

Implicit Solvent with Explicit Ions Generalized Born Model in Molecular Dynamics: Application to DNA

Egor S. Kolesnikov,^{||} Yeyue Xiong,^{||} and Alexey V. Onufriev*



Cite This: *J. Chem. Theory Comput.* 2024, 20, 8724–8739



Read Online

ACCESS |



Metrics & More

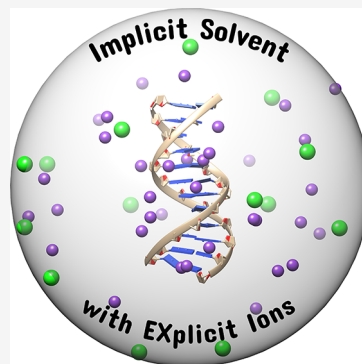


Article Recommendations



Supporting Information

ABSTRACT: The ion atmosphere surrounding highly charged biomolecules, such as nucleic acids, is crucial for their dynamics, structure, and interactions. Here, we develop an approach for the explicit treatment of ions within an implicit solvent framework suitable for atomistic simulations of biomolecules. The proposed implicit solvent/explicit ions model, GBION, is based on a modified generalized Born (GB) model; it includes separate, modified GB terms for solute–ion and ion–ion interactions. The model is implemented in the AMBER package (version 24), and its performance is thoroughly investigated in atomistic molecular dynamics (MD) simulations of double-stranded DNA on a microsecond time scale. The aggregate characteristics of monovalent (Na^+ and K^+) and trivalent (Cobalt Hexamine, CoHex^{3+}) counterion distributions around double-stranded DNA predicted by the model are in reasonable agreement with the experiment (where available), all-atom explicit water MD simulations, and the expectation from the Manning condensation theory. The radial distributions of monovalent cations around DNA are reasonably close to the ones obtained using the explicit water model: expressed in units of energy, the maximum deviations of local ion concentrations from the explicit solvent reference are within $1 k_B T$, comparable to the corresponding deviations expected between different established explicit water models. The proposed GBION model is able to simulate DNA fragments in a large volume of solvent with explicit ions with little additional computational overhead compared with the fully implicit GB treatment of ions. Ions simulated using the developed model explore conformational space at least 2 orders of magnitude faster than in the explicit solvent. These advantages allowed us to observe and explore an unexpected “stacking” mode of DNA condensation in the presence of trivalent counterions (CoHex^{3+}) that was revealed by recent experiments.



1. INTRODUCTION

Ions are essential for cellular function in plants, animals, and humans. It is not surprising that even a slight imbalance of the ionic environment may disrupt cellular operations, leading to severe consequences for the organism. The key role of metal ions in various biological processes, in particular their interactions with nucleic acids, is well-documented.^{1–5} These interactions are critical in determining the mechanical properties of nucleic acids; for instance, Mg^{2+} is essential for the structural stability of RNA molecules.^{6–8} Multivalent counterions surrounding DNA and RNA molecules play critical roles in important phenomena, such as counterion-induced nucleic acid condensation,^{9,10} which is essential for chromatin compaction,^{11,12} affecting gene expression. Understanding the mechanisms behind counterion-induced condensation is important for the development of mRNA vaccines^{13,14} and gene therapies, where targeted delivery of nucleic acid payloads is challenging.^{15–18} Ion–protein interactions affect protein stability^{19,20} and regulate the activity of transcription factors.^{21,22} Ions also modulate the structure of ion-sensing proteins,^{23–25} which play roles in various physiological processes, from inflammation and smooth muscle contraction to short-term and long-term memory and immune response. Specific and nonspecific explicit ion binding is

important for the structure of monoclonal antibodies.²⁶ Thus, the accurate treatment of ions is just as important as the accurate treatment of water in biomolecular simulations.

Atomistic molecular dynamics (MD) simulations are one of the most widely used theoretical tools in biomedical research,^{27–30} which requires accurate and, at the same time, computationally facile representations of the key elements of aqueous solvation—ions and water. Among existing classical models of aqueous solvation, the most accurate ones treat individual solvent molecules and biologically relevant ions explicitly, on par with the target biomolecule. However, this level of detail often entails a significant computational cost, making some simulations difficult or impractical. For instance, close to microsecond simulation times are necessary to properly average various parameters in nucleic acid systems with multivalent ions.^{31,32} A system that contains multiple

Received: June 27, 2024

Revised: August 27, 2024

Accepted: August 28, 2024

Published: September 16, 2024



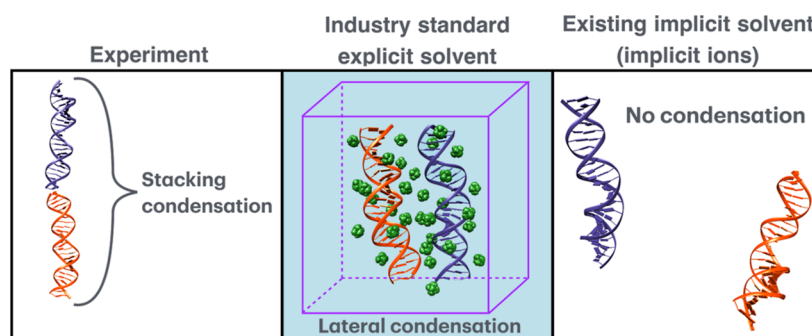


Figure 1. Highly approximate treatment of ions by mainstream implicit solvation models, such as the GB (right) or speed/size limitations of the traditional explicit solvent (middle), can lead to qualitatively wrong outcomes of atomistic simulations compared to experiment (left). In this example, an unusual stacking mode of DNA condensation is shown experimentally³³ to occur for certain DNA sequences (left panel). The implicit solvation methodology that treats the ion atmosphere at a mean-field level cannot describe counterion-induced attraction and condensation of nucleic acids in principle (right panel). Only the well-known lateral DNA condensation mode is likely to be found by a standard explicit solvent simulation (middle panel) due to the high computational costs associated with the significantly larger system size and simulation times needed to reach the stacking condensation mode.

DNA molecules that can form unusual states of counterion-induced condensation, Figure 1 (left), may require a very large solvent box to model properly; a typical “parsimonious” setup where the box is just big enough for the expected but wrong state, Figure 1 (middle), will miss the right one in a typical simulation. In general, an unanticipated conformational transition leading to a significant increase of the size of the “structural unit” would likely be missed in a parsimonious solvent box designed to simulate the original, smaller system.

Apart from the general issue of computational expense and the relatively slow speed of conformational sampling in the explicit solvent,³⁴ the traditional explicit solvent approach has several other, less obvious, but highly significant limitations. Simulating physiologically relevant (intracellular) conditions of small concentrations of many multivalent ions,^{35,36} e.g., 1 mM of free Mg^{2+} or 0.1 mM of Ca^{2+} , in the bulk solvent around a macromolecule is challenging to impossible in a reasonably sized box of explicit water as the number of explicit water molecules required to produce the necessary dilution of even a few tens of unbound ions in the bulk becomes enormous.

The implicit solvation model, an efficient alternative to explicit solvent methods, accelerates conformational sampling in simulations of biomolecules.³⁴ It represents the solvent as a continuum, capturing water’s dielectric and nonpolar properties,^{37–47} and is grounded in solid physical principles.⁴⁸ Currently, the generalized Born (GB) model,^{42,49–79} a variant of implicit solvation, is the de facto “engine” of molecular simulations in the implicit solvent,^a where it enhances computational and sampling efficiency in MD.^{86–118}

The GB is a “trail blazer” method, which can justify the additional approximations to reality made¹¹⁹ by this model relative to the traditional explicit solvation. For example, it was the GB model that enabled early simulations^{87–96} of “de novo” protein folding, long before the advent of highly specialized supercomputers made it possible in the explicit solvent.¹²⁰ Atomistic folding simulations using the GB model are now routine on GPUs,¹²¹ accessible to most computational laboratories. The GB-based MD simulations are particularly well suited for modeling of large-scale motions in biomolecules,^{97,98} including the nucleosome and its DNA;⁹⁹ for the latter, speed-ups of conformational sampling up to 2 orders of magnitude can be expected.³⁴ Recent GPU implementations¹²² have significantly boosted the GB model’s performance,

enabling faster simulations^{34,123} of large¹²⁴ structures previously challenging for explicit solvents.

Yet, the canonical GB framework,⁵³ foundational to various GB flavors¹²⁵ (models) available in major simulation packages, has a notable limitation: it is designed to approximate the effects of monovalent ions alone, representing them as a diffuse, mean-field continuum (Debye screening), neglecting ion correlation effects, as well as differences between ions such as Na^+ versus K^+ ,¹²⁶ or specific ion binding. Multivalent ions remain out of reach. Electrostatic potential distributions around biomolecules, obtained with explicit ions,^{31,127,128} often differ significantly from those produced using more approximate ion models such as the GB, which treat ions as a diffuse continuum. The fundamental limitations have their consequences for practical simulations: for example, the canonical GB model is not suitable for simulations of the DNA condensation process, see Figure 1. Straightforward attempts to use explicit ions directly within the canonical GB lead to qualitatively wrong results, see Figure 2. Collectively, these limitations make the canonical GB unsuitable for high-accuracy simulations of many biologically relevant systems and

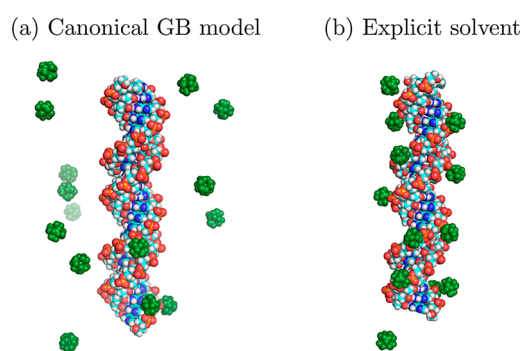


Figure 2. An MD simulation based on the canonical GB model generates a qualitatively wrong distribution of multivalent ions around DNA (a), in which explicit trivalent ions (CoHex^{3+} , green) fail to condense onto the double-stranded DNA. The expected counterion condensation is reproduced by a fully explicit solvent (b). The distributions are adapted from ref 128. The qualitative inadequacy of CoHex^{3+} distribution seen in (a) stems from a fundamental limitation of the canonical GB model when applied to disconnected, i.e., water-separated, topologies,¹²⁸ such as ions around a molecule.

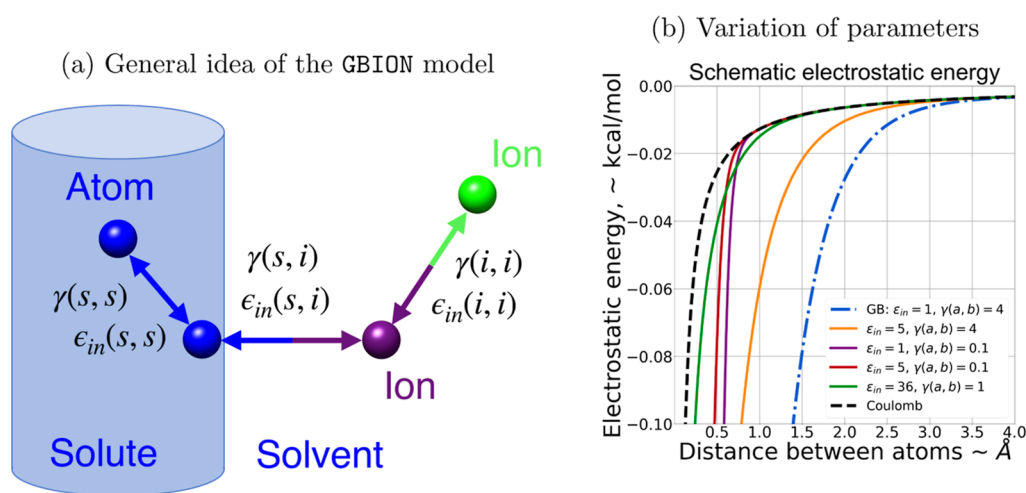


Figure 3. (a) Adaptation of the general implicit solvent/explicit ions framework for MD simulations. Each pair (a,b) of interacting atom types has its own parameter set describing their interactions, eq 3. (b) Dependence of the total electrostatic energy on the distance d_{ij} between two charged atoms for different values of $\gamma(a,b)$ and $\epsilon_{in}(a,b)$. Decreasing $\gamma(a,b)$ makes the interacting potential closer to Coulomb's law with a stronger effect for larger values of d_{ij} . The smaller the $\gamma(a,b)$, the closer $E_{ij}(d_{ij})$ is to the pure Coulomb electrostatics (black dashed line). Increasing $\epsilon_{in}(a,b)$ from 1 to ϵ_{out} makes $E_{ij}(d_{ij})$ deviate further from the canonical GB and closer to Coulomb's law for all d_{ij} . In all of our calculations, we use the default AMBER value of $\epsilon_{out} = 78.5$. An example functional dependence for one of the optimized parameter pairs [$\epsilon_{in}(a,b) = 36$ and $\gamma(a,b) = 4$] is also shown (green solid line).

processes—all where detailed ion interactions are crucial. The very approximate treatment of ions by the current GB model was likely adequate for a number of problems at much shorter time-scales of yesterday; we argue that it is no longer the case today.

Recognizing this gap, here we introduce the Implicit Solvent with EXplicit Ions approach and its practical implementation GBION, which enhances the capabilities of atomistic simulations by combining explicit ion treatment with the benefits of implicit water solvation. This work lays out the foundation of the new model; as a concrete example of its potential application here, we focus on the important case of DNA.

2. THEORY

2.1. Canonical GB Model: Brief Introduction. The hydration free energy plays a pivotal role in determining the interaction potential between solutes in environments characterized by water. This energy, denoted as ΔG_{solv} , can be decomposed into two main contributions: electrostatic, ΔG_{el} , and nonpolar, $\Delta G_{non-polar}$, contributions¹²⁹

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{non-polar} \quad (1)$$

which are typically estimated independently. A widely used, computationally efficient approximation to calculating the nonpolar term ($\Delta G_{non-polar}$) is based on the assumption that $\Delta G_{non-polar}$ is proportional to the solvent-accessible surface area (SASA).

The GB model provides an approximate way, relative to the exact solution of the Poisson equation, to calculate the electrostatic part, ΔG_{el} , of the solvation free energy of any singly connected solute or tightly bound complex.¹¹⁹ The canonical GB approximation is based on the equation originally proposed by Still et al.⁵³

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

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 and q_j 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 . Generally, R_i is larger than the atom's intrinsic radius ρ_i , the two becoming equal for a single isolated ion or atom. Multiple practical algorithms¹²⁵ exist for calculating R_i from the molecular geometry defined by a set of its $\{\rho_i\}$ and interatomic distances; fast analytical flavors of the GB are most often used in MD simulations.

2.2. Proposed GBION Model. The canonical GB model¹²⁵ exhibits remarkable accuracy when applied to single, topologically connected molecular structures. This accuracy arises from the fact that these structures are topologically equivalent to a sphere, where the GB model becomes essentially exact in the limit of the infinite solute dielectric.^{130–132} Conversely, it is not surprising that the model's accuracy deteriorates for topologies distinctly different from a sphere: “disconnected” topologies, such as solvent-separated molecules. Understanding and simulating these “disconnected” topologies are critical for elucidating various biologically relevant processes, including ligand binding and intricate interactions of DNA with the surrounding ions.

A general approach, suitable for Monte Carlo simulations of disconnected solutes, was proposed earlier by Tolokh et al.¹²⁸ However, that approach implies a discontinuous, and therefore nondifferentiable, solvation energy function. Consequently, forces cannot be calculated, and the approach cannot be

implemented “as is” in MD simulations. A version with a differentiable energy $\Delta G_{\text{el}}^{ij}$ is presented here.

We begin here by identifying a new charge-pairwise $\Delta G_{\text{el}}^{ij}$ suitable for multiconnected geometries (disconnected solutes) to replace the $\Delta G_{\text{el}}^{ij}$ of the canonical GB, eq 2. The key requirements are that the resulting functional form is differentiable and “simple” enough to make adaption for MD possible. In contrast to the canonical $\Delta G_{\text{el}}^{ij}$, the new function depends on the topology of the interacting atoms (i.e., ion–ion, ion–solute, or solute–solute), see Figure 3. The total interaction electrostatic energy for two charges with the proposed $\Delta G_{\text{el}}^{ij}$ takes the form

$$E_{ij}(d_{ij}) = \frac{q_i q_j}{\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b}) d_{ij}} - \left(\frac{1}{\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b})} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{q_i q_j}{\sqrt{d_{ij}^2 + R_i R_j \exp(-d_{ij}^2 / (\gamma(\mathbf{a}, \mathbf{b}) R_i R_j))}} \quad (3)$$

where $(\mathbf{a}, \mathbf{b})$ denotes the type (topology) of the pair of interacting atoms, e.g., $\mathbf{a}$ = ion and $\mathbf{b}$ = solute. This equation captures the general nature of charge–charge interactions mediated by the solute or solvent, introducing $\gamma(\mathbf{a}, \mathbf{b})$ and $\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b})$ coefficients for tuning the interaction energy. The parameter $\gamma(\mathbf{a}, \mathbf{b})$ controls how the function approximates the interaction between solvent-separated charges (see Figure 3); it was shown previously¹²⁸ that, unlike the canonical GB ($\gamma(\mathbf{a}, \mathbf{b}) = 4$ in eq 3), pure Coulomb interaction works reasonably well for solvent-separated charges, including ions^b. In eq 3, the pure Coulomb regime is approximated by setting $\gamma(\mathbf{a}, \mathbf{b}) \ll 1$; note that $E_{ij}(d_{ij}) \rightarrow \frac{q_i q_j}{\epsilon_{\text{out}} d_{ij}}$ when $\gamma(\mathbf{a}, \mathbf{b}) \rightarrow 0$.

Additional flexibility to fine-tune charge–charge interactions is afforded by the coefficient $\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b})$, which scales the part of the electrostatic energy (see Figure 3). Increasing $\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b})$ from 1 to ϵ_{out} makes $E_{ij}(d_{ij})$ further from canonical GB and closer to Coulomb’s law for all d_{ij} . The introduction of specific coefficients $\gamma(\mathbf{a}, \mathbf{b})$ and $\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b})$ for interactions between solute atoms and ions (both anions and cations) enables the model to fairly accurately mimic the localized effects of ions on solvation energetics. For solute–solute interactions, $\gamma(\mathbf{a}, \mathbf{b}) = 4$ and $\epsilon_{\text{in}}(\mathbf{a}, \mathbf{b}) = 1$, meaning that the canonical GB is used.

2.3. Setting the Salt Concentration within the Model.

Since there are no periodic boundary conditions in this implicit GB solvent model, an explicit ion in such a system would not be “imaged” back as it moves away from the solute. Consequently, over extended simulations, ions would gradually drift away from the solute; given an infinite volume and a finite number of particles, entropy always prevails. To address this issue, we have applied an NMR-type restraint between each ion and a user-selected “anchor”, which can be either a single atom near the center of the solute molecule or the center of mass of a group of atoms, see Figure 4 and the movie in Supporting Information. The restraint operates as follows: once the ion exceeds a predefined distance (parameter r_3 in the AMBER *DISANG* file) from the anchor point, the ion begins experiencing a harmonic restraining potential (switched on by *nmropt* setting in AMBER and specified in the *DISANG* file): the force constant is chosen to be large enough so that possible excursions of the ion beyond the distance r_3 should be negligible compared to r_3 . This approach allows one to define the volume of the simulation system as the volume of a sphere

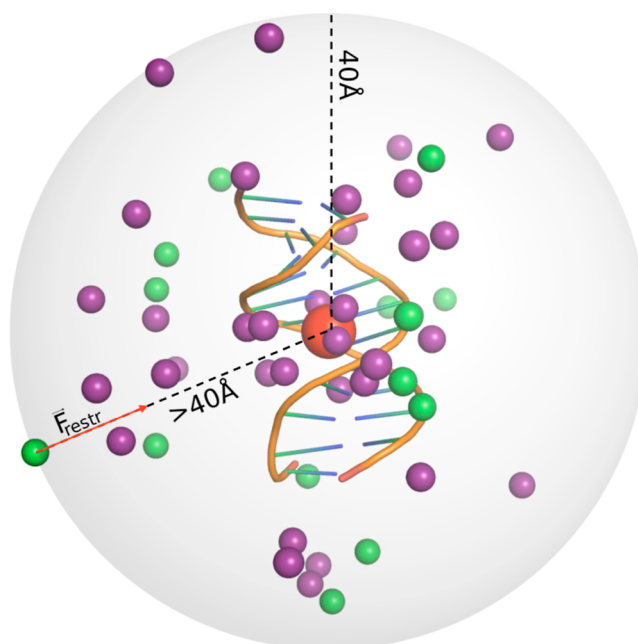


Figure 4. Setting the simulation volume and proper ion concentration in GBION. A restraining force F_{restr} (red arrow) prevents the ions (purple and green) from drifting beyond a user-specified distance (40 Å in this example) from the anchor point (center of the red sphere). Shown is an example of the force restraining the ions to the center of mass of all of the phosphorus atoms of the DNA fragment. The snapshot is taken from a simulation using GBION, see the movie in the Supporting Information.

of radius r_3 , see the example of the simulation system with the restraining sphere in Figure 4.

To set the ion concentration in a simulation, we use the SLTCAP¹³³ method. It utilizes the equation below to calculate the numbers of anions and cations that should be added to the system

$$N_{\pm} = V_w c_0 \exp \left[\mp \text{arsinh} \left(\frac{Q}{2V_w c_0} \right) \right] \quad (4)$$

where N_+ is the number of cations to be added, N_- is the number of anions to be added, Q is the total charge of the molecule (solute), V_w denotes the volume of the solvent in the system in $\text{\AA}^3$, and c_0 is the desired salt concentration in units of particles per $\text{\AA}^3$; to convert from the common M/L, one should multiply the M/L value by 6.02×10^{-4} . The solvent volume V_w is defined as the volume of the enclosing sphere minus the volume of the solute. For instance, in the case of our example restraining sphere (see Figure 4), the volume $V_w = \frac{4 \times \pi \times R^3}{3} - V_{\text{solute}}$, where V_{solute} is the volume of simulating DNA. Substituting 40 Å for R yields $V_w = \frac{4 \times \pi \times 40^3 \text{\AA}^3}{3} - 8800 \text{\AA}^3 = 2.59 \times 10^5 \text{\AA}^3$ and

Table 1. Aggregate Characteristics of Cation Distributions around DNA Obtained Using GBION Are Compared with Those Based on Explicit Water Models. Reference Values from the Experiment¹³⁵ and Manning Condensation Theory¹³⁶ Are Shown Where Available. The GBION Model Reproduces the Aggregate Parameters of Cation Distributions with Reasonable Accuracy.^a

water model	number of cations associated with DNA		degree of DNA charge neutralization, %		cation affinity, $k_B T$	
	Na ⁺	K ⁺	Na ⁺	K ⁺	Na ⁺	K ⁺
OPC	30.8 ± 0.5	30.9 ± 0.5	56 ± 4	54 ± 3	−2.28 ± 0.02	−2.22 ± 0.02
SPC/E	33.7 ± 0.5	N/A	62 ± 4	N/A	−2.45 ± 0.02	N/A
GBION	33.7 ± 0.5	34.2 ± 0.5	53 ± 2	54 ± 2	−2.07 ± 0.02	−2.09 ± 0.02
experiment	37 ± 2	N/A	N/A	N/A	N/A	N/A
manning theory	36.5	36.5	76	76	N/A	N/A

^aThe deviations are calculated as standard errors of the mean.

$$N_{\pm} = 2.59 \times 10^5 \text{ Å}^3 \times 6.02 \times 10^{-4} \frac{1}{\text{Å}^3 \times \text{M/L}} \times 0.145 \text{ M/L} \\ \times \exp \left[\mp \operatorname{arsinh} \left(\frac{-22}{2.59 \times 10^5 \text{ Å}^3 \times 6.02 \times 10^{-4} \frac{1}{\text{Å}^3 \times \text{M/L}} \times 0.145 \text{ M/L}} \right) \right] = \frac{+36}{-14}$$

In this specific example, we add 36 cations and 14 anions to achieve the bulk salt concentration of 0.145 M in the simulation volume.

However, in practice, it is often the case that $V_{\text{solute}} \ll V_w$ and therefore one can simply neglect the solute volume and calculate the solvent volume V_w as the sphere volume. Consequently, in most cases, one can use the approximate equation below to calculate the numbers of cations and anions

$$N_{\pm} = \frac{4\pi R^3}{3} \times 6.02 \times 10^{-4} \times c_0 \\ \exp \left[\mp \operatorname{arsinh} \left(\frac{Q}{2 \times \frac{4\pi R^3}{3} \times 6.02 \times 10^{-4} \times c_0} \right) \right] \quad (5)$$

where R denotes the radius of the confining sphere in Å and c_0 is the desired salt concentration in M/L. If the formula returns noninteger N_+ and N_- , one should take the closest integer for either N_+ or N_- and subtract it from the absolute value of Q to find N_- or N_+ , respectively.

3. PROPOSED IMPLICIT SOLVENT/EXPLICIT IONS GB MODEL APPLIED FOR SIMULATIONS OF DNA

To evaluate the proposed implicit solvent/explicit ions model, we focus on its ability to reproduce ion distributions around highly charged DNA molecules in water as well as the ability of the method to maintain the proper structure of the DNA in the simulation. Specific validation involves comparing the distributions of cations around DNA with those calculated in explicit solvent simulations and examining the DNA structure's stability during the simulation with the GBION model. The similarity of ion distributions and the preservation of DNA structure over extended periods emphasize the GBION model's suitability for MD simulations of DNA. Additionally, we assess the computational efficiency of the GBION model relative to traditional explicit water models.

Given the propensity of DNA to condense in solutions containing ions with a charge of +3e or higher,^{9,134} the model's

ability to simulate DNA condensation processes is crucial – testing the model for the expected DNA association in the presence of CoHex³⁺ ions is another critical test. As we shall see below, our simulations reveal not only the expected lateral but also, notably, “stacking” condensation of DNA duplexes—a phenomenon only very recently observed experimentally.³³

First, we compare distributions of sodium and potassium ions around DNA obtained using the GBION model with the corresponding distributions obtained using the explicit water model and check the stability of the DNA in these simulations. Following this, the GBION model is applied to simulate double-stranded DNA in the presence of trivalent ions (CoHex³⁺); we perform a qualitative comparison of the distribution to the published results. To check the applicability of the GBION model for the simulation of DNA condensation, we simulate a pair of DNA duplexes in the presence of CoHex³⁺ ions to observe DNA attraction. We observe lateral condensation of DNA duplexes, which agrees with experimental results.⁹ We also unexpectedly observe a “stacking” condensation mode of DNA duplexes. That condensation mode has been revealed recently³³ (and was not known to us when we ran the simulations). Details of the simulations and comparisons are presented below.

3.1. Distributions of Monovalent Na⁺ and K⁺ Ions around the DNA Duplex.

We use GBION to calculate monovalent Na⁺ and K⁺ ion distributions around the homopolymeric 25 base pair (bp) DNA duplex with the sequence: GCATCTGGGCTATAAAAGGGCGTCG. First, we have compared the overall characteristics of cation distributions around DNA obtained using the GBION model and explicit water models with each other and with available experimental results, see Table 1. The aggregate characteristics we compare with available references are the number of ions associated with the DNA, ion affinity for the DNA, and the degree of DNA charge neutralization. For the number of sodium ions associated with the DNA, we used the results of previous theoretical and experimental studies¹³⁵ and our recent simulations. The degree of DNA charge neutralization is compared with the results of the explicit solvent simulations and with predictions of a well-established theory.¹³⁶

One can see that the aggregate characteristics of the cation distribution around DNA obtained using the GBION model are close to the ones obtained using the explicit water model. A previous study¹³⁷ demonstrated that simulations using widely adopted explicit water models can approximate the number of Na⁺ ions associated with DNA to within ~12% of the experimental value. The GBION model achieves a comparable level of agreement with the experiment, Table 1. It is important

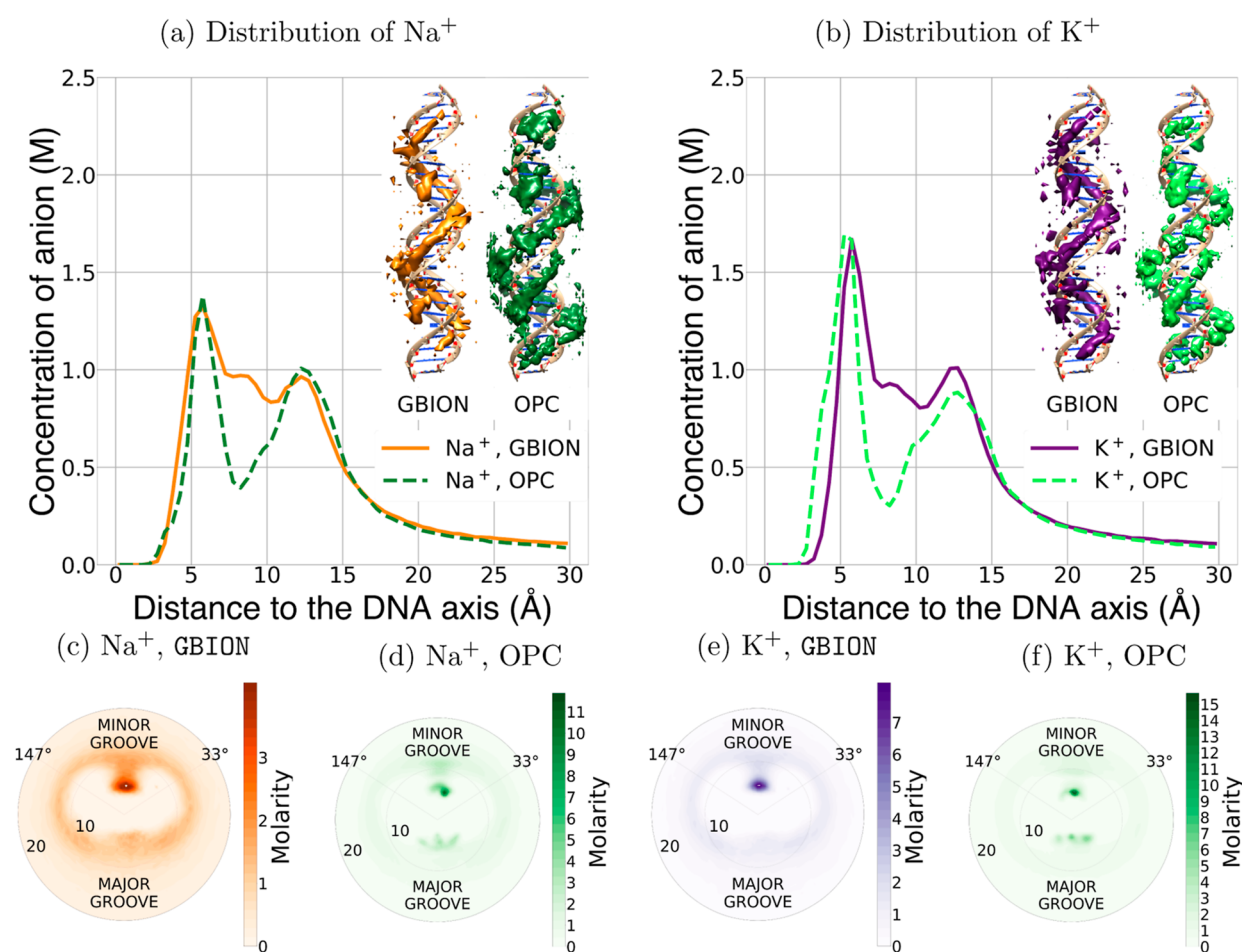


Figure 5. Radial distributions and ion density maps of Na⁺(a) and K⁺(b) obtained using the GBION model are compared with those obtained using the explicit solvent [OPC water¹³⁸ with J/C (Joung/Cheatham) ion parameters¹³⁹]. Radial-angle distributions of Na⁺ and K⁺ are presented in panels (c–f); the water model used to obtain each distribution is indicated above the panels. The distributions of sodium and potassium around DNA obtained with the GBION model reproduce those obtained using the explicit solvent with reasonable accuracy, with even the most significant difference corresponding to a variation in the ion–DNA affinity of less than $k_B T$. The GBION and explicit solvent distributions in panel (a) differ the most at a distance of ~ 8 Å from the DNA helical axis, the same as those in panel (b).

to note that differences between specific details of the simulation protocol can influence the simulation outcomes within that level of accuracy; these details include the types of restraints used (e.g., with the entire DNA restrained as in this work vs only the middle base pair kept fixed as in ref 137) and the geometry of the simulation enclosure (e.g., a truncated octahedron box for explicit water simulations in this study vs a rectangular box in ref 137 vs a restraining sphere for GBION simulations). With this caveat in mind, one can reasonably conclude that GBION reproduces the aggregate characteristics of the cation distribution around DNA at the same level of accuracy that may be expected from the established explicit water models.

Below is a more detailed comparison of the distributions of sodium and potassium obtained using GBION with the ones obtained using simulations with the explicit water model OPC, see Figure 5. One can see that the ion distributions generated by GBION reproduce the explicit water MD results with a reasonable accuracy, comparable with the accuracy of the explicit water simulations themselves.¹³⁷ The uncertainty of the explicit solvent models in reproducing the ion distributions can be estimated by comparison with the corresponding distributions obtained by using different established explicit water models. Specifically, in ref 137, the SPC/E water model

was found to yield ion distributions that differ the most from those obtained via other water models tested in that work. To estimate the uncertainty of the ion distribution expected from an explicit water model, we have compared the distribution of Na⁺ ions obtained using different explicit water models (see Figure S1). The uncertainty can be estimated as the ratio of the ion concentrations at the distribution peaks. The concentration of Na⁺ ions at the first peak in the SPC/E water model is about 1.6 times higher than that obtained in the OPC water. This concentration difference translates into the difference of Na⁺–DNA interaction potentials for these two explicit water models of an order of $0.55 \times k_B T$. Another representative metric of uncertainty can be the ratio between the areas integrated under the first peak of the ion distributions obtained by using different water models. In our simulations, this ratio is 1.62, consistent with the uncertainty estimate based on the peak concentration ratio estimated above. The uncertainty puts in perspective the accuracy of the proposed GBION model: the largest difference between the GBION-derived ion distribution and the one obtained using the OPC water model, Figure 5, is still less than $k_B T$ in energy units, which is reasonable and comparable to the difference between the ion distributions obtained in SPC/E vs OPC explicit water models.

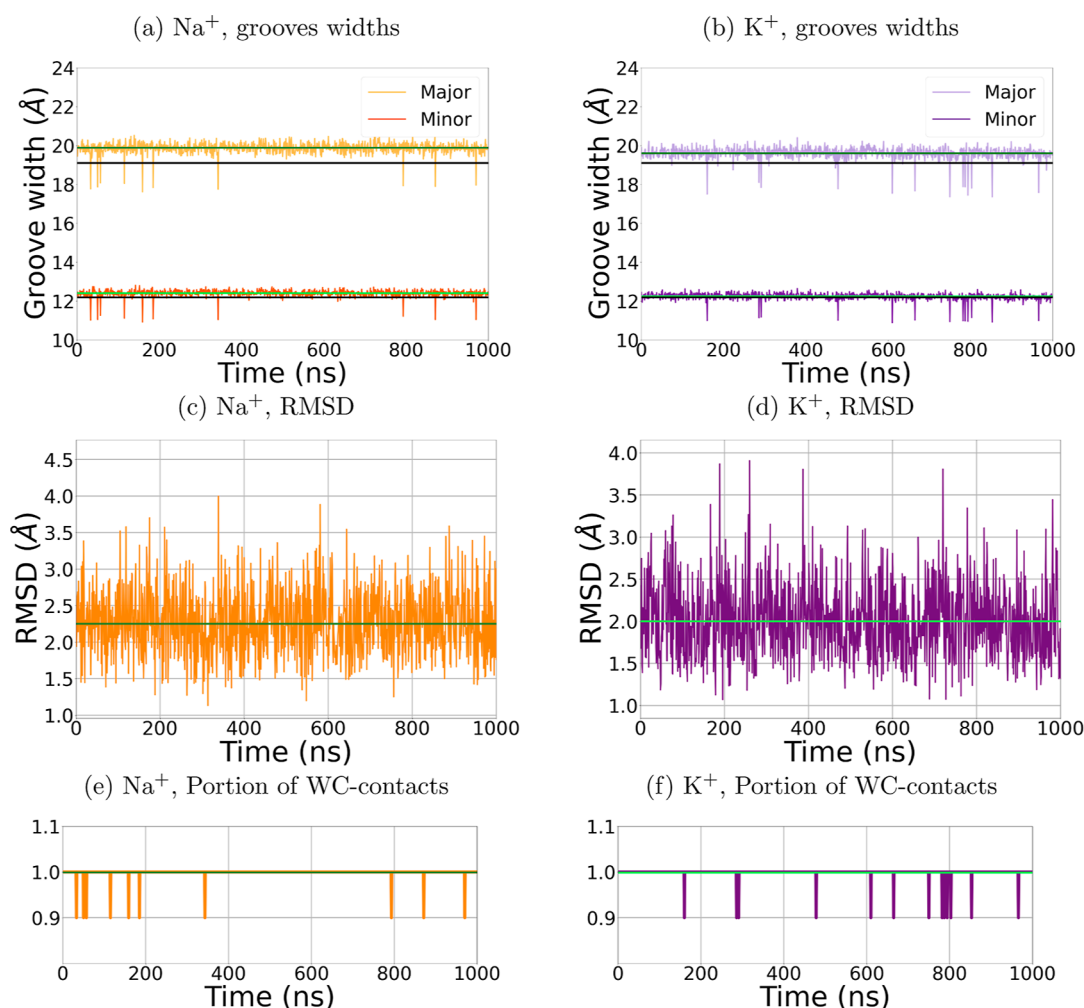


Figure 6. (a,b) Widths of the major (light orange and light purple) and minor (dark orange and dark purple) grooves in comparison with the experimental values¹⁴⁰ (black lines) in GBION simulations with NaCl (a) and KCl (b). Average values over the simulation time are indicated by green lines. The widths of the grooves agree with the corresponding experimental values. (c,d) All-atom (excluding terminal nucleotides) rmsd to crystal structure reference in GBION simulations with NaCl (c) and KCl (d). (e,f) Fraction of stable WC-pairs (averaged over the DNA and excluding the terminal nucleotides) in GBION simulations with NaCl (e) and KCl (f).

The most noticeable difference between the distributions of ions obtained using the GBION model and explicit solvent is at a distance of ~ 8 Å from the DNA helical axis. As observed from the radial-angle ion distributions shown in Figure S2, the distributions of sodium and potassium ions at approximately 8 Å from the DNA helical axis are more diffuse in GBION simulations compared to those in the explicit water simulations. The distributions reveal two binding regions, one inside and one outside the minor groove. In the explicit water simulations, there are almost no ions present between these regions. In contrast, GBION simulations show a significant presence of ions in the gap between the binding areas. It is this difference in ion distributions within the minor groove that accounts for the observed discrepancy between the radial distributions obtained using the explicit water model and the GBION model. Importantly, however, the overall number of ions bound to the DNA does not differ appreciably between GBION simulations and the explicit solvent ones. In fact, the difference is no more than 2 ions, which corresponds to $\sim 4\%$ of the degree of DNA charge neutralization. This small difference is less than that between different explicit solvents¹³⁷ in the same context. It means that the GBION model correctly

captures the overall physics of ion binding to the DNA. We have also calculated the distributions of Cl[−] around DNA in simulations of DNA in the presence of NaCl and KCl (see Figure S3). The distributions obtained in GBION are close to those in the explicit solvent and are insensitive to the cation type present in the solution.

A crucial advantage of the implicit compared to the explicit solvation is the high speed of conformational sampling that the implicit solvent model offers.³⁴ We find that the proposed GBION retains this advantage. To estimate how fast the ion distribution equilibrates with GBION, we have compared the distributions averaged over different time intervals: over 10 ps, 100 ps, 1 ns, and 10 ns for GBION simulations and over 100 and 300 ns for the explicit water simulations (see Figure S4). According to our estimates, the ion atmosphere converges at least 2 orders of magnitude faster in GBION than in the explicit solvent.

3.2. DNA Remains Stable in GBION MD Simulations.

To verify that the DNA structure remains stable in a typical MD simulation, we have simulated Dickerson–Drew dodecamer¹⁴⁰ using the GBION model for 1 μ s in the presence of NaCl and, separately, of KCl, at physiologically relevant

conditions. Throughout the simulations, the ions were restrained to remain within 40 Å of the center of mass of the phosphorus atoms of the DNA using a restraining force (see an example of the restraining force in Figure 4 and a movie in the Supporting Information). We assessed the DNA stability using three key metrics: the average portion of stable Watson–Crick (WC)-pairs, rmsd from the X-ray reference, and the widths of the DNA grooves, all evaluated over the entire microsecond-long trajectory, see Figure 6.

The average WC-value was 0.99 ± 0.03 for NaCl and 0.99 ± 0.04 for KCl, indicating that the WC pairs remained stable throughout the simulation. The minor groove width in the presence of NaCl and KCl was 12.4 ± 0.2 and 12.2 ± 0.2 Å, respectively. The width of the major groove in the presence of NaCl and KCl was 19.9 ± 0.3 and 19.6 ± 0.3 Å, respectively. These widths are in agreement with those calculated from the reference structure¹⁴⁰ (12.2 for the minor groove and 19.1 Å for the major groove), see Figure 6. The average rmsd from the X-ray reference structure was 2.0 ± 0.5 Å for NaCl and 2.3 ± 0.5 Å for KCl, respectively, slightly higher than the results obtained with the explicit water model (1.5 Å). The consistency of the structural parameters of the simulated DNA with the experimental references demonstrates the DNA structure's stability during simulations with the GBION model, confirming the model's applicability to simulating DNA duplexes in the presence of monovalent ions.

3.3. Distribution of Trivalent CoHex Ions around Double-Stranded DNA. To evaluate the suitability of the proposed implicit solvent/explicit ions GB method for simulating solutions with multivalent ions surrounding DNA, where ion–ion correlations and other explicit solvent effects may significantly affect the ion distributions,¹⁴¹ we applied the GBION model to investigate the interactions between trivalent CoHex³⁺ counterions and DNA duplexes. Accurately reproducing the distributions of multivalent ions around nucleic acids is crucial as subtle aspects of these distributions influence the propensity of nucleic acids to aggregate.^{31,142}

The CoHex³⁺ ion distribution around a DNA duplex, as determined by GBION simulations, is compared with these from all-atom MD simulations utilizing the explicit water model, Figure 7.

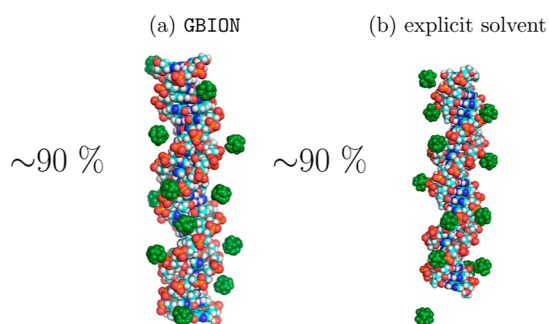


Figure 7. Snapshots from simulations of DNA in solution with CoHex³⁺ show the distribution of CoHex³⁺ ions (green) in simulation with the GBION model (a) and previously reported explicit solvent distribution^{31,143} (b), which agree with each other quantitatively. The degree of DNA charge neutralization obtained in the simulation carried out using the GBION model is equal to the one observed in the simulation with the explicit water model; the corresponding values are listed next to each structure.

The CoHex³⁺ distribution estimated based on the simulations carried out using the GBION model, as one can see in Figure 7, qualitatively reproduces the distribution of trivalent ions obtained using the explicit water model. The degree of DNA charge neutralization matches the one observed in simulation of DNA with CoHex³⁺ using the explicit water model¹⁴³ (~90%).

According to the previously developed multishell model of counterion-induced condensation of nucleic acids,^{31,143} CoHex³⁺ ions bound within the “external” ion-binding shell (12–16 Å from the helical axis) of the double helix facilitate the condensation. The probability of condensation increases with the number of ions attached to the external shell. The DNA charge neutralization is also a prerequisite for condensation. Our next step involves verifying whether DNA condensation occurs in the presence of trivalent ions, further assessing the effectiveness of GBION.

3.4. Counterion-Induced DNA Condensation. To assess the applicability of the GBION model for simulating the DNA condensation process, we employed the GBION model to simulate the condensation of DNA induced by CoHex³⁺ ions. The simulation began with the two DNA duplexes positioned at a distance from each other, followed by the removal of constraints to allow for the DNA attraction to lead to the condensation. The initial conformation and the observed condensed states are illustrated below.

During the simulations, we identified two distinct conformational states of condensed DNA: the expected lateral condensation and a “stacking” condensed state, see Figure 8.

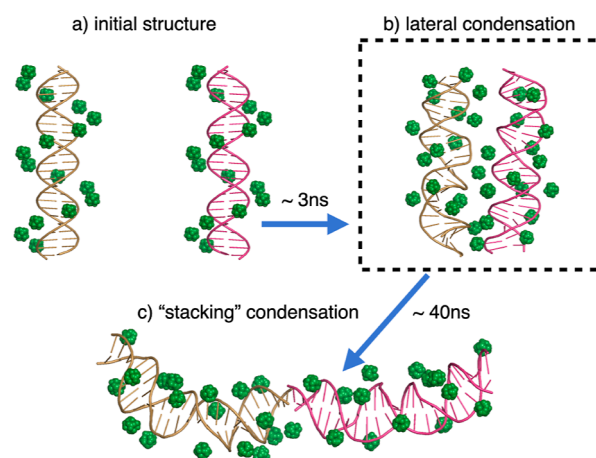


Figure 8. DNA condensation induced by CoHex³⁺ ions. Starting from two separated duplexes (a), a GBION MD simulation reaches the expected “lateral” condensed state (b) on a nanosecond time scale. Continuing the GBION run predicts an unexpected, novel “stacking” condensed state (c), recently validated experimentally.³³ This state would likely not have been identified in a conventional explicit solvation simulation, which would likely use the solvent box (black dashed outline) just large enough to contain the anticipated lateral configuration of the duplexes. A movie is available in Supporting Information.

The latter has been observed and studied in highly concentrated DNA solutions.³³ It was also observed using SAXS.¹⁴⁴ The simulation time required to observe the “stacking” condensed state in GBION is just ~43 ns. For this simulation, we have calculated the average Watson–Crick (WC) value over both duplexes, finding it to be 0.99 ± 0.03 ,

which indicates that the DNA remained stable throughout the condensation process induced by trivalent ions.

As we have already mentioned, sampling of the conformational space by the system in simulations carried out using the GBION model is faster than that of the explicit water simulations by 1.5–2 orders of magnitude, or more. This speed up translates, e.g., the simulation time of ~ 43 ns using the GBION model, Figure 8, into ~ 4 μ s of effective simulation time in the explicit water. Also, to observe the “stacked” condensation mode, the simulation box for the system with the explicit water should also be large enough to accommodate two stacked DNA duplexes. A back-of-the-envelope estimate^c suggests that to observe the stacking condensation mode in the explicit solvent, one would need at least ~ 0.5 years on the GPU we used (NVIDIA RTX 3080). On the other hand, the simulation carried out using the GBION model took a few days using the same computational resources. Also note that the unexpected stacked state would have been missed in a typical explicit solvation simulation designed to explore the expected condensation pattern; the solvent box size (black dashed outline in Figure 8) would have been just large enough to accommodate the expected lateral condensed state, too small for the stacked one.

4. METHODS

4.1. GBION Implementation Notes. The proposed framework can be effectively implemented in any computational environment that supports the general Born model. A key feature of this implementation is the differentiated treatment of various interacting atomic pairs. Specifically, each interacting atomic pair is characterized by unique coefficients, denoted as $\gamma(a,b)$ and $\epsilon_{in}(a,b)$, as shown in Figure 9.

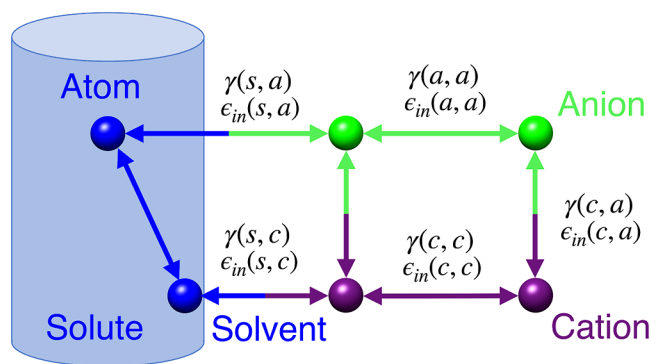


Figure 9. Detailed overview of the general implementation of the GBION framework for MD simulations. The specified coefficients are incorporated into eq 6.

This implementation involves a modified equation for calculating the electrostatic interactions between atomic pairs

$$E_{ij}(d_{ij}) = \frac{q_i q_j}{\epsilon_{in}(\mathbf{a}, \mathbf{b}) d_{ij}} - \left(\frac{1}{\epsilon_{in}(\mathbf{a}, \mathbf{b})} - \frac{1}{\epsilon_{out}} \right) \frac{q_i q_j}{\sqrt{d_{ij}^2 + R_i R_j \exp(-d_{ij}^2 / (\gamma(\mathbf{a}, \mathbf{b}) R_i R_j))}} \quad (6)$$

The current implementation details within the AMBER software package are provided in Supporting Information. Our analysis indicates that, in general, GBION parameters in the

following range yield an accurate representation of the modeled ion distributions:

- $\gamma(s,a)$, $\gamma(a,a)$, and $\gamma(c,c)$ are approximately 4, maintaining the canonical GB approximation for solute–anion, anion–anion, and cation–cation interactions.
- $\gamma(s,c)$ and $\gamma(a,c)$ are near 0, aligning solute–cation and cation–anion electrostatic interactions closely with Coulomb’s law.
- $\epsilon_{in}(s,a)$ and $\epsilon_{in}(a,a)$ are slightly lower than, but close to, the dielectric constant of water (ϵ_w).
- $\epsilon_{in}(s,c)$, $\epsilon_{in}(a,c)$, and $\epsilon_{in}(c,c)$ for monovalent ions are approximately equal to $\sqrt{\epsilon_w}$.

The specific values of these coefficients depend on other parameters within the employed GB model as well as on the particular DNA and ion models used in the implementation and on the ion charge.

4.1.1. Optimization of the Model Parameters. The following parameters were optimized for best fit against the reference explicit solvent values and distributions: radial distributions of sodium and potassium ions and the degree of the DNA charge neutralization by CoHex³⁺ ions. We carried out a number of simulations with different values of GBION parameters to find the sets of parameters that best reproduce the explicit solvent distributions of ions around the DNA. Each trajectory was made with the protocol described below. During the process of searching through the parameter space, we have revealed the following trends, which helped with the optimization process:

- The larger the $K_{GB}(s, *)$ (see Supporting Information for implementation details within AMBER), the stronger the ions bind to the phosphate backbone of DNA. Variation of $K_{GB}(s, *)$ modulates the height of the second peak of ion distribution around DNA (see Figure 5).
- The larger the $K_e(s, *)$ is, the better the ions accumulate in the ion-binding sites of the DNA structure
- The larger the $K_e(c, c)$ is, the smaller the absolute value of affinity of cations to DNA is—the stronger the ions repel from each other, so if some ions are already bound to the DNA, additional ions tend to bind less frequently.

Using the aforementioned trends as guidance, we varied the parameters K_{GB} in the range of 0–2 and K_e in the range of 1–78.5 to find the optimal values reported below in the simulation protocols.

4.2. Simulation Protocols. **4.2.1. DNA–NaCl and DNA–KCl Systems: Explicit Water MD Simulations.** The explicit water MD simulations of the 25 bp-long DNA duplex with the sequence (GCATCTGGGCTATAAAAGGGCGTCG) in solutions of NaCl and KCl were carried out using the bsc1 DNA force field using AMBER 18. The 25 bp duplex is constructed in the canonical B-form using Nucleic Acid Builder.¹⁴⁵ The DNA duplex was solvated with 29255 OPC water molecules, 104 cations (Na^+ or K^+), and 56 Cl^- ions. For the simulation of DNA with the NaCl and SPC/E water model 29213 water molecules, 103 Na^+ ions and 55 Cl^- ions were used. The truncated octahedron was set as a form of simulation box to approximate the results obtained using spherical restraints. To calculate how many ions we should add to the system to reach the desired bulk salt concentration (0.145 M), we used the SLTCAP¹³³ method^d. We used the OPC water model¹³⁸ and Joung/Cheatham (J/C) ion parameter set¹³⁹ for reference simulations. This solvent model was found earlier to be the

best compromise choice for the reproduction of the experimental Na^+ and K^+ distributions around DNA.¹³⁷ The same ion parameter set¹³⁹ was used in SPC/E simulation. First, we ran 2000 step-long initial minimization using the steepest descent algorithm for the first 1000 steps and the conjugate gradient algorithm for 1001–2000 steps. After the initial minimization, each system was heated from 0 to 300 K in a canonical ensemble (NVT) for 18 ps and then equilibrated for 2 ps in the same ensemble. Then, it was equilibrated for 40.06 ns in isothermal–isobaric ensemble (NPT) using Langevin dynamics with a collision frequency of 2 ps^{-1} to reach 1 atm pressure. Periodic boundary conditions and the particle mesh Ewald method were used. After the equilibration, 300 ns long production trajectories were generated for each system using the NPT ensemble. The integration time-step was 2 fs. During the heating and equilibration, every atom of 25 bp DNA was restrained to their original positions with harmonic restraints with a 5 kcal/mol/Å² force constant. During the production simulation, the constant was 0.01 kcal/mol/Å². Snapshots were saved every 10 ps along each trajectory. We use these simulations to examine the ion distributions around DNA alone, so the DNA was restrained to avoid the influence of DNA conformation change on ion distribution.

4.2.2. DNA–NaCl and DNA–KCl Systems: Implicit Water/Explicit Ions GBION Simulations. A 25 bp DNA duplex with the sequence (GCATCTGGGCTATAAAAGGGCGTCG) was constructed as described above for the explicit water simulations. The DNA was simulated with NaCl and KCl using the OL15 DNA force field. Ion parameters for Na^+ , K^+ , and Cl^- are taken from the Joung/Cheatham (J/C) set for the OPC water model,¹³⁹ and mbondi3 radii set was used; all the parameters of the ions are presented in Table S1. To match the concentrations of NaCl and KCl achieved in the explicit solvent simulations, the same amounts of ions (104 cations and 56 anions) were added to the systems simulated using the GBION model. For the enlarged restraining sphere, with a radius of 120 Å, 458 cations and 410 anions were added. The parameters of the GBION model for the DNA–NaCl simulations are listed below [$K_*(a,b)$ = parameter name as implemented in AMBER = its value, see Supporting Information for implementation details within AMBER]. In what follows, s = solute atoms, c = cation, and a = anion.

1. $K_{\text{GB}}(s, c) = \text{gi_coef_1_p} = 1$
2. $K_{\text{GB}}(s, a) = \text{gi_coef_1_n} = 0.05$
3. $K_{\text{GB}}(c, c) = \text{gi_coef_1_pp} = 1$
4. $K_{\text{GB}}(a, c) = \text{gi_coef_1_pn} = 0.05$
5. $K_{\text{GB}}(a, a) = \text{gi_coef_1_nn} = 1$
6. $K_c(s, c) = \text{intdiel_ion_1_p} = 54$
7. $K_c(s, a) = \text{intdiel_ion_1_n} = 8$
8. $K_c(c, c) = \text{intdiel_ion_1_pp} = 54$
9. $K_c(a, c) = \text{intdiel_ion_1_pn} = 8$
10. $K_c(a, a) = \text{intdiel_ion_1_nn} = 8$

For simulations with KCl, the parameters are the same except $K_c(s, c) = 36$ and $K_c(c, c) = 36$. First, we ran 2000 step-long initial minimization using the steepest descent algorithm for the first 1000 steps and the conjugate gradient algorithm for 1000–2000 steps. After the initial minimization, all of the systems were heated from 0 to 300 K over 20 ps. Each system was equilibrated for 0.4 ns using Langevin dynamics with a collision frequency of 0.05 ps^{-1} . After the equilibration, 10 ns long production trajectories were generated for each system. The integration time-step was 2 fs. Unless otherwise specified,

the nonpolar surface energy term (GB/SA) was not accounted for in the simulations ($\text{gbsa} = 0$). When GB/SA calculations are applied ($\text{gbsa} = 3$), the nonpolar term in eq 1 is proportional to the SASA of the solute, while the GB model is used to calculate the electrostatic term in eq 1. Distributions do not depend on taking into account the surface energy, see Figure S6. During the heating, equilibration, and production simulation, every atom of 25 bp DNA was restrained to their original positions with a 0.01 kcal/mol/Å² force. Snapshots were saved every 1 ps along the production trajectory. To limit the simulation volume, we applied the restraining potential that works as follows. If an ion's position exceeds a specified distance from the attractive center (in this case 67.7 Å from the 401th atom of the DNA duplex, which is atom C4 of the 13th residue, thymine), a harmonic restraining potential with a constant of 20 kcal/mol/Å² begins to act on the ion. For ions closer to the center than that distance, the restraining force is zero. A confining sphere with a radius of 67.7 Å was chosen to match the NaCl and KCl concentrations in the equilibrated truncated octahedron used in the explicit water simulations described above. The robustness of the ion distribution with respect to variations of the restraining sphere's radius was verified, see Figure S7.

The dielectric constants of the solutes and the solvent were chosen as $\epsilon_{\text{in}} = 1$ and $\epsilon_{\text{out}} = 78.5$, respectively. The DNA duplex atomic partial charges and the parameters for van der Waals interactions between the DNA, Na^+ , and Cl^- ions are taken from the AMBER OL15 force field.¹⁴⁶

4.2.2.1. Dickerson–Drew Dodecamer Simulations. To verify that the DNA remains stable during GBION simulations, we simulated Dickerson–Drew dodecamer (CGCGAATTCGCG) for 1 μs . To limit the simulation volume, we applied the restraining potential that works as follows. If an ion exceeds a specified distance of 40 Å from the center of mass of phosphorus atoms of the DNA backbone, harmonic restraining potential with a constant of 20 kcal/mol/Å² begins to act on the ion. For ions closer to the center than that distance, the restraining force is zero. In this case, the center of mass, as opposed to a single atom, was chosen as the reference center point to avoid a significant influence of the restraints on the DNA structure. The number of cations in this case is 36, anions –14. The parameters of minimization, heating, and equilibration simulations were the same as for 25 bp-long DNA duplexes. Production simulation parameters were the same except for the DNA restraints and simulation time. During the 1 μs production simulation, the DNA atoms were not restrained. Snapshots were saved every 100 ps along the production trajectory. The integration time-step was 2 fs.

4.2.3. DNA–CoHex³⁺ System: Explicit Water MD Simulations. The details of the all-atom MD simulations of DNA–CoHex³⁺ systems using the explicit TIP3P water model¹⁴⁷ are described elsewhere³¹ and are similar to the all-atom MD simulation of the DNA–NaCl and DNA–KCl systems described above.

4.2.4. DNA–CoHex³⁺ System: Implicit Water/Explicit Ions GBION Simulations. The construction of the 25 bp homopolymeric (GCATCTGGGCTATAAAAGGGCGTCG) DNA duplex is described above. The duplex (charge –48e) was neutralized by 16 CoHex³⁺ ions. Ion parameters are taken from ref 31. No other ions were included in the simulations. The DNA force field was the same as in the case of the DNA–NaCl and DNA–KCl systems described above (OL15). The parameters of the GBION model for DNA–CoHex³⁺

simulations are listed below [$K_*(a,b)$ = parameter as implemented to AMBER = value]:

1. $K_{GB}(s, c) = gi_coef_1_p = 0.01$
2. $K_{GB}(s, a) = gi_coef_1_n = 0.01$
3. $K_{GB}(c, c) = gi_coef_1_pp = 0.06$
4. $K_{GB}(a, c) = gi_coef_1_pn = 0.06$
5. $K_{GB}(a, a) = gi_coef_1_nn = 0.06$
6. $K_e(s, c) = intdiel_ion_1_p = 1$
7. $K_e(s, a) = intdiel_ion_1_n = 1$
8. $K_e(c, c) = intdiel_ion_1_pp = 2$
9. $K_e(a, c) = intdiel_ion_1_pn = 2$
10. $K_e(a, a) = intdiel_ion_1_nn = 2$

The equilibration step in the simulations of DNA with CoHex³⁺ ions was 500 ps long. During the equilibration, every atom of the DNA was restrained to its initial position with a 1 kcal/mol/Å² harmonic restraining potential. The production simulation was 10 ns long. During the production run, the DNA atoms were also restrained with 0.01 kcal/mol/Å² to match the conditions of the explicit solvent simulation. The integration time-step was 2 fs.

4.2.4.1. Simulation of Ion-Induced DNA Condensation.

To simulate counterion-induced aggregation of a pair of DNA duplexes, we used the same structure as that for the single DNA simulations. Two copies of the system simulated above were initially located at a distance of 74 Å from each other. The minimization and heating simulations were the same as those above. The equilibration step was 50 ps long, and the simulation time step for this equilibration step only was chosen as 0.5 fs to avoid possible numerical instability. The integration time-step was 2 fs in all of the following simulation steps. Production simulation was 300 ns long. During the production simulation, the DNA duplexes were not restrained. To limit the simulation volume, we applied the restraining potential to the ion if the distance from the 413rd atom of one of the duplexes exceeded 160 Å (in other words, the ion-restraining sphere was centered on one of the duplexes).

4.3. Trajectory Analysis and Calculation of Ion Distributions. For analysis of ion distributions around DNA, we have used Curves+ and Canion¹⁴⁸ programs. Curves+ translate coordinates into a convenient curvilinear helicoidal system. Canion analyzes the curvilinear coordinates of ions and generates 1D and 2D distributions and density maps of ions around DNA. To calculate groove widths, rmsd of DNA, and WC values, we used the program *nastruct*, included in the AMBER package.

4.4. Calculating the Number of Ions Associated with the DNA. As an experimental reference, here we utilized the findings of Bai et al., who measured the number of ions associated with DNA using the buffer exchange technique.¹³⁵ Their experiments showed that the number of ions associated with double-stranded DNA remains relatively constant across bulk salt concentrations ranging from 0.1 to 0.2 M. To align with our study, the experimental value for the number of DNA-associated ions was scaled by a factor of 24/23, reflecting the ratio of the total charge between our 25 base-pair DNA fragment and the 24 base-pair fragment used in the experiment. For a direct comparison with the experimental results, we calculated the number of DNA-associated ions as follows¹⁴⁹

$$N = \int_0^r (c(r') - c(\infty)) \times 2\pi hr' dr' \quad (7)$$

where $c(\infty)$ represents the bulk salt concentration, determined as the mean concentration between 28 and 30 Å from the DNA helical axis in this work, and h indicates the length of the DNA duplex, which is 85 Å in our case. The ion concentration as a function of distance from the DNA helical axis, $c(r)$, was computed using the Curves+ and Canion programs.¹⁴⁸

4.5. Generating Schematic Graphs of Electrostatic Energy between Charged Atoms. To illustrate the dependence of function $E_{ij}(d_{ij})$ on its parameters $\epsilon_{in}(a,b)$ and $\gamma(a,b)$, Figure 3b, for a generic pair of positive and negative charges of value ± 1 , we have used the following generic function

$$E_{ij}(d_{ij}) = -\frac{1}{\epsilon_{in}(a, b)d_{ij}} - \left(\frac{1}{\epsilon_{in}(a, b)} - \frac{1}{78.5} \right) \times \frac{1}{\sqrt{d_{ij}^2 + e^{-d_{ij}^2/\gamma(a,b)}}} \quad (8)$$

For Coulomb's law, we used the generic function

$$E_{ij}(d_{ij}) = -\frac{1}{78.5d_{ij}} \quad (9)$$

In eq 8, the Born self-terms were assumed constant (set to zero for simplicity), and the effective Born radii were set to 1. This, by no means general, approximation reflects the fact that the effective Born radii for the pair of small ions change relatively little over the relatively small range of d_{ij} of interest to us here. Also, our focus is on the general behavior of E_{ij} , eq 3, which is not affected by the approximation. The specific coefficients used as examples in that figure were chosen to generate distinguishable graphs.

4.6. Software Used. We used the AMBER¹⁴⁶ program package for MD simulations. The GBION model used here is available in AMBER 24, released in April of 2024. For visualization of structures and trajectories, we used PyMol¹⁵⁰ and CHIMERA.¹⁵¹ We used Curves+, Canion,¹⁴⁸ and Python3 libraries NumPy¹⁵² and Matplotlib¹⁵³ for ion distribution analysis and graph visualization, respectively. We used ChatGPT 4¹⁵⁴ for style editing.

5. CONCLUSIONS

The ionic atmosphere around a biomolecule can have a significant effect on its structure, dynamics, and function, necessitating accurate yet efficient representation of ions in simulation. Existing solvent models face various limitations with respect to representing ions in atomistic simulations: explicit solvent models can be computationally expensive, while the “engine” of the more efficient implicit solvent MD simulations, the GB model, provides only a crude, mean field approximation for the ionic atmosphere.

In this study, we have made substantial modifications to the framework of the canonical GB approximation to explicitly take into account ions around the solute molecule. The key insight is to consider distinct functional forms for the solute–solute, solute–ion, and ion–ion interactions.

The resulting “Implicit Solvent with EXplicit Ions” framework was implemented in AMBER (version 24, released on 04/30/2024, or later version) as a GBION model and evaluated through extensive MD simulations of DNA duplexes surrounded by either monovalent (Na⁺, K⁺, and Cl[−]) or trivalent (cobalt hexamine, CoHex³⁺) ions. The computed

ion distributions obtained from GBION simulations showed reasonable agreement with those estimated by traditional all-atom MD simulations in explicit water as well as in experiment, where available. The largest deviation between GBION and the explicit solvent distribution was about $\sim k_B T$ (in energy units), comparable to relevant differences between ion distributions obtained using different established explicit water models.

The GBION model accurately captures key features of CoHex³⁺ distribution around double-stranded DNA, including the preferential ion binding to the “outer surface” of the B-DNA backbone and the $\sim 90\%$ degree of DNA charge neutralization. In contrast to the canonical GB, the new approach can simulate the process of DNA condensation induced by multivalent counterions. Not only do these simulations yield the well-known “lateral” DNA condensation mode, but they also reveal an unexpected “stacking” condensation of DNA duplexes, supported by recent experimental evidence.³³ The efficiency of GBION in simulating large solvent volumes without severely compromising computation efficiency enables the observation of the transition from the lateral to the “stacking” condensed mode, which would have been highly challenging to simulate in the traditional explicit solvent.

Below is a brief summary of the key advantages and limitations of GBION.

5.1. Advantages vs Canonical GB. In contrast to the canonical GB, the GBION model allows simulating ions explicitly; this level of detail is desirable for many, if not most, biologically relevant systems on time scales that are becoming routine today. In particular, the canonical GB does not distinguish between Na⁺ and K⁺, but GBION does. The model can also handle multivalent ions naturally, which is completely out of reach to the canonical GB.

5.2. Advantages vs Traditional Fully Explicit Solvent. The GBION model inherits the key advantage of the canonical GB model relative to the explicit solvent: the speed of conformational sampling. Specifically, GBION simulations sample conformational space 1–2 orders of magnitude faster than the corresponding explicit solvent simulations.

Another notable advantage of the GBION model over the traditional explicit solvation is that the GBION computational cost is not very sensitive to the size of the simulation volume. This unique feature of the approach allows one to simulate physiologically relevant, very low salt concentrations of some ions without the exorbitant computational expense that such simulations would incur in the explicit solvent due to the need for a very large solvent box.

5.3. Limitations. With the main focus of this work being conceptual development of the new model, the set of molecular systems used for the current parameterization and testing is limited. Specifically, we optimized parameters for Na⁺, K⁺, Cl[−], and CoHex³⁺ to work with the DNA. The current parameters of the model will likely work reasonably well for other negatively charged solutes, but we have not verified that yet. We have not yet tested the ability of GBION to reproduce ion distributions around proteins; positively charged solutes may be particularly challenging to the current parameterization of Cl[−]. Therefore, we advise against using the current set of GBION parameters for protein simulations, especially those that carry a net positive charge. Also, biologically relevant divalent ions such as Mg²⁺ or Ca²⁺ have not yet been parameterized.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00833>.

Comparison of sodium distributions obtained using the SPC/E water model with those presented in the text; capped radial-angle distributions of Na⁺ and K⁺; comparison of Cl[−] distributions and dependence of cation distributions on the simulation times, taking into account solvent accessible surface energy and the radius of the restraining sphere; details of the model implementation in the AMBER package; and parameters of monovalent ions used in the simulations (PDF)

Movie of 225 ps-long Dickerson–Drew dodecamer simulation with KCl (MP4)

Movie of 100 ns-long simulation of DNA condensation in the presence of CoHex³⁺ (MP4)

■ AUTHOR INFORMATION

Corresponding Author

Alexey V. Onufriev — *Departments of Computer Science and Physics, Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia 24061, United States;*
✉ orcid.org/0000-0002-4930-6612; Email: alexey@cs.vt.edu

Authors

Egor S. Kolesnikov — *Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, United States*

Yeyue Xiong — *Department of Biomedical Engineering and Mechanics, Virginia Tech, Blacksburg, Virginia 24061, United States;* ✉ orcid.org/0000-0003-0348-1009

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jctc.4c00833>

Author Contributions

[†]E.S.K. and Y.X. made equal contribution.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) R01 GM144596. The authors also acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support.

■ ADDITIONAL NOTES

^aProof-of-concept MD simulations based on the numerical PB were reported,^{43,80–82} including an explicit ion model.⁸³ It is believed that the inherent algorithmic complexity of the numerical PB approach, conceptual difficulties associated with incorporating PB-based algorithms into MD,⁸⁴ and significant computational expense⁸⁵ preclude its wider adoption in atomistic MD.

^bPhysically, the interaction E_{ij} between solvent-separated charges must be zero in the conductor limit of $\epsilon_{\text{out}} \rightarrow \infty$, but this is not the case for the canonical GB. In contrast, the pure Coulomb interaction has the correct limiting behavior.

^cEvery dimension of the explicit solvent box needs to be at least 180 Å, which translates into $\sim 2 \times 10^5$ water molecules or $\sim 7 \times 10^5$ atoms. According to recent AMBER benchmarks, the

simulation speed of a system containing $\sim 4 \times 10^5$ atoms is ~ 45 ns/day (NVIDIA RTX 3080, NPT ensemble, 2 fs step). Extrapolating to $\sim 7 \times 10^5$ atoms yields ~ 25 ns/day. Consequently, a 4 μ s-long simulation of the system containing $\sim 7 \times 10^5$ atoms should take at least ~ 0.5 years.

^dThe actual bulk salt concentration inferred from an analysis of the ion distributions (Figure 5_i) is ~ 0.1 M, which is slightly less than 0.145 M targeted by SLTCAP. However, since the difference with the target is small and 0.1 M can still be considered physiological, we did not apply further corrections.

REFERENCES

- (1) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. Ionic effects on the elasticity of single DNA molecules. In *Proceedings of the National Academy of Sciences*, 1997; pp 6185–6190.
- (2) Wenner, J. R.; Williams, M. C.; Rouzina, I.; Bloomfield, V. A. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophys. J.* **2002**, *82*, 3160–3169.
- (3) Savelyev, A.; Materese, C. K.; Papoian, G. A. Is DNA's Rigidity Dominated by Electrostatic or Nonelectrostatic Interactions? *J. Am. Chem. Soc.* **2011**, *133*, 19290–19293.
- (4) Savelyev, A.; MacKerell, A. D. Differential Deformability of the DNA Minor Groove and Altered BI/BII Backbone Conformational Equilibrium by the Monovalent Ions Li⁺, Na⁺, K⁺, and Rb⁺ via Water-Mediated Hydrogen Bonding. *J. Chem. Theory Comput.* **2015**, *11*, 4473–4485.
- (5) Drozdetski, A. V.; Tolokh, I. S.; Pollack, L.; Baker, N.; Onufriev, A. V. Opposing Effects of Multivalent Ions on the Flexibility of DNA and RNA. *Phys. Rev. Lett.* **2016**, *117*, 028101.
- (6) Woodson, S. A. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* **2005**, *9*, 104–109.
- (7) Hayes, R. L.; Noel, J. K.; Whitford, P. C.; Mohanty, U.; Sanbonmatsu, K. Y.; Onuchic, J. N. Reduced Model Captures Mg²⁺-RNA Interaction Free Energy of Riboswitches. *Biophys. J.* **2014**, *106*, 1508–1519.
- (8) Fischer, N. M.; Polêto, M. D.; Steuer, J.; van der Spoel, D. Influence of Na⁺ and Mg²⁺ ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res.* **2018**, *46*, 4872–4882.
- (9) Bloomfield, V. A. Condensation of DNA by multivalent cations: Considerations on mechanism. *Biopolymers* **1991**, *31*, 1471–1481.
- (10) Bloomfield, V. A. DNA condensation by multivalent cations. *Biopolymers* **1997**, *44*, 269–282.
- (11) Morgan, J. E.; Blankenship, J. W.; Matthews, H. R. Polyamines and acetyl polyamines increase the stability and alter the conformation of nucleosome core particles. *Biochemistry* **1987**, *26*, 3643–3649.
- (12) Belyi, V. A.; Muthukumar, M. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17174–17178.
- (13) Hassett, K. J.; Benenato, K. E.; Jacquinet, E.; Lee, A.; Woods, A.; Yuzhakov, O.; Himansu, S.; Deterling, J.; Geilich, B. M.; Ketova, T.; Mihai, C.; Lynn, A.; McFadyen, I.; Moore, M. J.; Senn, J. J.; Stanton, M. G.; Almarsson, O.; Ciarabella, G.; Brito, L. A. Optimization of Lipid Nanoparticles for Intramuscular Administration of mRNA Vaccines. *Mol. Ther.—Nucleic Acids* **2019**, *15*, 1–11.
- (14) Sahin, U.; Muik, A.; Derhovanessian, E.; Vogler, I.; Kranz, L. M.; Vormehr, M.; Baum, A.; Pascal, K.; Quandt, J.; Maurus, D.; Brachtendorf, S.; Lörks, V.; Sikorski, J.; Hilker, R.; Becker, D.; Eller, A.-K.; Grützner, J.; Boesler, C.; Rosenbaum, C.; Kühnle, M. C.; Luxemburger, U.; Kemmer-Brück, A.; Langer, D.; Bexon, M.; Bolte, S.; Karikó, K.; Palanche, T.; Fischer, B.; Schultz, A.; Shi, P.-Y.; Fontes-Garfias, C.; Perez, J. L.; Swanson, K. A.; Loschko, J.; Scully, I. L.; Cutler, M.; Kalina, W.; Kyratsous, C. A.; Cooper, D.; Dormitzer, P. R.; Jansen, K. U.; Türeci, Ö. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. *Nature* **2020**, *586*, 594–599.
- (15) Verma, I. M.; Somia, N. Gene therapy - promises, problems and prospects. *Nature* **1997**, *389*, 239–242.
- (16) Anderson, W. F. Human gene therapy. *Nature* **1998**, *392*, 25–30.
- (17) Agrawal, N.; Dasaradhi, P. V. N.; Mohammed, A.; Malhotra, P.; Bhatnagar, R. K.; Mukherjee, S. K. RNA Interference: Biology, Mechanism, and Applications. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 657–685.
- (18) Mansoori, B.; Sandoghchian Shotorbani, S.; Baradaran, B. RNA Interference and its Role in Cancer Therapy. *Adv. Pharm. Bull.* **2014**, *4*, 313–321.
- (19) Baldwin, R. How Hofmeister ion interactions affect protein stability. *Biophys. J.* **1996**, *71*, 2056–2063.
- (20) Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of Ion-Protein Interactions Using the Drude Polarizable Force-Field. *J. Phys. Chem. B* **2015**, *119*, 9401–9416.
- (21) Rutherford, J. C.; Bird, A. J. Metal-Responsive Transcription Factors That Regulate Iron, Zinc, and Copper Homeostasis in Eukaryotic Cells. *Eukaryotic Cell* **2004**, *3*, 1–13.
- (22) Schreiter, E. R.; Drennan, C. L. Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* **2007**, *5*, 710–720.
- (23) Ashraf, K. U.; Josts, I.; Mosbahi, K.; Kelly, S. M.; Byron, O.; Smith, B. O.; Walker, D. The Potassium Binding Protein Kbp Is a Cytoplasmic Potassium Sensor. *Structure* **2016**, *24*, 741–749.
- (24) Chin, D.; Means, A. R. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* **2000**, *10*, 322–328.
- (25) Handel, T. M.; Williams, S. A.; DeGrado, W. F. Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **1993**, *261*, 879–885.
- (26) Al Qaraghuli, M. M.; Kubiak-Ossowska, K.; Mulheran, P. A. Thinking outside the Laboratory: Analyses of Antibody Structure and Dynamics within Different Solvent Environments in Molecular Dynamics (MD) Simulations. *Antibodies* **2018**, *7*, 21–34.
- (27) Adcock, S.; McCammon, J. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106*, 1589–1615.
- (28) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.
- (29) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (30) Raman, E. P.; MacKerell, A. D. Spatial Analysis and Quantification of the Thermodynamic Driving Forces in Protein-Ligand Binding: Binding Site Variability. *J. Am. Chem. Soc.* **2015**, *137*, 2608–2621.
- (31) Tolokh, I. S.; Pabit, S. A.; Katz, A. M.; Chen, Y.; Drozdetski, A.; Baker, N.; Pollack, L.; Onufriev, A. V. Why double-stranded RNA resists condensation. *Nucleic Acids Res.* **2014**, *42*, 10823–10831.
- (32) Katz, A. M.; Tolokh, I. S.; Pabit, S. A.; Baker, N.; Onufriev, A. V.; Pollack, L. Spermine Condenses DNA, but Not RNA Duplexes. *Biophys. J.* **2017**, *112*, 22–30.
- (33) Kodikara, S. G.; Gyawali, P.; Gleeson, J. T.; Jakli, A.; Sprunt, S.; Balci, H. Stability of End-to-End Base Stacking Interactions in Highly Concentrated DNA Solutions. *Langmuir* **2023**, *39*, 4838–4846.
- (34) Anandakrishnan, R.; Drozdetski, A.; Walker, R.; Onufriev, A.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108*, 1153–1164.
- (35) Romani, A. M. P. Magnesium homeostasis in mammalian cells. *Front. Biosci.* **2007**, *12*, 308.
- (36) Bagur, R.; Hajnóczky, G. Intracellular Ca²⁺ Sensing: Its Role in Calcium Homeostasis and Signaling. *Mol. Cell* **2017**, *66*, 780–788.
- (37) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (38) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144–1149.
- (39) Beroza, P.; Case, D. A. Calculations of proton-binding thermodynamics in proteins. *Methods Enzymol.* **1998**, *295*, 170–189.

- (40) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. Biological Applications of Electrostatic Calculations and Brownian Dynamics. *Rev. Comput. Chem.* **1994**, *5*, 229–267.
- (41) Gilson, M. K. Theory of Electrostatic Interactions in Macromolecules. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216–223.
- (42) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098–8106.
- (43) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann Calculations for Static and Dynamic Systems. *Journal of computational chemistry* **2002**, *23*, 1244–1253.
- (44) Simonson, T. Electrostatics and Dynamics of Proteins. *Rep. Prog. Phys.* **2003**, *66*, 737–787.
- (45) Baker, N.; Bashford, D.; Case, D. Implicit Solvent Electrostatics in Biomolecular Simulation. *New Algorithms for Macromolecular Simulation* **2006**, *49*, 263–295.
- (46) Bardhan, J. P. Biomolecular electrostatics – I want your solvation (model). *Comput. Sci. Discovery* **2012**, *5*, 013001.
- (47) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.
- (48) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691–1741.
- (49) Hoijtink, G.; De Boer, E.; Van der Meij, P.; Weijland, W. Reduction Potentials of Various Aromatic Hydrocarbons and Their Univalent Anions. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487–503.
- (50) Tucker, S. C.; Truhlar, D. G. Generalized Born Fragment Charge Model for Solvation Effects as a Function of Reaction Coordinate. *Chem. Phys. Lett.* **1989**, *157*, 164–170.
- (51) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
- (52) Jayaram, B.; Liu, Y.; Beveridge, D. L. A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J. Chem. Phys.* **1998**, *109*, 1465–1471.
- (53) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (54) Onufriev, A.; Bashford, D.; Case, D. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (55) Dominy, B. N.; Brooks, C. L. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.
- (56) Bashford, D.; Case, D. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (57) Calimet, N.; Schaefer, M.; Simonson, T. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 144–158.
- (58) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (59) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
- (60) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
- (61) Feig, M.; Im, W.; Brooks, C. L. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* **2004**, *120*, 903–911.
- (62) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (63) Lee, M. S.; Salsbury, F. R., Jr; Brooks, C. L. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- (64) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- (65) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* **2004**, *55*, 383–394.
- (66) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.
- (67) Tsui, V.; Case, D. Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.
- (68) Wang, T.; Wade, R. Implicit solvent models for flexible protein-protein docking by molecular dynamics simulation. *Proteins* **2003**, *50*, 158–169.
- (69) Gallicchio, E.; Levy, R. M. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **2004**, *25*, 479–499.
- (70) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934.
- (71) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.
- (72) Im, W.; Lee, M. S.; Brooks, C. L. Generalized born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
- (73) Haberthür, U.; Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* **2008**, *29*, 701–715.
- (74) Grant, J. A.; Pickup, B. T.; Sykes, M. J.; Kitchen, C. A.; Nicholls, A. The Gaussian Generalized Born model: application to small molecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913.
- (75) Tjong, H.; Zhou, H. X. GBr6: A Parameterization-Free, Accurate, Analytical Generalized Born Method. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.
- (76) Labute, P. The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* **2008**, *29*, 1693–1698.
- (77) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.
- (78) Onufriev, A. V.; Sigalov, G. A strategy for reducing gross errors in the generalized Born models of implicit solvation. *J. Chem. Phys.* **2011**, *134*, 164104.
- (79) Lange, A. W.; Herbert, J. M. Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator. *J. Chem. Theory Comput.* **2012**, *8*, 1999–2011.
- (80) Totrov, M.; Abagyan, R. Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide. *Biopolymers* **2001**, *60*, 124–133.
- (81) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *Journal of computational chemistry* **2001**, *22*, 608–640.
- (82) Prabhu, N. V.; Zhu, P.; Sharp, K. A. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *Journal of computational chemistry* **2004**, *25*, 2049–2064.
- (83) Prabhu, N. V.; Panda, M.; Yang, Q.; Sharp, K. A. Explicit ion, implicit water solvation for molecular dynamics of nucleic acids and

highly charged molecules. *Journal of computational chemistry* **2008**, *29*, 1113–1130.

(84) Xiao, L.; Wang, C.; Luo, R. Recent progress in adapting Poisson-Boltzmann methods to molecular simulations. *J. Theor. Comput. Chem.* **2014**, *13*, 1430001.

(85) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. Performance Comparison of Generalized Born and Poisson Methods in the Calculation of Electrostatic Solvation Energies for Protein Structures. *J. Comput. Chem.* **2004**, *25*, 265–284.

(86) Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238–250.

(87) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.

(88) Chen, J.; Im, W.; Brooks, C. L. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(89) Jang, S.; Kim, E.; Pak, Y. All-atom level direct folding simulation of a $\beta\beta\alpha$ miniprotein. *J. Chem. Phys.* **2008**, *128*, 105102.

(90) Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. Simulation of Folding of a Small Alpha-helical Protein in Atomistic Detail using Worldwide-distributed Computing. *J. Mol. Biol.* **2002**, *323*, 927–937.

(91) Jang, S.; Kim, E.; Shin, S.; Pak, Y. Ab Initio Folding of Helix Bundle Proteins Using Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2003**, *125*, 14841–14846.

(92) Lei, H.; Duan, Y. Ab Initio Folding of Albumin Binding Domain from All-Atom Molecular Dynamics Simulation. *J. Phys. Chem. B* **2007**, *111*, 5458–5463.

(93) Pitera, J. W.; Swope, W. Understanding folding and design: Replica-exchange simulations of Trp-cage² miniproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.

(94) Jagielska, A.; Scheraga, H. A. Influence of temperature, friction, and random forces on folding of the B-domain of staphylococcal protein A: All-atom molecular dynamics in implicit solvent. *J. Comput. Chem.* **2007**, *28*, 1068–1082.

(95) Lopes, A.; Alexandrov, A.; Bathelt, C.; Archontis, G.; Simonson, T. Computational sidechain placement and protein mutagenesis with implicit solvent models. *Proteins* **2007**, *67*, 853–867.

(96) Felts, A. K.; Gallicchio, E.; Chekmarev, D.; Paris, K. A.; Friesner, R. A.; Levy, R. M. Prediction of Protein Loop Conformations Using the AGBNP Implicit Solvent Model and Torsion Angle Sampling. *J. Chem. Theory Comput.* **2008**, *4*, 855–868.

(97) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915–920.

(98) Amaro, R. E.; Cheng, X.; Ivanov, I.; Xu, D.; McCammon, J. A. Characterizing Loop Dynamics and Ligand Recognition in Human and Avian-Type Influenza Neuraminidases via Generalized Born Molecular Dynamics and End-Point Free Energy Calculations. *J. Am. Chem. Soc.* **2009**, *131*, 4702–4709.

(99) Ruscio, J. Z.; Onufriev, A. A computational study of nucleosomal DNA flexibility. *Biophys. J.* **2006**, *91*, 4121–4132.

(100) Tsui, V.; Case, D. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56*, 275–291.

(101) Sorin, E.; Rhee, Y.; Nakatani, B.; Pande, V. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophys. J.* **2003**, *85*, 790–803.

(102) De Castro, L. F.; Zacharias, M. DAPI binding to the DNA minor groove: a continuum solvent analysis. *J. Mol. Recognit.* **2002**, *15*, 209–220.

(103) Allawi, H.; Kaiser, M.; Onufriev, A.; Ma, W.; Brogaard, A.; Case, D.; Neri, B.; Lyamichev, V. Modeling of flap endonuclease interactions with DNA substrate. *J. Mol. Biol.* **2003**, *328*, 537–554.

(104) Chocholoušová, J.; Feig, M. Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment. *J. Phys. Chem. B* **2006**, *110*, 17240–17251.

(105) Spassov, V. Z.; Yan, L.; Szalma, S. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B* **2002**, *106*, 8726–8738.

(106) Im, W.; Feig, M.; Brooks, C. L. An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* **2003**, *85*, 2900–2918.

(107) Tanizaki, S.; Feig, M. A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J. Chem. Phys.* **2005**, *122*, 124706.

(108) Ulmschneider, M. B.; Ulmschneider, J. P.; Sansom, M. S.; Di Nola, A. A generalized Born implicit-membrane representation compared to experimental insertion free energies. *Biophys. J.* **2007**, *92*, 2338–2349.

(109) Zheng, W.; Spassov, V.; Yan, L.; Flook, P.; Szalma, S. A hidden Markov model with molecular mechanics energy-scoring function for transmembrane helix prediction. *Comput. Biol. Chem.* **2004**, *28*, 265–274.

(110) Zhu, J.; Alexov, E.; Honig, B. Comparative study of generalized Born models: Born radii and peptide folding. *J. Phys. Chem. B* **2005**, *109*, 3008–3022.

(111) Liu, H. Y.; Zou, X. Electrostatics of Ligand Binding: Parametrization of the Generalized Born Model and Comparison with the Poisson-Boltzmann Approach. *J. Phys. Chem. B* **2006**, *110*, 9304–9313.

(112) Wittayanarakul, K.; Hannongbua, S.; Feig, M. Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *J. Comput. Chem.* **2008**, *29*, 673–685.

(113) Islam, M. A.; Bhayye, S.; Adeniyi, A. A.; Soliman, M. E. S.; Pillay, T. S. Diabetes mellitus caused by mutations in human insulin: analysis of impaired receptor binding of insulins Wakayama, Los Angeles and Chicago using pharmacoinformatics. *J. Biomol. Struct. Dyn.* **2017**, *35*, 724–737.

(114) Zou, X.; Yaxiong, S.; Kuntz, I. D. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.

(115) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.

(116) Pellegrini, E.; Field, M. J. A Generalized-Born Solvation Model for Macromolecular Hybrid-Potential Calculations. *J. Phys. Chem. A* **2002**, *106*, 1316–1326.

(117) Walker, R. C.; Crowley, M. F.; Case, D. A. The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.* **2008**, *29*, 1019–1031.

(118) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.

(119) Onufriev, A. In *Modeling Solvent Environments*; Feig, M., Ed.; Wiley: USA, 2010; pp 127–165.

(120) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.

(121) Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *J. Am. Chem. Soc.* **2014**, *136*, 13959–13962.

(122) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(123) AMBER GPU BENCHMARKS. <https://ambermd.org/GPUPerformance.php>.

(124) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Implicit Solvent Model for Million-Atom Atomistic Simulations: Insights into the Organization of 30-nm Chromatin Fiber. *J. Chem. Theory Comput.* **2016**, *12*, 5946–5959.

- (125) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (126) Kolesnikov, E. S.; Gushchin, I. Y.; Zhilyaev, P. A.; Onufriev, A. V. Why Na^+ has higher propensity than K^+ to condense DNA in a crowded environment. *J. Chem. Phys.* **2023**, *159*, 145103.
- (127) Collins, K. D. Why continuum electrostatics theories cannot explain biological structure, polyelectrolytes or ionic strength effects in ion–protein interactions. *Biophys. Chem.* **2012**, *167*, 43–59.
- (128) Tolokh, I. S.; Thomas, D. G.; Onufriev, A. V. Explicit ions/implicit water generalized Born model for nucleic acids. *J. Chem. Phys.* **2018**, *148*, 195101.
- (129) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1–20.
- (130) Grycuk, T. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. *J. Chem. Phys.* **2003**, *119*, 4817–4826.
- (131) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *J. Chem. Phys.* **2006**, *124*, 124902.
- (132) Sigalov, G.; Scheffel, P.; Onufriev, A. Incorporating Variable Dielectric Environments into the Generalized Born Model. *J. Chem. Phys.* **2005**, *122*, 094511.
- (133) Schmit, J. D.; Kariyawasam, N. L.; Needham, V.; Smith, P. E. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *J. Chem. Theory Comput.* **2018**, *14*, 1823–1827.
- (134) Widom, J.; Baldwin, R. L. Cation-induced toroidal condensation of DNA. *J. Mol. Biol.* **1980**, *144*, 431–453.
- (135) Bai, Y.; Greenfeld, M.; Travers, K. J.; Chu, V. B.; Lipfert, J.; Doniach, S.; Herschlag, D. Quantitative and Comprehensive Decomposition of the Ion Atmosphere around Nucleic Acids. *J. Am. Chem. Soc.* **2007**, *129*, 14981–14988.
- (136) Manning, G. S. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions II. Self-Diffusion of the Small Ions. *J. Chem. Phys.* **1969**, *51*, 934–938.
- (137) Kolesnikov, E. S.; Gushchin, I. Y.; Zhilyaev, P. A.; Onufriev, A. V. Similarities and Differences between Na^+ and K^+ Distributions around DNA Obtained with Three Popular Water Models. *J. Chem. Theory Comput.* **2021**, *17*, 7246–7259.
- (138) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
- (139) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (140) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179–2183.
- (141) Sushko, M. L.; Thomas, D. G.; Pabit, S. A.; Pollack, L.; Onufriev, A. V.; Baker, N. A. The role of correlation and solvation in ion interactions with B-DNA. *Biophys. J.* **2016**, *110*, 315–326.
- (142) Wu, Y.-Y.; Zhang, Z.-L.; Zhang, J.-S.; Zhu, X.-L.; Tan, Z.-J. Multivalent Ion-Mediated Nucleic Acid Helix-Helix Interactions: RNA versus DNA. *Nucleic Acids Res.* **2015**, *43*, 6156–6165.
- (143) Tolokh, I. S.; Drozdetski, A. V.; Pollack, L.; Baker, N.; Onufriev, A. V. Multi-shell model of ion-induced nucleic acid condensation. *J. Chem. Phys.* **2016**, *144*, 155101.
- (144) Pollack, L. SAXS Studies of Ion–Nucleic Acid Interactions. *Annu. Rev. Biophys.* **2011**, *40*, 225–242.
- (145) Macke, T.; Case, D. *Molecular Modeling of Nucleic Acids*. Chapter 24; American Chemical Society: WA, 1997; pp 379–393.
- (146) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (147) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (148) Pasi, M.; Maddocks, J. H.; Lavery, R. Analyzing ion distributions around DNA. *Nucleic Acids Res.* **2015**, *42*, 2412.
- (149) Yoo, J.; Aksimentiev, A. Competitive Binding of Cations to Duplex DNA Revealed through Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116*, 12946–12954.
- (150) Schrödinger, L.; DeLano, W. *PyMOL*, 2020. <http://www.pymol.org/pymol>.
- (151) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (152) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.
- (153) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (154) *OpenAI, ChatGPT*. Mar 14 2023 version, 2023; <https://www.openai.com/chatgpt>.



CAS BIOFINDER DISCOVERY PLATFORM™

**PRECISION DATA
FOR FASTER
DRUG
DISCOVERY**

CAS BioFinder helps you identify
targets, biomarkers, and pathways

Unlock insights

CAS
A division of the
American Chemical Society