, and A. V. Onufriev, "Protein–ligand electrostatic binding free energies from explicit and implicit solvation," J. Chem. Theory Comput. **11**, 4450–4459 (2015).

[71]F. Richards, "Areas, volumes, packing, and protein structure," Annu. Rev. Biophys. Bioeng. **6**, 151–176 (1977).

[72]M. L. Connolly, "Analytical molecular surface calculation," J. Appl. Crystallogr. **16**, 548–558 (1983).

[73]M. S. Lee, M. Feig, F. R. Salsbury, and C. L. Brooks III, "New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations," J. Comput. Chem. **24**, 1348–1356 (2003).

[74]J. Mongan, C. Simmerling, J. McCammon, D. Case, and A. Onufriev, "Generalized Born model with a simple, robust molecular volume correction," J. Chem. Theory Comput. **3**, 156–169 (2007).

[75].W. A. Svrcek-Seiler, "Force field based investigations on structure and dynamics of RNA molecules," Ph.D. thesis, University of Vienna, Vienna, Austria, 2003.

[76]H. Tjong and H. X. Zhou, "GBr6: A parameterization-free, accurate, analytical generalized Born method," J. Phys. Chem. B **111**, 3055–3061 (2007).

[77]B. Aguilar, R. Shadrach, and A. V. Onufriev, "Reducing the secondary structure bias in the generalized Born model via R6 effective radii," J. Chem. Theory Comput. **6**, 3613–3630 (2010).

[78]R. B. Hermann, "Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area," J. Phys. Chem. **76**, 2754–2759 (1972).

[79]K. A. Sharp, A. Nicholls, R. Fine, and B. Honig, "Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects," Science **252**, 106–109 (1991).

[80]D. Sitkoff, K. A. Sharp, and B. Honig, "Accurate calculation of hydration free energies using macroscopic solvent models," J. Phys. Chem. **98**, 1978–1988 (1994).

[81]E. Gallicchio, M. M. Kubo, and R. M. Levy, "Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation," J. Phys. Chem. B **104**, 6271–6285 (2000).

[82]J. A. Wagoner and N. A. Baker, "Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms," Proc. Natl. Acad. Sci. U. S. A. **103**, 8331–8336 (2006).

[83]A. V. Onufriev and B. Aguilar, "Accuracy of continuum electrostatic calculations based on three common dielectric boundary definitions," J. Theor. Comput. Chem. **13**, 1440006 (2014).

[84]L. Wang, L. Li, and E. Alexov, "pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa," Proteins: Struct., Funct., Bioinf. **83**, 2186–2197 (2015).

[85]H. Tjong and H. X. Zhou, "On the dielectric boundary in Poisson–Boltzmann calculations," J. Chem. Theory Comput. **4**, 507–514 (2008).

[86]M. Nina, D. Beglov, and B. Roux, "Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations," J. Phys. Chem. B **101**, 5239–5248 (1997).

[87]A. Bondi, "Van der Waals volumes and radii," J. Phys. Chem. **68**, 441–451 (1964).

[88]V. Tsui and D. Case, "Theory and applications of the generalized Born solvation model in macromolecular simulations," Biopolymer **56**, 275–291 (2001).

[89]A. Onufriev, D. Case, and D. Bashford, "Effective Born radii in the generalized Born approximation: The importance of being perfect," J. Comput. Chem. **23**, 1297–1304 (2002).

[90]A. A. Rashin, "Electrostatics of ion-ion interactions in solution," J. Phys. Chem. **93**, 4664–4669 (1989).

[91]L. Pratt, G. Hummer, and A. Garcia, "Ion pair potentials-of-mean-force in water," Biophys. Chem. **51**, 147–165 (1994).

[92]B. M. Pettitt and P. J. Rossky, "Alkali halides in water: Ion–solvent correlations and ion–ion potentials of mean force at infinite dilution," J. Chem. Phys. **84**, 5836–5844 (1986).

[93]G. Hummer, D. Soumpasis, and M. Neumann, "Pair correlations in an NaCl-SPC water model simulations versus extended RISM computations," Mol. Phys. **77**, 769–785 (1992).

[94]Y. Marcus, "Thermodynamics of solvation of ions," J. Chem. Soc., Faraday Trans. **87**, 2995–2999 (1991).

[95]Y. Marcus, *Ion Properties* (Marcel Dekker, New York, 1997).

[96]R. Schmid, A. M. Miah, and V. N. Sapunov, "A new table of the thermodynamic quantities of ionic hydration: Values and some applications (enthalpy–entropy compensation and Born radii)," Phys. Chem. Chem. Phys. **2**, 97–102 (2000).

[97]A. K. Soper and C. J. Benmore, "Quantum differences between heavy and light water," Phys. Rev. Lett. **101**, 065502 (2008).

[98]J. Mähler and I. Persson, "A study of the hydration of the alkali metal ions in aqueous solution," Inorg. Chem. **51**, 425–438 (2012).

[99]V. Vivcharuk, B. Tomberli, I. S. Tolokh, and C. G. Gray, "Prediction of binding free energy for adsorption of antimicrobial peptide lactoferricin B on a POPC membrane," Phys. Rev. E **77**, 031913 (2008).
[100]H. Shinto, S. Morisada, M. Miyahara, and K. Higashitani, "A reexamination of mean force potentials for the methane pair and the constituent ion pairs of NaCl in water," J. Chem. Eng. Jpn. **36**, 57–65 (2003).
[101]A. D. MacKerell, Jr. *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," J. Phys. Chem. B **102**, 3586–3616 (1998).
[102]V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, "Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory," Biophys. J. **93**, 3202–3209 (2007).
[103]S. P. Meisburger, S. A. Pabit, and L. Pollack, "Determining the locations of ions and water around DNA from x-ray scattering measurements," Biophys. J. **108**, 2886–2895 (2015).
[104]N. Wang, S. Zhou, P. M. Kekenes-Huskey, B. Li, and J. A. McCammon, "Poisson–Boltzmann versus size-modified Poisson–Boltzmann electrostatics applied to lipid bilayers," J. Phys. Chem. B **118**, 14827–14832 (2014).
[105]A. V. Onufriev and G. Sigalov, "A strategy for reducing gross errors in the generalized Born models of implicit solvation," J. Chem. Phys. **134**, 164104 (2011).
[106]H. Nguyen, D. R. Roe, and C. Simmerling, "Improved generalized Born solvent model parameters for protein simulations," J. Chem. Theory Comput. **9**, 2020–2034 (2013).
[107]G. Sigalov, A. Fenley, and A. Onufriev, "Analytical electrostatics for biomolecules: Beyond the generalized Born approximation," J. Chem. Phys. **124**, 124902 (2006).
[108]G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Pairwise solute descreening of solute charges from a dielectric medium," Chem. Phys. Lett. **246**, 122–129 (1995).
[109]A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized Born model," Proteins: Struct., Funct., Bioinf. **55**, 383–394 (2004).
[110]U. Haberthür and A. Caflisch, "FACTS: Fast analytical continuum treatment of solvation," J. Comput. Chem. **29**, 701–715 (2008).
[111]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[112]H. W. Horn *et al.*, "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew," J. Chem. Phys. **120**, 9665–9678 (2004).
[113]I. S. Tolokh *et al.*, "Why double-stranded RNA resists condensation," Nucleic Acids Res. **42**, 10823–10831 (2014).
[114]Y. Mizuhara, D. Parkin, K. Umezawa, J. Ohnuki, and M. Takano, "Overdestabilization of protein–protein interaction in generalized Born model and utility of energy density integration cutoff," J. Phys. Chem. B **121**, 4669–4677 (2017).
[115]I. S. Tolokh *et al.*, "Multi-shell model of ion-induced nucleic acid condensation," J. Chem. Phys. **144**, 155101 (2016).
[116]T. Macke and D. Case, *Molecular Modeling of Nucleic Acids* (American Chemical Society, Washington, 1998), pp. 379–393.
[117]T. E. Cheatham III, P. Cieplak, and P. A. Kollman, "A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat," J. Biomol. Struct. Dyn. **16**, 845–862 (1999).
[118]A. Pérez *et al.*, "Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformer," Biophys. J. **92**, 3817–3829 (2007).
[119]J. Åqvist, "Ion-water interaction potentials derived from free energy perturbation simulations," J. Phys. Chem. **94**, 8021–8024 (1990).
[120]T. Fox and P. A. Kollman, "Application of the RESP methodology in the parametrization of organic solvents," J. Phys. Chem. B **102**, 8070–8079 (1998).
[121]I. S. Joung and T. E. Cheatham, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations," J. Phys. Chem. B **112**, 9020–9041 (2008).
[122]T. E. Cheatham III and P. A. Kollman, "Insight into the stabilization of A-DNA by specific ion association: Spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]$_2$ in the presence of hexaamminecobalt(III)," Structure **5**(10), 1297–1311 (1997).