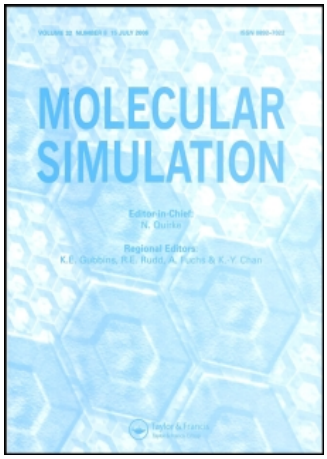

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

ISIM: A Program for Grand Canonical Monte Carlo Simulations of the Ionic Environment of Biomolecules

Andreas Vitalis [a]; Nathan A. Baker [b]; J. Andrew McCammon‡ [c]
[a] Department of Chemistry and Biochemistry, University of California at San Diego, San Diego, CA, USA
[b] Department of Biochemistry and Molecular Biophysics, Center for Computational Biology, Washington University School of Medicine, St. Louis, MO, USA
[c] Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, Department of Pharmacology, University of California at San Diego.Gilman Dr., San Diego, CA, USA

Online Publication Date: 01 January 2004
To cite this Article: Vitalis, Andreas, Baker, Nathan A. and McCammon‡, J. Andrew (2004) 'ISIM: A Program for Grand Canonical Monte Carlo Simulations of the Ionic Environment of Biomolecules', Molecular Simulation, 30:1, 45 - 61
To link to this article: DOI: 10.1080/08927020310001597862
URL: http://dx.doi.org/10.1080/08927020310001597862

Molecular Simulation, VOL. 30 (1), 15 January 2004, pp. 45–61

Taylor & Francis
Taylor & Francis Group

ISIM: A Program for Grand Canonical Monte Carlo Simulations of the Ionic Environment of Biomolecules

ANDREAS VITALIS[a,*], NATHAN A. BAKER[b,†] and J. ANDREW MCCAMMON[c,‡]

[a]*Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Dr., Mail Code 0365, San Diego, CA 92093-0365, USA;* [b]*Department of Biochemistry and Molecular Biophysics, Center for Computational Biology, Washington University School of Medicine, 700 S. Euclid Avenue, St. Louis, MO 63110, USA;* [c]*Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, Department of Pharmacology, University of California at San Diego. 9500 Gilman Dr., Mail Code 0365, San Diego, CA 92093-0365, USA*

(Received September 2002; In final form October 2002)

In this work we present a new software package (ISIM), which represents a flexible, computational tool for simulations of electrolyte solutions via a grand canonical Monte Carlo procedure (GCMC) with a specific capability of treating biomolecules in solution. The GCMC method provides a powerful tool for studying the ionic environments of highly charged macromolecules with attention to the atomic detail of both the solute and the mobile counterions. The ISIM software differs from previous schemes mainly by treating different ion types independently and offering a new parameterization procedure for calibrating excess chemical potentials and bulk ion concentrations. Additionally, ISIM leverages the APBS software package to provide accurate descriptions of the biomolecular electrostatic potential through the efficient solution of Poisson's equation. ISIM has been validated on a variety of test systems; we successfully reproduce elementary properties of electrolyte solutions as well as theoretical and experimental results for challenging test systems like Calmodulin and DNA.

Keywords: Grand canonical Monte Carlo; Ion simulation; ISIM software; Electrostatics

INTRODUCTION

The grand canonical ensemble (GCE) has proven to be an appropriate description of the ionic species in electrolyte solutions [1] when treating them within the primitive model [2,3], which employs a continuum description of the solvent as a dielectric medium (McMillan-Mayer level [4,5]) and a pairwise additive potential of mean force (PMF) for ion–ion interactions. The GCE is a μVT ensemble, which includes configurations with fixed chemical potential (μ), volume (V), and temperature (T), but allows particle numbers (N) to fluctuate [6].

Monte Carlo (MC) methods use probability distributions derived from statistical mechanics to generate members of a particular ensemble [7]. The most common of these methods is Metropolis MC [8], which generates members of the canonical ensemble (NVT) by randomly displacing particles with specific probabilities to obtain new, statistically relevant configurations. Since GCEs contain systems with differences in both particle configuration and number, grand canonical MC (GCMC) methods supplement Metropolis configuration space sampling with random particle insertions and deletions [9,10]. GCMC procedures have been employed in a wide range of technical fashions for example by taking advantage of parallelization [11,12] or by coupling them to force-based methods like molecular dynamics (MD) [13,14] or Brownian dynamics (BD) [15] and for a variety of applications, including the simulation of electrolyte solutions [1,16], prediction of water positions in crystal structures [17,18], adsorption processes [19], systems dominated by quantum effects [20], dynamical treatment of lipid bilayers [21] and finally biomolecular ion atmospheres. These latter simulations have typically focused on DNA [22,23] and rarely included other macromolecular polyions

*E-mail:avitalis@mccammon.ucsd.edu
†Corresponding author. E-mail:baker@biochem.wustl.edu
‡E-mail:jmccammon@ucsd.edu

ISSN 0892-7022 print/ISSN 1029-0435 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/08927020310001597862

(one exception is the treatment of an ion channel in Ref. [15]). The effectiveness of GCMC methods typically depends on the particle densities [1,24] which has limited its applicability in explicit solvent models. However the so-called cavity-biased scheme was developed and used to improve the efficiency for such systems significantly [24,25].

Every GCMC simulation has to deal with the fundamental issues of obtaining and maintaining an average target concentration (convergence) and, in case of charged particles, maintaining an average net charge of zero (electroneutrality). The latter condition is usually introduced as a fundamental constraint: the simulation is started with a neutral solution and only net neutral combinations of particles are inserted or deleted. However, except for its purpose in maintaining strict electroneutrality (especially in the case of periodic simulations), there is no justification for this restraint though. In fact, this restriction can be unrealistic; for example, the simulation of divalent cations and a univalent cationic coions around a charged macromolecule imposes the necessity to either have a huge bulk reservoir of the electrolyte mixture to simulate the right average composition or to dynamically adjust the composition of the neutral combination to be inserted or deleted, for which no *a priori* rule can be thought of. One notable exception to those strict electroneutrality constraints is the methodology developed by Im *et al.* [15], whereby all particle types are treated independently yet the correct average concentration and neutrality conditions are maintained. Im and co-workers handle the concentration problem elegantly by successfully using results obtained with the hypernetted chain equation (HNC) theory instead of the "classical" parameterization.

However, the independent treatment of ionic species poses two (related) concerns. The first one is the necessary splitting rule for the excess chemical potential of solvation, which is frequently used as a calibration parameter in earlier works by associating specific values with the concentrations yielded by the corresponding simulations all within the restriction of pairwise insertion and deletion. There is no theoretically rigorous algorithm to break up the electrolyte's excess chemical potential into its single ion contributions, although these quantities can be directly obtained from integral equation theories [15]. The second concern is the success of such an algorithm for maintaining the global electroneutrality for an arbitrarily chosen simulation system. One part of the solution of the electroneutrality problem is to work within a true "droplet" model instead of applying periodic boundary conditions, so that net charge fluctuations can easily be tolerated, as long as the average value is strictly zero (in the presence of a highly charged macroion this necessity holds only for sufficiently large "droplets").

When working within the primitive model one has to be concerned about its limitations. Commonly applied interaction potentials for ions in solution are spherically symmetric and consist of an electrostatic part and a hard- or soft-core (Lennard-Jones) contribution which accounts for steric interactions [6]. Electrostatic interactions are typically obtained from solutions to Poisson's equation, the canonical equation for electrostatic interactions in a dielectric continuum [26]. Clearly, these simple interaction potentials do not capture the full range of influences on the ion. Specifically, the primitive model typically neglects molecular effects due to the solvent, polarization effects, angular dependencies, and many-body contributions [2,3]. However, despite these approximations, previous work has shown that it is capable of reproducing fundamental properties of electrolyte solutions [27–29]. The previously mentioned applications showed both the sufficiency of the model [15,30] and its limitations, which were therefore addressed by variations [22,30]. However as we discuss later, some of the errors might have to be assigned to the restrictions (as described above) in the GCMC scheme applied in most of these publications. Apart from this ongoing discussion, the conclusion might be that the use of the primitive model is almost certainly the most serious approximation made in the course of the whole methodology described here and must, therefore, be subject to close and continuous inspection.

Ions play an important role in many aspects of biology, including (for example) the structure, folding, and function of numerous proteins such as hemoglobin, cytochromes or zinc-finger proteins; the regulation of the transmembrane potential by the interplay of various transport mechanisms; or the multiple events in cellular signaling triggered by calcium ions. Ion effects in biology have often been treated at the mean-field level through solutions to the Poisson-Boltzmann equation [31–33]. However, detailed computational and theoretical studies of ion behavior, both in bulk [34–36] and in biological systems [36–39] have indicated that such treatments can be *qualitatively* incorrect in a variety of situations relevant to biology, including high ionic strength, asymmetric electrolytes, and large biomolecular charge densities. While "extended" versions of these mean-field electrolyte solution models are available [34,36,40–42], the underlying equations can be very difficult to solve in the complicated geometries presented by most biomolecules. Furthermore, it is not clear to what extent these corrections will ameliorate the shortcomings of the existing model. Since the main approximation of primitive model GCMC methods lies in its treatment of the solvent and since the "beyond mean-field"

phenomena are included in GCMC simulations, these methods offer a natural framework for exploring the ion atmospheres of biomolecules at higher levels of detail and accuracy than Poisson-Boltzmann-like models.

METHODS

This section gives a short description of the models and algorithms we use in our software package. One focus here is on the general issues of a GCMC procedure of estimating how selected parameters represent a consistent description of ionic solutions. Moreover, the program's applicability towards biologically interesting as well as computationally challenging systems like highly charged biopolymers such as DNA/RNA is addressed. Finally, some additional aspects of the code itself are discussed.

GCMC Scheme

The GCMC scheme we use in our code is essentially the same as proposed by Im *et al.* [15]. This refers especially to the explicit choice of probabilities, since the derivation of the general formulas is established knowledge of statistical mechanics. We start with the grand partition function for a system composed of M different types of particles [6]:

$$\Xi = Z(n_1, n_2, \ldots, n_M) \prod_{i=1}^{M} \sum_{n_i=0}^{\infty} \frac{(\bar{n}_i\, e^{\beta\mu_i})^{n_i}}{n_i!} \tag{1}$$

Here the n_i and $\bar{n}_i$ give the actual and equilibrium numbers of particles of type i respectively, β is the reciprocal thermal energy, and Z is the volume-normalized configuration integral

$$Z(n_1, n_2, \ldots, n_M) = V^{-N} \int_\Omega \int_\Omega \cdots \int_\Omega e^{-\beta W(\{\mathbf{x}_i\},\{n_i\})} \times \mathrm{d}(\mathbf{x}_1^{n_1}, \mathbf{x}_2^{n_2}, \ldots, \mathbf{x}_M^{n_M}) \tag{2}$$

where N denotes the total number of particles, Ω is the spatial domain of interest, the $\{\mathbf{x}_i\}$ are the particles' generalized coordinates, and $W(\{\mathbf{x}_i\},\{n_i\})$ is the system's mean free energy or, in other words, the N-particle PMF. Consequently, the probability of a particular configuration is given as:

$$P(\{\mathbf{x}_i\}\{n_i\}) = \Xi^{-1} e^{-\beta W(\{\mathbf{x}_i\},\{n_i\})} \prod_{i=1}^{M} \frac{(\bar{n}_i\, e^{\beta\mu_i})^{n_i}}{n_i!} \tag{3}$$

It is now possible to create a Markov chain for changes in numbers of particles with the restraint of detailed balance as expressed by the equality

$$P(\{\mathbf{x}_i\},\{n_i\})\, k_{(\{n_i\}\to\{n_i\}+\{\Delta n_i\})} = P(\{\mathbf{x}_i\},\{n_i\}+\{\Delta n_i\})\, k_{(\{n_i\}+\{\Delta n_i\}\to\{n_i\})} \tag{4}$$

which serves to ensure microscopic reversibility (the k-values are the *transition* probabilities) yielding the (still generalized) expression

$$\frac{k_{(\{n_i\}\to\{n_i\}+\{\Delta n_i\})}}{k_{(\{n_i\}+\{\Delta n_i\}\to\{n_i\})}} = e^{-\beta[W(\{\mathbf{x}_i\},\{n_i\}+\{\Delta n_i\})-W(\{\mathbf{x}_i\},\{n_i\})]} \prod_{i=1}^{M} \frac{n_i!\,(\bar{n}_i\, e^{\beta\mu_i})^{\Delta n_i}}{(n_i+\Delta n_i)!} \tag{5}$$

When limiting allowed transitions to one-particle insertions or deletions for one single species of type r, the expression simplifies to

$$\frac{k_{(n_r\to n_r+1)}}{k_{(n_r+1\to n_r)}} = e^{-\beta[W(\{\mathbf{x}_i\},(n_1,\ldots,n_r+1,\ldots,n_M)-W(\{\mathbf{x}_i\},(n_1,\ldots,n_r,\ldots,n_M))]} \times \left[\frac{\bar{n}_r\, e^{\beta\mu_r}}{n_r+1}\right] \tag{6}$$

which can be satisfied by a number of probability choices. The following two expressions [15] have the advantage of yielding transition probabilities mapped onto the interval [0;1]:

$$P_{\mathrm{creat},i} = \frac{\left(\frac{\bar{n}_i}{n_i+1}\right) e^{-\beta(\Delta W-\mu_i^{\mathrm{ex}})}}{1+\left(\frac{\bar{n}_i}{n_i+1}\right) e^{-\beta(\Delta W-\mu_i^{\mathrm{ex}})}} \tag{7}$$

and

$$P_{\mathrm{destr},i} = \frac{1}{1+\left(\frac{\bar{n}_i}{n_i}\right) e^{\beta(\Delta W-\mu_i^{\mathrm{ex}})}} \tag{8}$$

where μ_i^{ex} denotes the excess chemical potentials of solvation for each species and ΔW is the change in free energy for the transitions $n_i \to n_i+1$ and $n_i \to n_i-1$, respectively.

The elementary step of a simulation is to attempt to create or destroy one particle of one type. The change is accepted if a random number equally distributed over the interval [0;1] is less or equal to the transition probability specified by Eqs. (7) or (8). In terms of statistics, it makes no difference if the choice of the particle type or the choice to insert/delete are determined by an equally distributed random decision or by a systematic procedure; therefore, we generally chose the latter alternative.

Usually, grand canonical schemes also rely on ordinary Metropolis movement steps [8], as the sampling of configuration space by exclusively creating and destroying particles is expected to be highly ineffective, if not insufficient. This is understood when considering the GCE as an extension of the canonical ensemble or actually as an ensemble of ordinary, canonical (NVT) ensembles [6]. The decision whether to accept such a displacement or not is governed by the Boltzmann-weighted probability distribution

$$P_{\mathrm{mov}} = e^{-\beta\Delta W} \tag{9}$$

where ΔW gives the change in free energy associated with the displacement. Once again, if a random number equally distributed over the interval [0;1] is less or equal to the above probability, the change is accepted, otherwise the original configuration is restored. All movements out of the simulation domain are categorically rejected.

Simulation Parameters

The above description contains a number of parameters which control the behavior of a GCMC simulation. In general, the ISIM software (described below) was designed to be highly flexible and allows nearly arbitrary choice for all of the simulation parameters.

The first set of parameters govern the behavior and frequency of Metropolis steps for altering particle configurations. These "shuffling" steps may occur for single particles (as originally proposed) or with multiple particles. The former selection was frequently applied to various problems within the canonical ensemble (for instance one frequently cited reference is Ref. [43]) as well as within GCMC schemes [1,9,10]. The latter is less well established in literature, because the fundamental concern is that the probability of acceptance scales with $(P_{\text{single}})^N$, where N is the number of particles to be moved simultaneously. In the given form, this is effectively not true; in fact N has to be replaced by a smaller exponent [44]. Therefore, while it *is* possible to move several particles within one elementary step and still maintain a reasonable acceptance rate [44], the failure rate rapidly increases for moves displacing larger numbers of particles lowering the overall efficiency of the simulation thereby [45]. However, in GCMC methods, the use of multiple-particle moves dealing with a few particles at a time does have the advantage of providing additional relaxation [13]. Our own results (data not shown) clarify that the acceptance probability is a function strongly dependent on the number and the particle density. The latter is necessarily correlated to the chosen interaction potential, and the acceptance probability can actually be maintained at a reasonable level (approximately 0.1–0.25 for a system of about 100 particles) for relevant densities for our specific purpose (i.e. the simulation of electrolyte solutions in the primitive model). An entirely analogous discussion holds for the maximum displacement one or several particles are allowed to undergo before a decision-making energy evaluation. These two, conceptually-independent simulation parameters determine the position sampling efficiency-just as they would for an ordinary canonical MC procedure.

Another set of parameters governs the balance between the two types of elementary steps, i.e. changes in spatial configurations and numbers of particles. Norman and Filinov [9] give the arbitrary choice of the probability of 1/3 to perform either a creation attempt, a destruction attempt or a moving attempt, whereas Adams [10] proposed a systematic procedure equivalent to probabilities of 1/4, 1/4 and 1/2, respectively. However, it should be a clear indication of validity if structural parameters like pair correlation functions are correctly reproduced (using previous MC work [16,46,47], integral equation theories such as HNC [48,49],[¶] or even experimental results as Ref. [50]). We encountered the largest difficulties with ion–ion pair correlation functions when reducing the proportion of moving steps, especially in the limit towards zero. This is no surprise, considering the fundamental, algorithmic (in terms of energy evaluation and probability) difference between the attempt to delete one ion out of a stable pair and to re-insert it afterward at a similar position in comparison to just slightly perturbing its coordinates. In other words, the number sampling procedure provides limited sampling of the positional configuration space but is not sufficient for complete sampling. As before, the software provides maximum flexibility by supporting a wide range of parameter choices.

The most important parameter in a GCMC simulation is the excess chemical potential which determines the probability for inserting and deleting particles from the simulation and therefore, governs the equilibrium concentration. There have been a variety of ways for determining the correct chemical potential, including direct calculation from experimental measurements (directly or of activity or osmotic coefficients such as in Ref. [51]), semi-empirical fitting equations (Pitzer equations [52,53]), or simple integration equation schemes such as the Percus-Yevick or mean spherical approximations (see Ref. [54] for an overview). One of the more popular methods uses chemical potentials obtained from the hypernetted chain HNC equation in its formulation for electrolyte solutions by Allnatt [55]. However, this approach suffers from the small range of applicability of HNC calculations, which are typically limited to relatively simple systems, i.e. MC reference results are almost perfectly reproduced for an aqueous 1:1 electrolyte for concentrations up to 1 M [56–58]. On the other hand, problems occur with 2:2 electrolytes [59–61], which can be partially overcome with new closure relations [49,62]. Still the generalizability both in terms of microscopic parameters (charge/size-ratio) and the number of particle types is not satisfying.

[¶]It is a problem that most HNC calculations use MC simulations as their own reference.

To provide a simulation framework which is applicable over a wide range of concentrations, we have developed a new procedure for *parameterizing* the chemical potential for a particular electrolyte at a specified concentration. The procedure for a particular composition has to be considered as a (multi-dimensional) variation to solve the N problems

$$\left(X^{\text{comp}}_{i,\text{sim}}(\mu_1,\ldots,\mu_N) - X^{\text{comp}}_{i,\text{exp}}\right) = 0 \qquad (10)$$

where the $X^{\text{comp}}_{i,\text{sim}}$ and $X^{\text{comp}}_{i,\text{exp}}$ are the simulated and expected numbers of ions of type i, the μ_i are the single-ion excess chemical potentials (variables) and N denotes the number of different ion types. The existence of phone, consistent minimum is assumed. The N-dimensionality of the problem makes it already challenging for more than two particle types. In addition, all collected data points are for finite simulations necessarily disturbed by random noise, so that rational minimization algorithms are inapplicable. We, therefore, apply the following strategy: apart from the limitations in configuration space, there is no fundamental need to include all present particle types within the insertion/deletion cycles. Consequently by creating a system, which satisfies the specified composition, and which allows number fluctuations for only one particular type, it is possible to reduce the problem to a simple one-dimensional minimization. Then $N-1$ concentration constraints are always fulfilled (fixed number species) and only one parameter (the excess chemical potential of the selected type) is relevant. The electroneutrality constraint is always a dependent condition and is not considered any further here. For one particular composition, N of these variations are necessary, which is still computationally demanding, but significantly less costly than the full multi-dimensional optimization. For the parameterization of an individual species within such a composition, we first apply a rough search strategy to define an interval, which will most likely include the point of interest (i.e. run simulations while varying the excess chemical potential, until two values, one significantly overestimating the particle expectation number and the other one significantly underestimating it, are obtained). Six additional data points within this identified interval are then used in a simple regression on the results to determine the best value (i.e. the value yielding precisely the expectation number based according to the fitted function). Assuming this unknown function is roughly symmetric around the final value (a local harmonic approximation) this procedure works reasonably well for arbitrary functions, as long as the points are symmetrically distributed around the target value. The search algorithm finds the boundaries of the corresponding interval by a directed perturbation of the seed value, until the result is well above or below the expectation number. Here the term "well" hides a tradeoff between accuracy and efficiency, since a narrow range guarantees exact symmetry, but might make require too many refinements. In the automation of this procedure a user only specifies the composition schemes to be parametrized and initial guesses to seed the search. It might be objected that these seeds already require rough prior knowledge of the results (which would point to an extremely expensive, iterative process). However, for various simple systems (for example a single salt), it is usually possible to find results for similar cases just by browsing appropriate literature, which can serve as an initial guess. Even activity data (see for example Ref. [51] can help estimating values by calculating the mean activity coefficient for the electrolyte and then using experience-based splitting rules. For more complicated systems, it seems reasonable to build up guesses from data for the constituting elementary systems (as an illustration: in a system containing K^+, Na^+ and Cl^- one would estimate the excess chemical potential for Na^+ by taking the value from pure NaCl using a concentration, which is the sum of both K^+ and Na^+).

Microscopic Model

As mentioned above, a primitive model (dielectric continuum) is used to describe the microscopic particle-solvent interactions. Each particle experiences influences from the other particles in the simulation as well as any "external" field (i.e. as generated by a fixed biomolecule):

$$V_i = V_i^{\text{ext}}(r_i) + \frac{1}{2}\sum_{i\neq j} V_{ij}(r_{ij}) \qquad (11)$$

The particle–particle potential V_{ij} combines electrostatic contributions via Coulomb's law and steric contributions either via 6–12 Lennard-Jones (LJ) potential or pure hard-sphere (HS) potentials. The resulting expressions dependent on the inter-nuclear distance r_{ij} are

$$V_{ij}(r) = \frac{q_i q_j}{4\pi\epsilon\epsilon_0}\,\frac{1}{r_{ij}} + 4\epsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right\} \qquad (12)$$

and

$$V_{ij}(r) = \frac{q_i q_j}{4\pi\epsilon\epsilon_0}\,\frac{1}{r_{ij}} + \begin{cases} 0 & \text{for} \quad r_{ij} > a_i + a_j \\ \infty & \text{for} \quad r_{ij} \leq a_i + a_j \end{cases} \qquad (13)$$

respectively, where q_i and q_j are the particles' charges, ϵ denotes the medium's relative permittivity, ϵ_{ij} and σ_{ij} are the mean Lennard-Jones parameters according to their usual definition as geometric and arithmetic average respectively and a_i and a_j are the particles hard-sphere radii. The nature

of the solvent and ions is determined by the user-specified charge, dielectric, well-depth, and radius parameters. It is important to notice that the whole simulation relies on this microscopic description and that the literature offers a wide variety of possible choices. Choices for ubiquitous species might come from commonly used force fields [63–67].[§] Work on interaction parameters for solvated ions include free energy perturbation simulations [68] or adoption from noble gases data [69]. We feel no necessity to restrict our software to specific models in the view of the wide range of possible applications.

The external potential V_i^{ext} is defined by the nature of the fixed solute which is determined by a collection of atoms with specified position, charge, and radius. As with popular Poisson-Boltzmann models [31,33], steric interactions between ions and the fixed solute are treated in a hard-sphere fashion with ion centers completely excluded from a shell extending $\sigma_i/2$ away from the solute surface. The electrostatic potential $\phi(x)$ due to the solute charges is obtained from the Poisson equation [26].

$$-\nabla\,\epsilon(x)\nabla\phi(x) = f(x), \tag{14}$$

where $\epsilon(x)$ is the dielectric function, which typically assumes low values (2–20) inside the biomolecule and bulk solvent values outside of the biomolecule, and $f(x)$ is the solute charge distribution, usually treated as a collection of delta functions placed at the atomic centers and scaled by the partial atomic charges. Ion-solute electrostatic interactions are treated *approximately* as:

$$V_i^{\text{elec}}(r_i) \approx \frac{1}{2} q_i \phi(r_i) \tag{15}$$

which neglects the reaction field (or desolvation) potential induced by bringing a charge close to the low-dielectric biomolecular interior [26]. This approximate treatment of the ion-solute electrostatic interactions is preferred because it requires only one solution of Poisson's equation for the entire GCMC simulation. This approach is validated in the simulation results and can be justified by noting that desolvation energetics are significant only over ranges of a few solvent lengths—well beyond the level of detail appropriate to the primitive model. Solutions to Poisson's equation are obtained from the APBS software (see http://agave.wustl.edu/apbs) [70] which uses efficient parallel multigrid methods to obtain electrostatic potentials for biomolecules of arbitrary size.

System and Analysis Parameters

As implemented in the current software, every simulation takes place in either a spherical or a cylindrical simulation "droplet". The latter choice was introduced to establish a better compatibility with cylindrically-symmetric macromolecules like DNA and includes an automated rotation routine to best exploit the symmetry. The initial number of explicit ions is determined by a crude estimation of the solvent-accessible volume and the bulk concentrations. For biomolecules carrying a net charge, the initial state will be far from equilibrium and it will be necessary to discard a portion of the simulation for equilibration purposes; this equilibration phase does not enter the calculation of averaged quantities. Analysis functions include the grid-based PMF, grid-based number densities for individual ion types, specific pair correlation functions for all combinations of different species, and the essential monitoring of particle numbers and global energies. The random number generator we use [71] was specifically designed for the use in MC simulations and is a fixed parameter PRNG providing a period length of $2^{19937} - 1$.

THE ISIM SOFTWARE

The algorithms, simulation methods, and parameterization schemes described above have been implemented in ISIM, a software package designed for the GCMC simulation of ion atmospheres around atomically-detailed biomolecules. Unlike previous GCMC software packages, ISIM has been written in a very flexible manner and is designed to perform simulations for electrolyte solutions of various compositions and concentrations. ISIM was programmed in "Clean Object-Oriented C" fashion [72] which is the subset of ISO C which is compatible with C++. This choice of object-oriented programming style makes the ISIM package easy to maintain, modify, and integrate with other software while maintaining the portability and speed of traditional C code. Finally, ISIM has been written using the MALOC library [72], a hardware-abstraction package which ensures the portability of the ISIM software across a wide range of machine architectures and operating systems (see http://www.scicomp.ucsd.edu/~ mholst/).

More information about ISIM, including detailed usage instructions and availability, can be found on its website: http://mccammon.ucsd.edu/isim

EXAMPLES

In this section, we present results obtained with the software for a variety of test systems, ranging from

[§]These references most often do not represent the latest version of the force-field though.

simple ion models to challenging DNA structures. There are frequently occurring terms describing the simulation setups. A "GCMC cycle" includes one attempt to create *and* destroy *one* particle of each type, whereas a "moving cycle" always implies that *every* particle has undergone a moving attempt. The "maximum displacement" parameter represents the side length of the particle-centered cube displacements are restricted to, and it was consistently chosen as 2.0 Å in all simulations. Furthermore, the temperature and the relative dielectric constant of the medium were generally set to 298.0 and 78.36 K (pure water), respectively.

TABLE I Microscopic parameters for ionic species in this work

Species	*Charge*	ϵ_i *(kcal/mol)*	σ_i*(Å)*
Cation (univ., symm.)	+1	0.300	3.50
Anion (univ., symm.)	−1	0.300	3.50
Cation (div., symm.)	+2	0.300	3.50
Anion (div., symm.)	−2	0.300	3.50
Na^+	+1	0.358	2.73
Cl^-	−1	0.168	4.86
Ca^{2+}	+2	0.450	2.41
Mg^{2+}	+2	0.875	1.64

Data for Na^+ and Cl^- taken from Ref. [69] and for Ca^{2+} and Mg^{2+} from Ref. [68].

Parameterization

We have chosen several model electrolyte systems to demonstrate the usefulness of the excess chemical potential parameterization scheme described above. In general, each system was simulated for a total of 32000 steps with each consisting of ten GCMC and one one-particle moving cycle. The first 2000 steps were always discarded for equilibration purposes. To enforce statistical relevance the simulation sphere sizes were adjusted to ensure a minimum number of 35 particles of every type.

A symmetric model electrolyte with the microscopic parameters given in Table I was chosen to establish some data used in subsequent parts of this work. The results for the individual excess chemical potentials are shown on the left side of Fig. 1. As one would expect for a symmetrical case, both curves are identical within the limits of accuracy (equivalent to a 1:1 splitting rule). The limited amount of "noise" in the curves is due to finite sampling in the GCMC procedure and approximations inherent in the parameterization scheme; this noise could be eliminated by increasing the numbers of GCMC steps, increasing the sampling of the chemical potential/concentration profile in the parameterization scheme, or simply smoothing the existing results. The current results indicate the success of the method, which is fully automatic except for the initial seeds, which were taken from earlier "by-hand" work for a similar system in this case.

The parameterization scheme was also applied to an electrolyte with more realistic ionic radii (i.e. different ones for cation and anion). The case chosen here mimics NaCl with microscopic parameters also given in Table I. Because of the asymmetric *steric* parameters for Na^+ and Cl^- the observation of identity for the individual excess chemical potential would point to domination of all interactions by electrostatic contributions, whereas significant deviations would indicate dominating steric interactions. Figure 1 shows the corresponding plot and it is clearly visible that, up to concentrations of 200 mM, short-range interactions seem to have only negligible impact (the approximate identity to the curves for the model electrolyte is discussed below in terms of activity). However, for higher concentrations, the individual values differ (with the implications mentioned above) with the smaller ion (Na^+) yielding the more negative results. We have

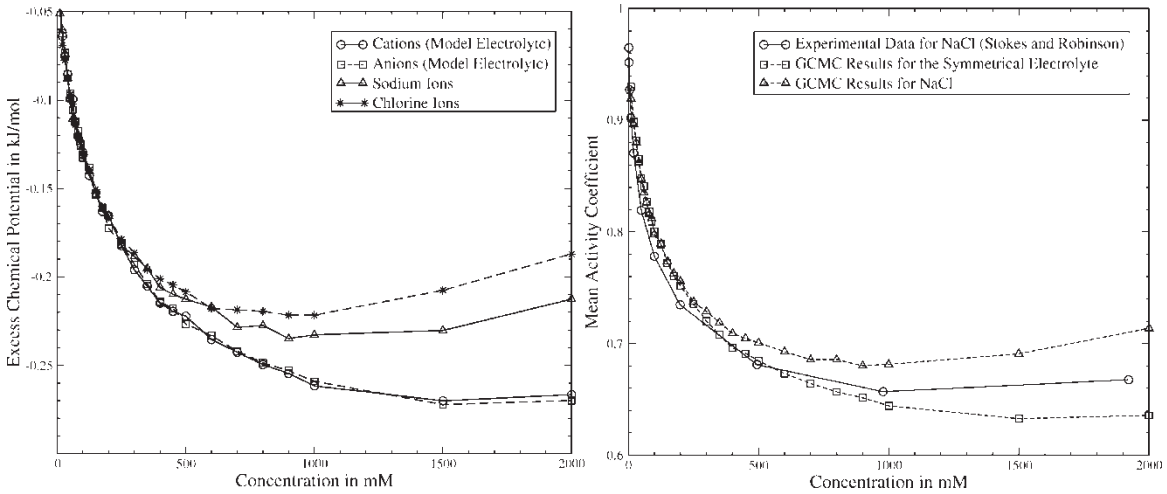

FIGURE 1 Parameterization results for the model Electrolyte and for NaCl (see text).

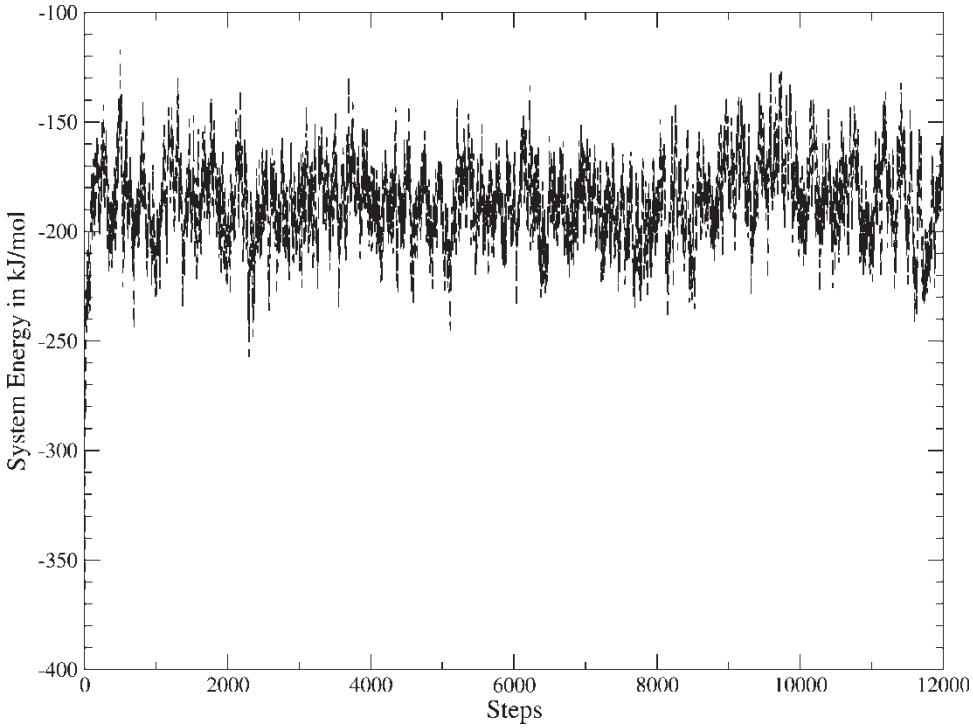

FIGURE 2 Development of total system energy throughout an empty-box simulation.

also compared activity coefficients obtained from our simulations with experimental data. In terms of individual excess chemical potentials, the mean ionic activity coefficient for an electrolyte $A_{\nu_+}B_{\nu_-}$ is given as:

$$\gamma_{\pm} = \exp\left[\frac{\beta\,\mu_{\pm}^{\mathrm{ex}}}{\nu_+ + \nu_-}\right] = \exp\left[\frac{\beta\,(\nu_+\mu_+^{\mathrm{ex}} + \nu_-\mu_-^{\mathrm{ex}})}{\nu_+ + \nu_-}\right] \quad (16)$$

A comparison of the results for the model electrolyte and for NaCl with experimental data for the latter [51] is given on the right side of Fig. 1. Obviously, the agreement is much better than the results reported for the more established GCMC approach to unify the ionic species in deletion and insertion events. Reportedly, [22], it was necessary to introduce Gurney sphere terms for the potential to be able to reproduce experimental activity curves even at a qualitative level, whereas simple adjustments for the primitive model parameters failed. A similar observation was made for osmotic coefficients [73] in earlier work. The qualitative agreement is promising and the use of other parameter sets with the current GCMC method may lead to better quantitative agreement, despite previous unsuccessful attempts with the concerted insertion/deletion GCMC methods [22]. Additionally, the GCMC data at low concentrations (approximately up to 200 mM) for both systems are very similar, indicating that small changes in the parameters will *not* affect the results. Still this is consistent with the fact that the experimental data sets for many alkali halides and other comparable 1:1-electrolytes also show rough agreement in that range [51].

Finally, the parameterization procedure was applied to a more complicated system composed of NaCl and $MgCl_2$ in various ratios. The above data for NaCl and previous previous results for pure $MgCl_2$ solutions (data not shown) were used as initial guesses. We focus on the reliability of the results here by presenting the simulation of an arbitrary mixture of 178 mM NaCl and 36 mM $MgCl_2$ (microscopic parameters for magnesium in Table I) in a spherical system with 70Å radius. A total of 12000 steps were performed each consisting of ten GCMC cycles and three one-particle moving cycles with the first 2000 steps being ignored (equilibration). The data for pair correlation functions was collected in every tenth step. Based on the parameterization run, we used -0.1821, -0.9416 and -0.2222 kcal/mol for Na^+, Mg^{2+} and Cl^-, respectively. Figure 2 shows that the general convergence (here in terms of system energy, but equally valid for ion numbers) is fast and that the results are stable throughout the whole simulation. However, this does *not* indicate that *all* observable quantities are well-averaged after a several hundred steps in full spatial resolution.

Another form of general validation is provided by pair correlation functions. They provide insight on a microscopic level and are fundamental quantities of the theory of electrolyte solutions [54]. Figure 3 shows our results for the various possible combinations of ions in the system. In general the noise is still significant, which is not surprising considering that only 1000 snapshots entered the calculation. The distances of closest approach precisely reflect

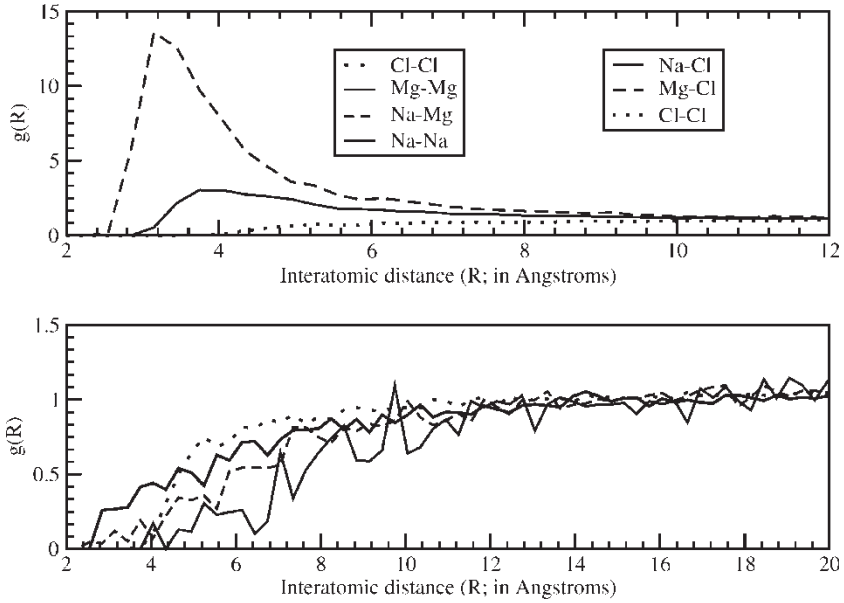

FIGURE 3 Pair correlation functions as determined by an empty-box simulation. **Top**. Pair correlations for Mg^{++}-Cl^{-} (dashed line), Na^{+}-Cl^{-} (solid line), and Cl^{-}-Cl^{-} (dotted line) pairs. **Bottom**. Pair correlations for Cl^{-}-Cl^{-} (dotted line), Mg^{++}-Mg^{++} (thin solid line), Na^{+}-Mg^{++} (dashed line), and Na^{+}-Na^{+} (thick solid line) pairs.

the Lennard-Jones radii of the individual species. The curves actually show the typical shapes for the primitive model and are qualitatively in agreement with results obtained via HNC [48,56,57], Poisson-Boltzmann studies [46,47] or most important with other MC studies [15,16,46–48] The quantitative agreement cannot be fully tested, since a system of that type is already difficult to address with continuum studies, but the peak values of about 3.0 for the 1:1 like-pair and 13.0–14.0 (not visible in the plot for better resolution) for the 2:1 like-pair are in the "normal" range, proving that the configuration space sampling is sufficiently balanced.

The open aspects of parameterization are the comparison between actual and "target" concentrations as well as maintenance of electroneutrality. The system described above contained on average 155.12 Na^{+} ions, 32.48 Mg^{2+} ions and 220.04 Cl^{-} ions equivalent to bulk concentrations of about 179 mM for NaCl and 37.5 mM for $MgCl_2$. Consequently, both conditions are consistently fulfilled, which appears even more promising when considering that, in case of larger deviations for the particle densities, all simulation results could easily be re-interpreted using the *actual* bulk concentrations.[‖]

The Born Ion

The task of obtaining a PMF for a system composed of arbitrary charge distributions is usually a serious one and accomplished by numerically solving the Poisson-Boltzmann equation. However, in case of a spherical, symmetrically charged particle in an electrolyte solution the problem has a rather simple analytical solution, which is just a slight re-interpretation of the work by Debye and Hückel [74] for a single ion as part of an electrolyte solution:

$$\phi(r) = \frac{q_{mm}}{4\pi\epsilon\epsilon_0}\,\frac{1}{r}\,\frac{e^{[\kappa(a_{mm}-r)]}}{1+\kappa a_{mm}} \quad \text{for } r \geq a_{mm}$$

$$\phi(r) = \text{for } r \leq a_{mm}$$

Here q_{mm} denotes the macroion's charge, a_{mm} is the distance of closest approach between it and surrounding ions (making a size-symmetrical electrolyte necessary), r is the (radial) distance from the macroion's center and κ is the so called Debye parameter

$$\kappa = \frac{\beta N_A e^2}{\epsilon\epsilon_0}\sum_j c_j z_j^2 \tag{17}$$

However, the derivation of these results makes some assumptions and, as a consequence [51], the Debye-Hückel theory works best for symmetric low-charge electrolytes at small to very small concentrations. Apart from mathematical approximations, the rigorously imposed model (point charges within a continuum solvent) is the main reason for deviations from experimental, more sophisticated theoretical or simulation data. The GCMC treatment with explicit ions in a primitive solvent includes the ions' spatial extension. Therefore, one might expect good agreement between computer experiment and

[‖]This must not be confused with a re-interpretation in terms of parameterization though. It is impossible to get quantitative insight about possible corrections, as the concentration enters the algorithm in two distinct ways.

Debye-Hückel theory only for a high ratio of the macroion's radius versus the electrolyte species' radii, for a moderate macroion charge, and for relatively low concentrations of an electrolyte composed of univalent ions. Unfortunately, it is usually difficult to separate all possible deviations within a given result. However, we still believe that it *is* possible to show some trends with a few simulations, which are described and analyzed in the following.

General Simulation Setup

All simulations were run a total of 12000 steps with the first 2000 of them as equilibration, with each step consisting of ten GCMC cycles and three one-particle moving cycles. All simulation systems were spherical with a Born ion of radius 15.0 Å at their respective centers. Grid-based potential analysis was performed every 24th step.

Reference Case

A Born ion carrying a net charge of $+5e$ within a 120.0 Å radius simulation sphere containing 40 mM of the symmetrical electrolyte represents a case for which Debye-Hückel theory works well. The necessary parameters for the cations and anions are given in Table I. The excess chemical potentials used are taken from the data set described in the previous section.

The simulation shows excellent agreement with Debye-Hückel theory (as shown in Fig. 4A) for the mean potential. It is possible to transform the Debye-Hückel solution into ion number densities via the (more general) relation

$$n_{\pm}(r) = n_{\infty,\pm}\, e^{(-\beta c_{\pm}\phi(r))} \tag{18}$$

where $n_{\infty,\pm}$ and $c_{\pm}$ are the ionic species' bulk density and charge, respectively. This formulation holds for 1:1 electrolytes. A comparison of simulated number densities and these calculated values can be found in Fig. 4B showing very good agreement within the limits of accuracy dictated by the MC noise and the concentration error. The edge effect visible in all plots of this kind is due to the "droplet" model, which treats the surroundings as vacuum. The missing component of the attractive field around ions close to the boundary unrealistically lowers the bulk density.

Higher Charge

If the reference case is modified by increasing the macroion's charge from $+5$ to $+30e$, the assumption that the potential close to the macroion is small compared to the particles' mean thermal energy becomes critical. We used an otherwise unchanged setup to obtain the results shown in Fig. 4C. There are serious deviations between "experiment" and theory, with the analytical solution tending to underestimate the screening effect of the surrounding solution. We do not see any reason to blame the GCMC results for the deviations as the Debye-Hückel model is even more simple and the approximation made in its derivation is apparently crude for the present system (actually the simulation predicts that the electrostatic energy still equals the thermal energy for univalent counterions at 298.0 K at about 10 Å beyond the distance of closest approach). Within this interpretation, a comparison of number densities obviously makes no sense.

Higher Concentration

The validity of the primitive model also decreases with increasing concentration. The steric effects of ions in the GCMC eventually lead to moderate deviations from the Debye-Hückel results. Once again we used a similar setup only changing the concentration to 800 mM and (roughly) proportionally reducing the simulation sphere size to 45.0 Å. The corresponding results for the mean potential and ion number densities are shown in Figs. 4D and E, respectively. Obviously there is a significant bulk concentration error of about 6–7%. While this may be partially due to inaccuracies in parameterization, it is mainly caused by the previously described edge effect in combination with the criterion to meet total ion numbers instead of effective bulk concentrations within the parameterization scheme. Necessarily this effect has much more impact at high concentrations. This certainly points to future improvements in terms of introducing a mean field replacing the vacuum currently "seen" by the ions. Still it is possible to obtain the reference data according to Debye-Hückel theory with the effective bulk concentration (these are the plots called "corrected" in Figs. 4D and E, whereas the ones labeled "original" use the "target" concentration). Obviously, the agreement is reasonably good with the slight deviations (especially compared to the high charge case) most likely being a result of steric pressure close to the macroion. Interestingly, the result is *increased* screening of the Born ion's charge (see Fig. 4D), which might be interpreted as a consequence of the higher likeliness for the anion concentration to be significantly decreased below the "ideal" Debye-Hückel value than for the counterion concentration.

Calmodulin

An excellent demonstration of the software's capability to handle biomolecules would be the correct identification of (exposed) ion binding sites in proteins. Calmodulin, an ubiquitous protein in living

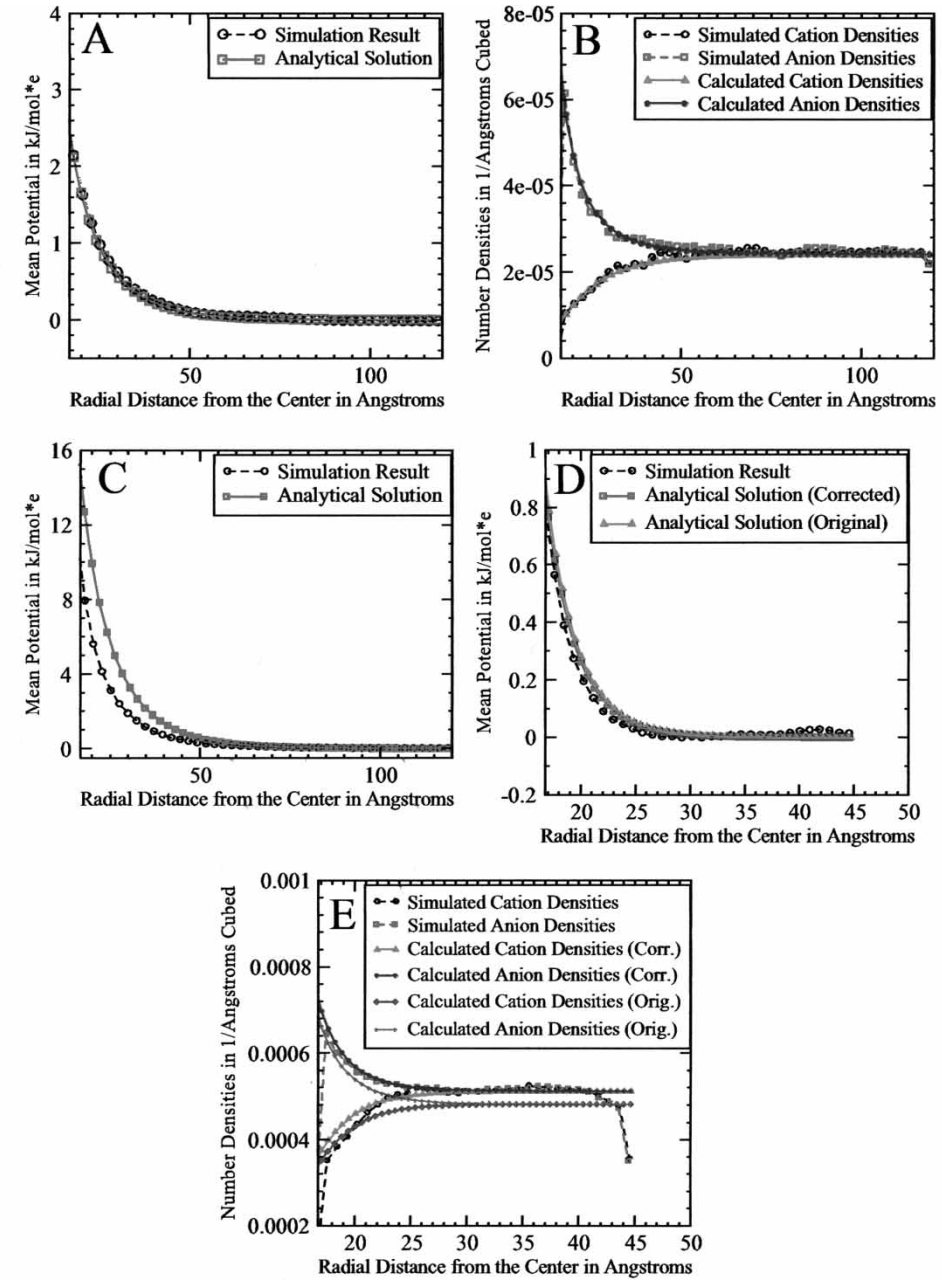

FIGURE 4 Mean potential and number densities for a born ion in different simulation setups (see text).

cells, was chosen as an example, because it has four high-affinity Ca^{2+} binding sites and high resolution X-ray structures for the calcium-bound form are available (PDB entries 3CLN [75], 1CLL [76], and 1EXR [77]).[#] The calcium-free form represents an entirely different problem due to the extensive structural change known to occur upon calcium binding and is not addressed here.

In the metal-bound form calmodulin is a dumbell-shaped protein composed of 148 amino acids. Both the N- and the C-terminal lobes feature two EF hand motifs [78,79], which are the primary calcium binding sites. One might be concerned about the concept to approach such a problem via the method presented here, as it accounts neither for specific binding phenomena apart from electrostatics nor for the possibly essential flexibility of the polypeptide to faciliate binding nor for any explicit solvent effects. However, the assumption that the interaction is primarily mediated by electrostatics seems reasonable.

We chose to simulate calmodulin in a cylindrical simulation domain of 65.0 Å radius and 150.0 Å length containing 75 mM $CaCl_2$. The salt concentration is

[#]The structure we use is the one from Chattopadhyaya *et al.* the more recent one from Wilson *et al.* is less well-documented, as the publication has a slightly different focus.

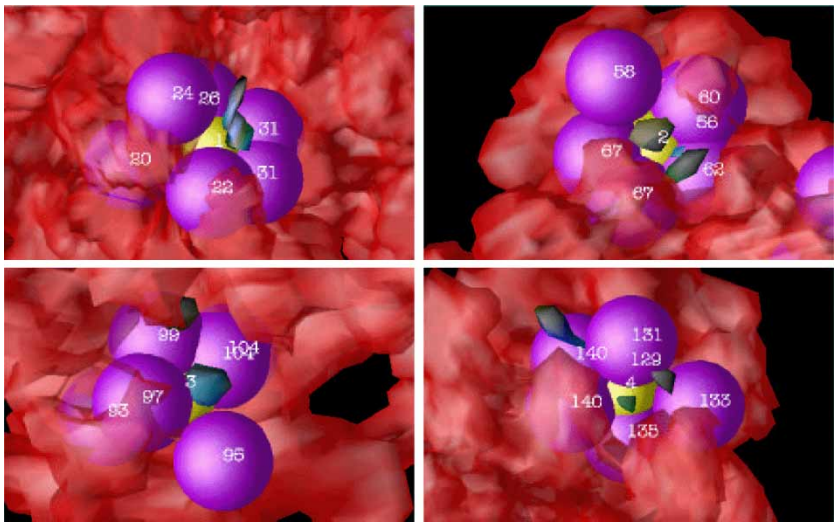

FIGURE 5 The four EF-hand motifs with calcium number density isosurfaces of 0.055 $Å^{-3}$ (see text for details).

different from cellular levels, but it is neither our aim here to gain insight into the concentration dependency of calcium binding nor to prove orders of specificity among different counterions. The simulation consisted of 205000 steps (50000 equilibration steps,[**] 10 GCMC cycles, 3 one-particle moving cycles), the microscopic parameters for the ions can be found in Table I. The excess chemical potentials used were -0.6844 and -0.2205 kcal/mol for Ca^{2+} and Cl^{-}, respectively. The macromolecule was prepared by using a conversion routine to obtain PQR- from PDB-format [80] thereby removing all metal ions, water molecules and the single ethanol molecule. This was followed by an APBS calculation in pure water (multigrid method on a cubic grid with 129.0 Å side length and a resolution of 1.0 Å, other parameters chosen consistently) writing out the necessary input file for the electrostatic potential.

The simulation system contained on average 197.76 Cl^{-} and 110.90 Ca^{2+} ions providing (together with the protein's net charge of $-24e$) an approximately neutral solution. The estimated bulk concentration was 83 mM with a relative error of 11%. Figure 5 shows detailed views of the four calcium binding sites. For all of them, the position of the metal ion in the crystal structure is indicated by a yellow sphere, the accessibility surface of the protein is shown in red, and the ligand oxygen atoms are depicted as purple spheres with a size approximately corresponding to their force-field radii [67]. Finally the isosurface for a calcium number density of 0.055 $Å^{-3}$ (representing more than 1000 times the bulk level) is shown as a solid, colored surface (mostly black and blue). All single pictures are rather similar in a sense that the actual position of the calcium ion in the crystal structure does not seem to be accessible within the simulation, which is not surprising comparing the actual distances from the ligands in the crystal structure [75] with the employed radii. Still in all four cases significant condensation of calcium ions obviously with high affinity at slightly different positions can be observed, which is very satisfying when considering that, apart from the surfaces shown, no significant binding site candidates (i.e. visible isosurfaces at the same level) were discovered. Therefore, the simulation clearly predicts four major calcium sites in agreement with structural knowledge. These results can hardly be misinterpreted in terms of the exact positions of the ions, if one is aware of the fundamental restrictions previously outlined. In this particular case, it is possible to derive the hypothesis, that the EF-hand motif provides a certain spatial domain, which in terms of electrostatics seems to be very attractive for small, divalent cations and is more or less accessible for the surroundings. For the same system (among others) the task of predicting binding sites has been previously addressed by empirical, knowledge-based methods [81,82]. These authors obtained the similar conclusion that very high values for their scoring function occurred around the precise locations of calcium ions in protein crystal structures.

DNA

The ion atmosphere around DNA is a very well-established and at the same time challenging problem. Structural and functional diversity might often be mediated by sequence but, for general

[**]This number was chosen, because the type of microscopic equilibrium dealt with here cannot be monitored via bulk quantities and might require a lot more random events.

properties, counterion and solvation effects are thought to be a crucial factor [83]. Every nucleic acid is a polyanion at physiological pH, which consequently emphasizes the importance of the counterion and solvent surroundings and their role in screening the net charge.

The problem has been studied by various approaches (for an extensive review with valuable references see Ref. [83]), starting with the elementary description in terms of a general polyelectrolyte theory by Manning ([84] and references therein). According to this theory, the counterion distribution at homogeneously charged, cylindrical macromolecules is characterized by a charge compensation due to condensation within the empirical Manning radius (7 Å for B-DNA) of 0.76 for a univalent and 0.88 for a divalent cation in the limit of vanishing ionic strength. Many computational studies have been carried out on the topic including Metropolis MC studies on both model polyelectrolytes of different complexity [85,86] and on detailed models of DNA [30], GCMC studies on model polyelectrolytes [23], HNC and Poisson-Boltzmann studies [87,88], and more recently MD studies with explicit solvent and counterions [89,90].

Especially in the treatment of nucleic acids the question, whether to use hydrated radii for the size parameters of the ions, must be addressed. We chose to stay with the model used so far, because modeling hydration shells around ions via hard-sphere contacts with the DNA molecule might-among other incompatibility issues to the primitive model-introduce a serious approximation itself. Still we are aware of the fact that hydration phenomena play an important role in terms of improving the model in the scope of future work.

To further demonstrate the capabilities of our software, we present results on a short piece of B-DNA (PDB entry 1D8G, [91]), which has the sequence $d(CCAGTACTGG)_2$ in a solution of approximately 150 mM NaCl and 20 mM $MgCl_2$ with the microscopic parameters again given in Table I. The excess chemical potentials used (based on parameterization data not shown) were -0.1726 kcal/mol for Na^+, -0.8637 kcal/mol for Mg^{2+}, and -0.1977 kcal/mol for Cl^-. The simulation was carried out in a cylindrical tube of 50 Å radius and 80 Å length having the DNA's helical and the system's cylindrical axis aligned. Each of the 250000 steps consisted of ten GCMC cycles and three one-particle moving cycles with the first 25000 steps treated as equilibration. For analysis purposes a grid resolution of 80 (for all dimensions) was selected. The PDB-file preprocessing was identical to that for calmodulin except for a different size of the grid ($74 \times 63 \times 61$ Å with a resolution of 129 in all dimensions) in the pure solvent APBS calculation.

The system contained on average 56.42 Na^+, 14.05 Mg^{2+} and 66.57 Cl^- ions, almost perfectly fulfilling the electroneutrality condition to compensate the macromolecule's net charge of $-18e$. The estimated bulk concentrations were very close to the proposed values with ranges of 184–188 mM for Cl^-, 144–150 mM for Na^+, and 18–20 mM for Mg^{2+}. The results on ion condensation are presented in the following.

Bulk View

In a bulk view there are two indicative properties to monitor. One is the compensation of the polyanion charge by counterions, which should approximately follow Manning's theory, and the other one is given by the corresponding ion distribution functions around the cylindrical DNA. Both are addressed here in Fig. 6A by plotting the individual, radially averaged number densities together with the net charge, which is the sum of the macromolecule charge (assumed to be localized on the cylindrical axis) and the integrated charge density caused by the specific distribution of ionic species. The whole simulation system is included in these data sets, although this implies that edge effects (i.e. contributions from the axially outer regions of the cylinder) might be able to disturb the plots, but corresponding tests showed only negligible impact.

Apparently the counterion condensation is limited almost solely to Mg^{2+} ions, as the (radially averaged) concentration level for Na^+ only weakly exceeds the bulk value at a small distance interval around 15.0 Å. On the other hand, the divalent cations clearly condense in close proximity to the macromolecule, as the standard radius for B-DNA is usually given as 10 Å (in fact the DNA piece's *maximum* radius for the present simulation is even bigger with approximately 15 Å) around the central axis and all main peaks obviously lie within this distance. There are at least three distinguishable maxima at about 4.7, 6.0 and 9.0 Å, a further interpretation is not given in this bulk view. Finally, the presence of Cl^- ions closely follows that of Na^+ ions, again emphasizing the fact that the charge compensation and ion condensation are more or less restricted to the divalent counterions here. Young *et al.* [89] reported distinct peaks at distances of about 5 and 13 Å for the distribution of sodium ions around DNA in absence of other counterions. A qualitative comparison with our results shows a general agreement for the main peak region in Fig. 6A and the first peak observed of these authors, but the weights are entirely different, which is not surprising considering the fundamental differences in the simulations.

The black line in Fig. 6A demonstrates the reasonable agreement of the results with Manning's condensation theory for divalent cations [84] with

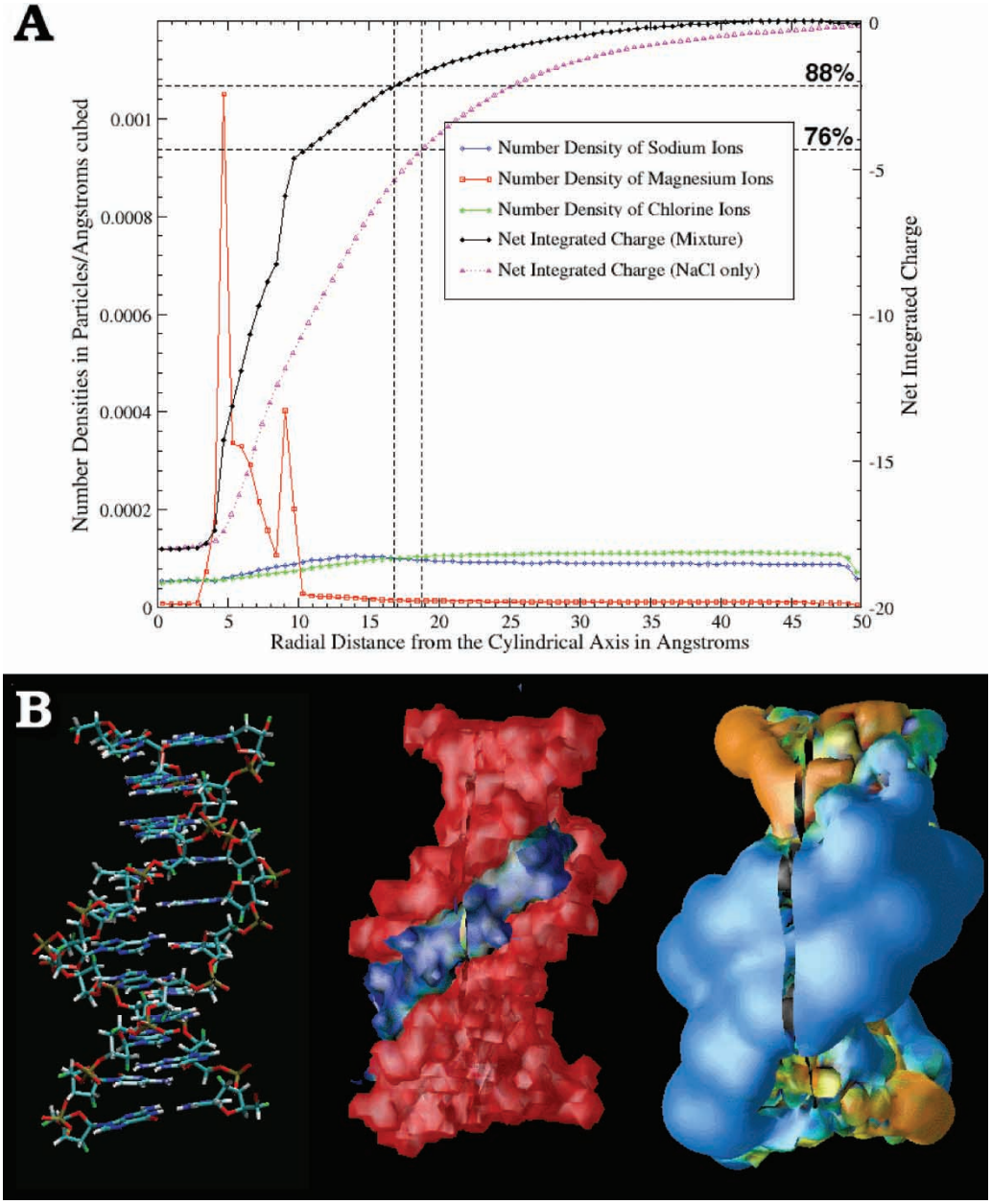

FIGURE 6 Analysis of the simulation with a piece of B-DNA in both bulk and microscopic view (see text for explanations).

the characteristic value of 88% charge compensation (upper dashed, horizontal line) within 7 Å from the DNA's surface. The left dashed, vertical line shows the empirical Manning radius for our simulation (i.e. the point of 88% charge compensation in the plot) and the value of roughly 17 Å corresponds perfectly with the theoretical prediction. Partially this agreement might be coincidental especially considering the fact that the values given by Manning vary significantly with the type of DNA used [84]. We simply assume that the piece of DNA in our simulation behaves like native DNA. Moreover edge effects are included. However, qualitatively it is important to notice that the "tight" counterion condensation indeed is not sufficient to compensate the DNA charge and that an extended environment with slight, "diffuse" ion condensation is predicted by the simulation. Apart from that it might be noticed that the peaks in the radial number density of Mg^{2+} ions coincide with the steeper regions in the charge compensation plot. For a further proof of our method, an analogous simulation in 150 mM NaCl was done. The only result shown here is the charge compensation plot as magenta, dotted line in Fig. 6A. For this case the theory predicts a value of 76% for the same radius. Our simulation would indicate a slightly extended Manning radius of 19 Å, which is certainly within the limits of accuracy for the reasons mentioned above. It appears promising that the software consistently reproduces results of

the analytical theory in independent simulations with reasonable, quantitative agreement.

Microscopic View

In this paragraph, we present pictures showing the actual distribution of counterions in terms of number density isosurfaces and the resulting mean, screened electrostatic potential both averaged throughout the simulation. A structural basis for the bulk observations in the previous paragraph is given.

The left side of Fig. 6B shows a molecular representation of the piece of B-DNA used in the simulation with the minor groove oriented towards the viewer. The middle image[††] shows the accessibility surface resulting from the atomic arrangement, which is "seen" by the explicit ions, as red and slightly transparent shape. Apart from that the blue shape represents a number density isosurface for Mg^{2+} ions with an absolute value of 0.00035 $Å^{-3}$, which is $\sim$30 times bulk level. The orientation is the same as in the left picture. Obviously the main mechanism of counterion condensation is an occupation of the minor groove itself by the divalent species. Apparently it seems to provide an attractive environment for these ions, which out-competes the helically aligned phosphate charges as possible, more localized binding sites. Despite the tight contact between macromolecule and ions here, this seemingly has to be understood as a diffuse layer formed by moving ("delocalized") particles, as can be concluded from the smooth and continuous shape of the isosurface. Reduction of the number density isosurface value (visualization not shown) generally reveals a broad environment of slightly increased counterion condensation around the DNA molecule, which still has a shape following the helical pattern defined by the minor groove. This observation holds analogously for Na^+ ions, but only for levels of up to $\sim$5 times the bulk value (maximum in the minor groove). Stronger condensation was not observed, which is in agreement with the (averaged) bulk results presented in the previous paragraph. Apparently sodium ions are playing a weak, supporting role in screening the charge of nucleic acids for the system chosen here. Still the conclusion that counterion condensation and charge compensation are mostly mediated my Mg^{2+} ions is verified here on a microscopic level.

Finally the right image—also in the same orientation-shows—the mean electrostatic potential in the simulation system by giving two isosurfaces with values of $\sim \pm 1kT/e$ (blue is negative, red is positive). It should be noticed that the potential is highly screened, i.e. the isosurfaces are found rather close to the DNA molecule, the rest of the system has PMF values decaying from these values to the bulk state (zero). The second important observation is that the counterion condensation does not yield a net neutralization of the phosphate charges or even a potential inversion, which agrees with Manning's theory and the results shown in the previous paragraph. It implies that the asymmetric screening effect provided by counterion condensation is not able to completely eliminate the underlying charge asymmetries, which means that approaching charged species "see" DNA more like a helical polyanion than like a simple cylindrical one.

FUTURE PROSPECTS

In this work we presented a new software package (ISIM), which is capable of doing explicit simulations of ionic atmospheres around fixed solutes within a primitive solvent model and the statistical mechanical framework of the GCE. We have introduced a new parameterization strategy for our independent treatment of different ionic species. Technically the software is designed to be flexible, portable and easily extensible.

We showed that the parameterization yields consistent results in terms of maintaining the prospected concentrations (and thereby electroneutrality) as well as reasonable, absolute values by comparing to experimental activity data. We have proven the general validity by successfully reproducing Debye-Hückel theory for simple electrolyte solutions. The more challenging test cases of actual biomolecules gave results, which met our expectations built up from structural and theoretical reference data and the inherent limitations of the approach.

Future work might include improvements of the model such as introducing an outer field to eliminate the edge effect or to allow for a more sophisticated and consistent model in terms of steric interactions between explicit ions and macromolecule atoms. Other straightforward extensions could include the support of explicit solvent models (with the solvent molecules being handled in a slightly separate fashion) or the refinement of the generic interaction potential (polarization effects, reaction field effects). Longer-term improvements might be the coupling to force-based simulation methods with the help of local control [13] to be able to study off-equilibrium systems or in a limited sense dynamical aspects. We also believe that it should be possible to implement suggestions and initiatives from different fields of science addressing problems within a similar framework.

[††]There is a cleft in the middle an right pictures visible, which is due to missing interpolation in the graphics output and does not have any significance.

Acknowledgements

The authors would like to thank P. Wolynes for helpful discussions on limitations of the Poisson-Boltzmann equation, W. Im and B. Roux for comments on grand canonical simulations, and H. Resat for advice on Monte Carlo methods. This work has been supported in part by grants from NIH, NSF, CTBP, NBCR, SDSC, and the W. M Keck Foundation.

References

[1] Valleau, J.P. and Cohen, L.K. (1980) "Primitive model electrolytes. I. Grand canonical Monte Carlo computations", *J. Chem. Phys.* **72**, 5935–5941.
[2] Friedman, H.L. (1960) "Mayer's ionic solution theory applied to electrolyte mixtures", *J. Chem. Phys.* **32**, 1134–1149.
[3] Høye, J.S. and Stell, G.R. (1977) "Ionic solution theory for non-ideal solvents. Potential of mean force between ions", *Faraday Discuss. Chem. Soc.* **64**, 16–21.
[4] McMillan, W.G. and Mayer, J.E. (1945) "The statistical thermodynamics of multicomponent systems", *J. Chem. Phys.* **13**, 276–305.
[5] Friedman, H.L. (1977) "Introduction", *Faraday Discuss. Chem. Soc.* **64**, 7–15.
[6] McQuarrie, D.A. (1973) *Statistical mechanics* (Harper Collins Publishers, New York, NY).
[7] Allen, M.P. and Tildesley, D.J. (1987) *Computer simulation of liquids* (Clarendon Press, Oxford).
[8] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.E. and Teller, E. (1953) "Equation of state calculations by fast computing machines", *J. Chem. Phys.* **21**, 1087–1092.
[9] Norman, G.E. and Filinov, V.S. (1969) "Investigations of phase transitions by a Monte Carlo methods", *High Temp. (USSR)* **7**, 216–222.
[10] Adams, D.J. (1975) "Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid", *Mol. Phys.* **29**, 307–311.
[11] Heffelfinger, G.S. and Lewitt, M.E. (1996) "A comparison between two massively parallel algorithms for Monte Carlo computer simulation: an investigation in the grand canonical ensemble", *J. Comp. Chem.* **17**, 250–265.
[12] Zara, S.J. and Nicholson, D. (1990) "Grand canonical ensemble Monte Carlo simulation on a transputer array", *Mol. Simul.* **5**, 245–261.
[13] Papadopoulou, A., Becker, E.D., Lupkowski, M. and van Swol, F. (1993) "Molecular dynamics and Monte Carlo simulations in the grand canonical ensemble: local versus global control", *J. Chem. Phys.* **98**, 4897–4908.
[14] Heffelfinger, G.S. and van Swol, F. (1994) "Diffusion in Lennard-Jones fluids using dual control volume grand canonical molecular dynamics simulation (dcv-gcmd)", *J. Chem. Phys.* **100**, 7548–7552.
[15] Im, W., Seefeld, S. and Roux, B. (2000) "A grand canonical Monte Carlo-Brownian dynamics algorithm for simulating ion channels", *Biophys. J.* **79**, 788–801.
[16] Wu, G.W., Lee, M. and Chan, K.Y. (1999) "Grand canonical Monte Carlo simulation of an electrolyte with a solvent primitive model", *Chem. Phys. Lett.* **307**, 419–424.
[17] Resat, H. and Mezei, M. (1994) "Grand-canonical Monte-Carlo simulation of water positions in crystal hydrates", *J. Am. Chem. Soc.* **116**, 7451–7452.
[18] Resat, H. and Mezei, M. (1996) "Grand canonical ensemble Monte Carlo simulation of the dCpG/proflavine crystal hydrate", *Biophys. J.* **71**, 1179–1190.
[19] Hribar, B., Vlachy, V. and Pizio, O. (2001) "Equilibrium properties of a model electrolyte adsorbed in quenched disordered charged media: the ROZ theory and GCMC simulations", *J. Phys. Chem. B* **105**, 4727–4734.
[20] Wang, Q. and Johnson, J.K. (1997) "Path integral grand canonical Monte Carlo", *J. Chem. Phys.* **107**, 5108–5117.
[21] Jedlovszky, P. and Mezei, M. (1999) "Grand canonical ensemble Monte Carlo simulation of a lipid bilayer using extension biased rotations", *J. Chem. Phys.* **111**, 10770–10773.
[22] Jayaram, B. and Beveridge, D.L. (1991) "Grand canonical Monte Carlo simulations on aqueous solutions of NaCl and NaDNA: Excess chemical potentials and sources of non-ideality in electrolyte and polyelectrolyte solutions", *J. Phys. Chem.* **95**, 2506–2516.
[23] Mills, P., Anderson, C.F. and Record, M.T. Jr., (1986) "Grand canonical Monte Carlo simulations of thermodynamic coefficients for a primitive model of DNA-salt solutions", *J. Phys. Chem.* **90**, 6541–6548.
[24] Mezei, M. (1980) "A cavity-biased (T, V, μ) Monte Carlo method for the computer simulation of fluids", *Mol. Phys.* **40**, 901–906.
[25] Mezei, M. (1987) "Grand-canonical ensemble Monte Carlo study of dense liquid Lennard-Jones, soft spheres and water", *Mol. Phys.* **61**, 565–582.
[26] Jackson, J.D. (1975) *Classical electrodynamics*, 2nd Ed. (John Wiley and Sons, Inc., New York, NY).
[27] Pitzer, K.S. (1984) "Ionic fluids", *J. Phys. Chem.* **88**, 2689–2697.
[28] Rasaiah, J.C. (1973) "A view of electrolyte solutions", *J. Solution Chem.* **2**, 301–338.
[29] Friedman, H.L. (1981) "Electrolyte solutions at equilibrium", *Ann. Rev. Phys. Chem.* **32**, 179–204.
[30] Jayaram, B., Swaminathan, S., Beveridge, D.L., Sharp, K. and Honig, B. (1990) "Monte Carlo simulation studies on the structure of the counterion atmosphere of B-DNA. Variations on the dielectric primitive model", *Macromolecules* **23**, 3156–3165.
[31] Davis, M.E. and McCammon, J.A. (1990) "Electrostatics in biomolecular structure and dynamics", *Chem. Rev.* **94**, 7684–7692.
[32] Sharp, K.A. and Honig, B. (1990) "Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation", *J. Phys. Chem.* **94**, 7684–7692.
[33] Honig, B. and Nicholls, A. (1995) "Classical electrostatics in biology and chemistry", *Science* **268**, 1144–1149.
[34] R.R. Netz, H. Orland. Beyond Poisson-Boltzmann: fluctuations and correlations, 1999. arXiv:cond-mat/9902085.
[35] Podgornik, R. and Žekš (1988) "Inhomogneous Coulomb fluid: a functional integral approach", *J. Chem. Soc. Faraday Trans. 2* **84**, 611–631.
[36] Nguyen, T.T., Grosberg, A.Yu and Shklovskii, B.I. (2000) "Screening of a charged particle by multivalent counterions in salty water: strong charge inversion", *J. Chem. Phys.* **113**, 1110–1125.
[37] Holm, C., Kékicheff, P., Podgornik, R., eds, (2001) *Electrostatic Effects in Soft Matter and Biophysics* NATO Science Series, (Kluwer Academic Publishers, Boston).
[38] Tanaka, M. and Grosberg, A.Y. (2001) "Giant charge inversion of a macroion due to multivalent counterions and monovalent coions: molecular dynamics sutdy", *J. Chem. Phys.* **115**, 567–574.
[39] Ha, B.-Y. and Liu, A.J. (2001) "Effect of nonzero chain diameter on DNA condensation", *Phys. Rev. E.* **63**, 021503.
[40] Coalson, R.D. and Duncan, A. (1992) "Systematic ionic screening theory of ions", *J. Chem. Phys.* **97**, 5653–5661.
[41] Shklovskii, B.I. (1999) "Screening of a macroion by multivalent ions: correlation-induced inversion of charge", *Phys. Rev. E.* **60**, 5802–5811.
[42] Abbas, Z., Gunnarsson, M., Ahlberg, E. and Nordholm, S. (2002) "Corrected Debye-Hückel theory of salt solutions: size asymmetry and effective diameters", *J. Phys. Chem. B.* **106**, 1403–1420.
[43] Brush, S.G., Sahlin, H.L. and Teller, E. (1966) "Monte Carlo study of a one-component plasma. i", *J. Chem. Phys.* **45**, 2102–2118.
[44] Chapman, W. and Quirke, N. (1985) "Metropolis Monte Carlo simulations of fluids with multiparticle moves", *Physica B* **131**, 34–40.
[45] G.S. Heffelfinger. In *Proceedings of the 1992 DAGS/PC Symposium*, page 229, Hanover, 1993. Dartmouth Institute for Advanced Graduate Studies.
[46] Valleau, J.P., Cohen, L.K. and Card, D.N. (1980) "Primitive model electrolytes. II. The symmetrical electrolyte", *J. Chem. Phys.* **72**, 5942–5954.

[47] Reščič, J., Vlachy, V., Outhwaite, C.W., Bhuiyan, L.B. and Mukherjee, A.K. (1999) "A Monte-Carlo simulation and symmetric Poisson-Boltzmann study of a four-component electrolyte mixture", *J. Chem. Phys.* **111**, 5514–5521.

[48] Vlachy, V., Ichiye, T. and Haymet, A.D.J. (1991) "Symmetric associating electrolytes: GCMC simulations and integral equation theory", *J. Am. Chem. Soc.* **113**, 1077–1082.

[49] Ichiye, T. and Haymet, A.D.J. (1990) "Integral equation theory of ionic solutions", *J. Chem. Phys.* **93**, 8954–8962.

[50] Certailler, T., Kunz, W., Turq, P. and Bellissent-Funel, M.C. (1990) "Lithium bromide in acetonitrile and water: a neutron scattering study", *J. Phys. Condens. Matter* **3**, 9511–9520.

[51] Robinson, R.A. and Stokes, R.H. (1959) *Electrolyte solutions* (Butterworths Scientific Publications, London).

[52] Pitzer, K.S. (1973) "Thermodynamics of electrolytes. 1. Theoretical basis and general equations", *J. Phys. Chem.* **77**, 268–277.

[53] Pitzer, K.S. (1991) *Activity coefficients in electrolyte solutions*, 2nd Ed. (CRC Press, Boca Raton).

[54] Barthel, J.M.G., Krienke, H. and Kunz, W. (1998) *Physical chemistry of electrolyte solutions* (Steinkopff/Springer, Darmstadt/New York).

[55] Allnatt, A.R. (1964) "Integral equations in ionic solution theory", *Mol. Phys.*, 8.

[56] Rasaiah, J.C. and Friedman, H.L. (1968) "Integral equation methods in the computation of equilibrium properties of ionic solutions", *J. Chem. Phys.*, 48.

[57] Rasaiah, J.C. and Friedman, H.L. (1969) "Integral equation computations for aqueous 1-1 electrolytes. Accuracy of the method", *J. Chem. Phys.*, 50.

[58] Carley, D.D. (1967) "Radial distributions of ions for a primitive model of an electrolyte solution", *J. Chem. Phys.*, 46.

[59] Rossky, P.J., Dudowicz, J.B., Tembe, B.L. and Friedman, H.L. (1980) "Ionic association in model 2-2 electrolyte solutions", *J. Chem. Phys.* **73**, 3372–3383.

[60] Bacquet, R. and Rossky, P.J. (1983) "Corrections to the HNC equation for associating electrolytes", *J. Chem. Phys.* **79**, 1419–1426.

[61] Rasaiah, J.C. (1972) "Computations for higher valence electrolytes in the restricted primitive model", *J. Chem. Phys.*, 56.

[62] Babu, C.S. and Ichiye, T. (1994) "New integral equation theory for primitive model ionic liquids: From electrolytes to molten salts", *J. Chem. Phys.* **100**, 9147–9155.

[63] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", *J. Am. Chem. Soc.* **117**, 5179–5197.

[64] Lindahl, E., Hess, B. and van der Spoel, D. (2001) "GROMACS 3.0: a package for molecular simulation and trajectory analysis", *J. Mol. Mod.* **7**, 306–317.

[65] Brooks, B.R., Bruccoleri, R.E., Olafson, B.A., States, D.J., Swaminathan, A. and Karplus, M. (1983) "Charmm: A program for macromolecular energy, minimization, and dynamics calculations", *J. Comput. Chem.* **4**, 187–217.

[66] MacKerell, Jr., A.D., Brooks, B., Brooks, III, C.L., Nilsson, L., Roux, B., Won, Y. and Karplus, M. (1998) "CHARMM: The energy function and its parameterization with an overview of the program", *The Encyclopedia of Computational Chemistry*, 1st Ed. (John Wiley & Sons, Chichester), pp 271–277.

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

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

[69] Heinzinger, K. (1985) "Computer simulations of aqueous electrolyte solutions", *Physica B* **131**, 196–216.

[70] Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) "Electrostatics of nanosystems: application to microtubules and the ribosome", *Proc. Natl Acad. Sci.* **98**, 10037–10041.

[71] Matsumoto, M. and Nishimura, T. (1998) "Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator", *ACM Trans. Mod. Com. Sim.* **8**, 3–30.

[72] Holst, M. (2001) "Adaptive numerical treatment of elliptic systems on manifolds: Theory and implementation in the finite element package MC", *Adv. Comput. Math.*, Submitted.

[73] Ramanathan, P.S. and Friedman, H.L. (1971) "Study of a refined model for aqueous 1–1 electrolytes", *J. Chem. Phys.* **54**, 1086–1099.

[74] Debye, P. and Hückel, E. (1923) "Zur theorie der elektrolyte i", *Phys. Z* **24**, 185–206.

[75] Chattopadhyaya, R., Meador, W.E., Means, A.R. and Quiocho, F.A. (1992) "Calmodulin structure refined at 1.7 Å resolution", *J. Mol. Biol.* **228**, 1177–1192.

[76] Babu, Y.S., Bugg, C.E. and Cook, W.J. (1988) "Structure of calmodulin refined at 2.2 Å resolution", *J. Mol. Biol.* **204**, 191–204.

[77] Wilson, M.A. and Brunger, A.T. (2000) "The 1.0 Å crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity", *J. Mol. Biol.* **301**, 1237–1256.

[78] Kretsinger, R.H. and Nockolds, C.E. (1973) "Carp muscle calcium-binding protein", *J. Biol. Chem.* **248**, 3313–3326.

[79] Kretsinger, R.H. (1975) "Hypothesis: Calcium modulated proteins contain EF-hands", In: Carafoli, E., ed, *Calcium transport in contraction and secretion* (North-Holland Publishing, Amsterdam).

[80] Nielsen, J.E., Andersen, K.V., Honig, B., Hooft, R.W., Klebe, G., Vriend, G. and Wade, R.C. (1999) "Improving macromolecular electrostatics calculations", *Protein Eng.* **12**, 657–662.

[81] Nayal, M. and DiCera, E. (1994) "Predicting Ca^{2+}-binding sites in proteins", *Proc. Natl Acad. Sci. USA* **91**, 817–821.

[82] Nayal, M. and DiCera, E. (1994) "Predicting Ca^{++}-binding sites in proteins", *Biophys. J.* **66**(part 2), A62–A62.

[83] Jayaram, B. and Beveridge, D.L. (1996) "Modeling DNA in aqueous solutions: theoretical and computer simulation studies on the ion atmosphere of DNA", *Annu. Rev. Biophys. Biomol. Struct.* **25**, 367–394.

[84] Manning, G.S. (1978) "The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides", *Q. Rev. Biophys.* **2**, 179–246.

[85] Le Bret, M. and Zimm, B.H. (1985) "Ions about a cylindrical polyelectrolyte", *Biopolymers* **23**, 271–285.

[86] Gordon, H. and Goldman, S. (1992) "Simulations on the counterion and solvent distribution functions around two simple models of a polyelectrolyte", *J. Phys. Chem.* **96**, 1921–1932.

[87] Murthy, C.S., Bacquet, R.J. and Rossky, P.J. (1985) "Ionic distributions near polyelectrolytes. A comparison of theoretical approaches", *J. Phys. Chem.* **89**, 701–710.

[88] Mills, P., Anderson, C.F. and Record, M.T. Jr., (1985) "Monte Carlo studies of counterion-DNA interactions. Comparison of the radial distribution of counterions with predictions of other polyelectrolyte theories", *J. Phys. Chem.* **89**, 3984–3994.

[89] Young, M.A., Jayaram, B. and Beveridge, D.L. (1997) "Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets", *J. Am. Chem. Soc.* **119**, 59–69.

[90] Feig, M. and Pettitt, B.M. (1999) "Sodium and chlorine ions as part of the DNA solvation shell", *Biophys. J.* **77**, 1769–1781.

[91] Kielkopf, C.L., Ding, S., Kuhn, P. and Rees, C. (2000) "Conformational flexibility of B-DNA at 0.74 Å resolution: $d(CCAGTACTGG)_2$", *J. Mol. Biol.* **296**, 787–801.