

Reproducible Estimation of Osmotic Coefficients Using the Inverse Monte Carlo Method

Jaydeep P. Bardhan,¹ Gary K. Leaf,² and Bob Eisenberg^{1,2}

*¹Department of Molecular Biophysics and Physiology
Rush University, Chicago, IL 60612, USA*

*²Mathematics and Computer Science Division
Argonne National Laboratory, Argonne, IL 60439, USA*

Abstract

Osmotic coefficients of solutions represent experimentally observable measures of the interactions between dissolved species and are particularly important metrics for biophysical experiments in which the concentrations of solutes control biological function, as in ion-channel proteins. The computational expense associated with all-atom molecular-dynamics (MD) simulations makes it difficult to connect these data to detailed molecular models directly. It is therefore generally impractical to ensure that MD force fields are parameterized consistently with the full range of available data. Several approaches to coarse graining these systems make the estimation of osmotic coefficients more tractable. In this paper we explore one such approach, the inverse Monte Carlo (IMC) method of Lyubartsev and Laaksonen, for estimating osmotic coefficients of aqueous solutions with dissolved sodium chloride. The IMC method is used to calculate effective pairwise potential functions for the ion-ion interactions, taking as input the ion-ion radial distribution functions calculated from all-atom MD simulations. Our results show that the IMC method converges robustly and reproducibly, and we conclude that IMC and related approaches hold promise for validating force fields against experimental data, though further refinements will be important to ensure practical viability for use in parameterization. In particular, we have found that the estimates of the osmotic coefficient, which is an intensive quantity, are size dependent, but the size dependence can be corrected with a simple linear fitting procedure. Also, the Monte Carlo procedure explores the phase space extremely slowly, and it may be valuable to use accelerated sampling procedures; performance may also be improved significantly by using more sophisticated Newton–Raphson methods. It is well known that osmotic coefficients are highly sensitive to the pair potentials involved, and thus methods for their estimation must be studied and validated carefully.

1. FINITE-SIZE CORRECTION

Some time after writing this manuscript, we found that the osmotic pressure P_{osm} , defined by the virial expression

$$P_{\text{osm}} = \frac{NkT}{V} - \left\langle \frac{\partial U}{\partial r_{ij}} r_{ij} \right\rangle \quad (1)$$

where the ensemble average involving the pair potential $U(r_{ij})$ incorporates the volume-dependence of the Ewald expression¹, was usually not just a linear function of the number of particles, but affine. That is, the osmotic coefficient $\phi = P_{\text{osm}}V/NkT$, which is an intensive quantity, instead has a size dependence that looks like

$$\phi_{SD}(N) = \phi + \frac{B}{N}, \quad (2)$$

where $\phi_{SD}(N)$ denotes the (size-dependent) estimate obtained from a simulation with N particles and B is the offset. The second term obviously decreases in magnitude as the Monte Carlo (MC) simulation uses increasing numbers of particles. Thus, ultimately one recovers the correct osmotic coefficient as the number of particles grows. We investigated the causes of this size dependence and attempted to find ways to eliminate it. We found that B is dependent on the particular inverse Monte Carlo (IMC) protocol used, including truncation parameters associated with the Ewald summation (as would be expected), the number of MC steps at each iteration, and the criterion used to choose a final IMC solution. More work will be needed to establish definitive protocols for removing this size dependence. This report does not describe details regarding the new finding.

2. INTRODUCTION

Aqueous ionic solutions represent an ongoing modeling challenge for theoretical and computational scientists², with substantial motivation provided by problems throughout biology and chemistry³. Gradients of ionic solutions are the energy source of many of life's functions, from signaling in the nervous system to vital parts of oxidative phosphorylation in mitochondria^{4,5} and photosynthesis in chloroplasts⁶. Ions carry the charge and current responsible for the electrical properties of neurons and muscle fibers, which (for example) allow the heart to function as a pump⁷. Ions can activate and deactivate proteins⁸ and act as messengers carrying signals that control a wide range of biochemical pathways⁹. One can hardly find

a biological function that does not involve ions, and so chemists have long dreamed that a physical theory of ions near and in proteins could provide decisive help in understanding biological function^{10–17}.

In physical chemistry, aqueous electrolyte solutions have been studied experimentally and theoretically for more than a century (see historical Refs. 2,18–26). The *osmotic coefficient* of a solution is a thermodynamic variable that captures a sense of the solution’s deviation from ideality—that is, how strongly the solute ions interact with one another. The osmotic coefficients of myriad solutions have been tabulated for decades^{27–31}, with aqueous ionic solutions representing a significant subset, although many other solutes and solvents have been studied. For example, osmotic coefficients have been measured for solutions of amino acids and short peptides³² as well as monosaccharides³³.

For experimental studies of ion-channel proteins, the measurement of osmotic coefficients (or their thermodynamic equivalents, the activity coefficients) of the solutions on both sides of the membrane is a critical step in verifying that the desired ionic mixtures have been achieved. Furthermore, a characteristic signature of ion-channel proteins is the reversal potential (the electrical potential at which the current is zero³⁴), which depends sensitively on the electrochemical potential gradients created by the different ionic compositions and therefore on the osmotic coefficients. Complementary to the thermodynamics of aqueous solutions, experimental methods such as neutron and x-ray scattering provide substantial structural information about the solvent molecules surrounding the ions and about the distributions of the ions themselves (see, e.g. Refs. 35–37). However, it is still difficult to obtain significant information beyond the first few nearest neighbors, and the thermodynamics can be sensitive functions of the structural details³⁸.

Rapid growth in computer power and the advent of all-atom explicit-solvent molecular dynamics (MD) simulation methods have been enabled by programs such as NAMD³⁹, CHARMM⁴⁰, and AMBER⁴¹, augmented with either non-polarizable^{42–46} or, in some cases, polarizable^{47–51} force fields. Together these technologies bring unprecedented possibilities for studying molecular phenomena such as ion solvation^{43,52–54}. However, fully atomistic simulations present challenges because of the large number of parameters that must be self-consistently determined prior to simulation⁵⁵. Many widely used force fields have been designed so that solvation free energies of particular solutes match experiment. The solvation free energy is a conceptually simple quantity and therefore an appealing metric for

parameterization, but unfortunately this quantity can be difficult to measure accurately⁵⁶. Shortcomings in current force fields have been pointed out^{57,58}, providing more motivation to improve and verify parameters. In a similar spirit of parameterizing ionic solutions on the basis of systematic measurements, Smith and collaborators have parameterized an atomistic force field using Kirkwood–Buff (KB) integrals^{59,60}. The KB force field has met with remarkable success, but the necessary experiment measurements can be difficult to obtain.

In contrast to fully atomistic methods, models with few parameters such as the primitive model of electrolytes^{61–63} and statistical-mechanical integral equations based on the Ornstein–Zernike equation have proven to be highly successful in predicting the osmotic coefficients of ionic solutions^{61,62,64,65}. The success of simple, reduced models in predicting osmotic coefficients makes it imperative that more detailed MD models predict these quantities at least as well. Unfortunately, the computational expense associated with estimating osmotic and activity coefficients has limited the opportunity for their use; the difficulty of calculating activity coefficients was reviewed several years ago by Lazaridis and Paulaitis⁶⁶. The rapid growth in available computing resources since then has greatly alleviated this constraint and led to a recent upswing in publications detailing calculations of osmotic and activity coefficients⁶⁷.

Much recent theoretical work has explored estimates of osmotic coefficients of salt solutions, most often sodium chloride^{56,68,69}. However, numerous studies have been performed on larger systems. Druchok et al. have studied protein solutions⁷⁰. Yu et al. have studied DNA-electrolyte solutions using classical density functional theory (DFT)⁷¹, and Li and Wu⁷² have also used classical DFT to study asymmetric electrolytes. Hansen, Podgornik, and Parsegian⁷³ have shown the inadequacy of simple charge-condensation theories to predict the osmotic coefficients of solutions of B-DNA. Binding affinities can also be analyzed by using activity coefficients⁷⁴, and therefore methods for estimating quantities such as osmotic or activity coefficients may be useful in studying molecular binding.

Estimating the osmotic coefficient of an aqueous electrolyte requires large numbers of ions and extensive sampling of phase space^{56,68}. Direct MD is feasible, as demonstrated recently by Kalcher and Dzubiella⁶⁹, though it is extremely computationally intensive. An alternative is to estimate *effective potentials* that can then be used in reduced-model simulations such as Monte Carlo (MC) methods^{68,75} or stochastic dynamics⁵⁶, which account implicitly for the solvent. These effective potentials arise when one replaces the original Hamiltonian with

a reduced Hamiltonian by integrating out the solvent degrees of freedom (DOF)^{75,76}. Even if the original Hamiltonian is pairwise additive, however, the reduced Hamiltonian will not be, in general. That is, by integrating out the solvent DOF, one introduces new three-body potentials, new four-body potentials, and so forth, up to the full N-body potential. Because the computational work required to determine and to employ the non-pairwise potentials can be prohibitive, the reduced Hamiltonian is often approximated instead as a sum of effective pairwise potentials.

Lyubartsev and Laaksonen presented an inverse Monte Carlo (IMC) algorithm for finding effective pairwise potentials (EPPs) given the radial distribution functions between the ionic species^{68,75,77,78}. This is a computational approach to solving the inverse problem in which many-body effects are incorporated directly into the pairwise potential functions. That is, the calculated EPPs are not actually the potentials of mean force at infinite dilution, which would be the actual pairwise potentials from the reduced Hamiltonian. The IMC approach is therefore one estimator for the thermodynamics of the solution. This approach is similar⁷⁹ to hypernetted chain-based approaches^{80,81}. In contrast, Hess et al. explicitly determine the ion-ion potentials of mean force at (effectively) infinite dilution, and include many-body effects (that is, the effects of the other ions for finite concentration simulations) using a concentration-dependent dielectric constant⁵⁶. Kalcher and Dzubiella found empirically that MD-generated potentials of mean force at different concentrations could be separated into a short-range potential that was concentration independent and a (concentration-dependent) Debye–Hückel long-range interaction to describe screening⁶⁹. Several other types of coarse-graining methods have been reported recently, most notably the force-matching approaches of Parrinello, Voth, Izvekov, and collaborators^{82–84}.

In this paper, we focus on the robustness of the IMC approach for estimating osmotic coefficients, with a primary goal of establishing the parameters required to obtain converged estimates of osmotic coefficients from MD simulations. This assessment seems appropriate given the enormous growth in computing power since the original IMC work was presented, along with results suggesting size-independence and convergence requirements^{68,75}. Furthermore, a detailed reproducibility analysis of much larger systems for the MD, MC, and IMC combination appears timely given the increasing interest in the use of coarse-graining methods as an approach to force-field validation. Another interest is the development of reduced-model approaches that can include solvent-packing effects for inhomogeneous sys-

tems such as ion-channel proteins, where the primitive model of electrolytes has proven surprisingly successful at describing selectivity^{85–89} despite the primitive model’s neglect of the seemingly important structural and chemical details.

We explore the effects of the Monte Carlo parameters, such as the number of steps and the number of ions employed, as well as the IMC parameters themselves, such as the number of IMC iterations. We obtain for 4 M NaCl solution an estimate of the osmotic coefficient that appears to have an uncertainty of approximately 0.03, or a few percent of a typical experimental value between 0.80 and 1.20; however, this level of convergence should be sufficient to illustrate whether a given MD force field captures essential features of the osmotic coefficient as a function of salt concentration.

The following section describes the inverse Monte Carlo method for deriving effective interaction potentials from MD simulations at finite concentrations, as well as the estimation of osmotic coefficients from Monte Carlo simulations. Section 4 presents the results of our calculations, establishing simulation parameters under which the IMC-calculated osmotic coefficients appear to be converged. Section 5 concludes the paper with a discussion.

3. THEORY

3.1. Inverse Monte Carlo Algorithm

The central theoretical basis for the IMC method is due to R. Henderson, who showed that for any set of radial distribution functions (RDFs), a set of pairwise potentials that reproduces the RDFs must be unique⁹⁰. Chayes, Chayes, and Lieb later supplied the proof that such a set of potentials always exists^{91,92}. Lyubartsev and Laaksonen proved that these effective pair potentials (EPPs) maximize entropy over all reduced energy functions⁶⁸. The IMC algorithm is essentially a Newton–Raphson method for determining the EPPs associated with a given set of RDFs, which in this work we take from molecular dynamics simulation. The utility of the relation between EPPs and RDFs also depends on practical issues such as sensitivity to error in the input data, as do all inverse problems^{93,94}. We investigate some of those here as we compute and use EPPs and RDFs.

For simplicity of notation we describe the IMC process in the context of a single-component system for which we know a priori that the particle-particle RDF is $g(r)$, where

r is the Euclidean distance between particles. This model assumes that the particles are interacting in bulk; that is, we are studying the homogeneous fluid, and the interaction between two particles does not depend on their locations but only on the relative distance between them. The Hamiltonian associated with the single EPP $\Psi(r)$ may be written

$$H = \sum_{i,k} \Psi(|q_i - q_k|), \quad (3)$$

where q_i denotes the position of the i th particle and the sum is taken over all (nonidentical) pairs⁷⁵. In the multicomponent case, such as in studying sodium and chloride, each type of interaction has its own EPP.

The IMC software of Lyubartsev approximates an EPP as a weighted combination of piecewise-constant basis functions⁷⁵; we discuss the advantages and disadvantages of this approach, as well as other possible representations, in Section 5. Representing the EPP as a scaled sum allows the single-component Hamiltonian to be written as

$$H = \sum_l K_l S_l(q_1, q_2, \dots), \quad (4)$$

where term l of the sum is associated with the scaling constant (K_l) and a function of the particle positions, $S_l(\{q_i\})$. In the present implementation of IMC⁷⁵, the basis function $S_l(\{q_i\})$ captures the number of particle pairs separated by a distance r satisfying

$$l\Delta r < r < (l + 1)\Delta r, \quad (5)$$

where Δr is the width of the piecewise-constant basis functions. Solving the inverse problem—that is, finding the correct values of K_l —requires that the MC calculation estimates average counts at a distance r , which we denote by $\langle S(r) \rangle$, proportional to the radial distribution function $g(r)$:

$$\langle S(r) \rangle = 4\pi r^2 g(r). \quad (6)$$

In practice, these conditions are enforced by ensuring the pair counts are matched in the bins defined by Eq. 5.

The iterative Newton–Raphson procedure for calculating the zero of a vector-valued function of several variables $f(x)$ proceeds as follows. In the IMC problem, we seek to find the zero of the function that maps the weights K_l to the difference between the reference and

computed RDFs. Writing the iterate at the k th iteration as x^k , one calculates the function value $f(x^k)$ and the Jacobian of f at x^k , J_f . The update Δx^k is calculated according to

$$J_f \Delta x^k = -f(x^k), \quad (7)$$

and the next iterate is set to

$$x^{k+1} = x^k + \gamma \Delta x^k, \quad (8)$$

where γ can be set to one for all iterations. One can also set γ to a smaller value to improve robustness and trade off the rate of convergence⁷⁵. For the IMC problem, the initial guess for the weights K_l is taken to be the sampled potential of mean force:

$$K_l^0 = -k_B T \ln g(l\Delta r). \quad (9)$$

At every iteration a Monte Carlo simulation is performed to calculate the RDFs by using the current set of weights (i.e., the current EPPs), from which the error relative to the reference RDFs is easily computed to provide the right-hand side of Eq. 7. Using piecewise-constant basis functions, the i, j entry of the Jacobian matrix is defined by the statistical-mechanical relationship⁷⁵

$$J_{i,j} = -\beta (\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle), \quad (10)$$

where $\beta^{-1} = k_B T$ and one averages over all of the sampled MC configurations. Note that other types of basis functions can also be used to approximate the effective pair potentials⁹⁵.

Lyubartsev's implementation of IMC represents the short-range potential (the potential at particle separations $r < r_c$ for some cutoff distance r_c using a weighted combination of basis functions as described above. The long-range potential for $r > r_c$ is represented with a Coulombic potential, where the dielectric constant used to scale the $1/r$ potential is an input to the IMC software⁷⁵. The IMC software uses Ewald summation to evaluate these potentials.

3.2. Osmotic Coefficient Definition and Estimation

Kalcher and Dzubiella have discussed the two approaches to estimation of ϕ : the *compressibility* approach and the *virial* approach⁶⁹. Here we use the virial method, in which the

use of concentration-independent effective potentials allows ϕ to be written as

$$\phi = 1 - \frac{Nk_B T}{3} \left\langle \sum_{i,j} \frac{\partial U}{\partial r_{ij}} r_{ij} \right\rangle, \quad (11)$$

where N is the number of particles (here, ions), k_B is the Boltzmann constant, and T is the temperature. Because the total effective potential is written as a sum of short-range and long-range potentials, the virial can be decomposed easily into short-range and long-range contributions^{68,69}. Hummer et al. have described a correction to the virial estimation procedure for Ewald summation methods¹, which reduces the dependence of estimated results on the system size. The IMC code incorporates this correction, and therefore the virial estimates should not suffer from this size-dependence issue.

The average in Eq. 11 is taken over all the MC snapshots saved over the entire IMC process, rather than just the snapshots saved during the current Newton iteration. Thus, at each successive iteration, the osmotic coefficient is estimated more accurately (with respect to the EPP at the iteration in question).

4. RESULTS

Our results are organized as follows. We first establish the parameters needed to ensure that the Monte Carlo calculations are reasonably converged. We then demonstrate that the IMC software, given essentially identical inputs, generates highly reproducible results but that one should assess possible pitfalls in the MC procedure by running an ensemble of IMC calculations. This is an important factor in considering the use of a Newton–Raphson method driven by stochastic simulation, particularly when the inputs are necessarily noisy RDFs taken from MD simulations. We also present results that suggest the origin of the slow convergence of the IMC procedure. In the final set of results we explore the convergence properties of the IMC approach for RDFs computed from independent MD trajectories.

4.1. Monte Carlo Parameters

Three user-specified parameters govern the Monte Carlo procedure that drives the IMC process and is used to estimate the osmotic coefficient: the system size (that is, the number of ions), the number of Monte Carlo steps used in the equilibration phase, and the number

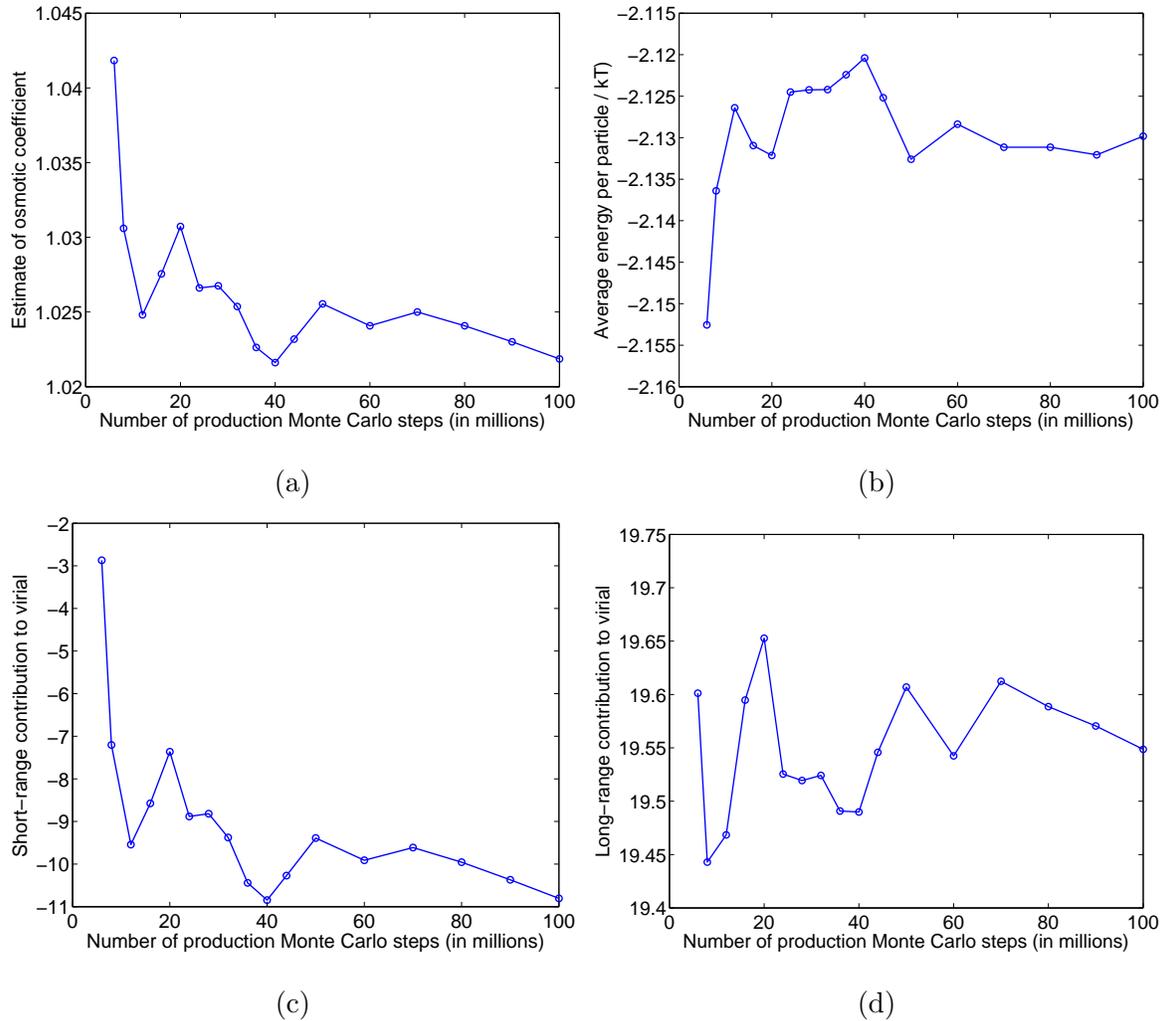


FIG. 1: Convergence of quantities derived from EPP-driven Monte Carlo simulation, as a function of the number of production MC steps attempted, using 200 ion pairs and attempting 8 million equilibration steps. See text for details on the EPPs used. Symbols denote where measurements were taken. Lines are included only as an aid to visualize trends. (a) Osmotic coefficient estimate. (b) Average energy per particle. (c) Short-range contribution to the virial. (d) Long-range contribution to the virial.

of steps in the production phase. The size of the simulation box used in the MC procedure is fixed by the combination of the desired salt concentration and the number of ions to be used. To establish the parameters required to converge the osmotic coefficient, we used an effective pair potential obtained from 50 iterations of IMC using 18 salts, following Lyubartsev and Laaksonen⁷⁵. The input RDFs for this simulation were those generated previously⁷⁵. Our assumption in using this EPP is that the MC process will exhibit essentially the same

convergence behavior for all reasonable pair potentials.

Figure 1(a) is a plot of the osmotic coefficients estimated after varying numbers of production MC steps, using 200 ion pairs and 8 million equilibration steps. The osmotic coefficient appears to oscillate between 1.02 and 1.025 after 30 million iterations. Similar requirements for the number of iterations were observed by Abbas et al.⁶⁵, who used 95 million iterations in their primitive-model calculations, which used 300 ion pairs. Figure 1(b) is a plot of the average energy per particle as a function of the number of MC steps. This plot suggests that at least 50 million iterations are required for convergence. Figures 1(c) and (d) are plots of the short-range and long-range contributions to the osmotic coefficient as more MC steps are employed. Note that the scales on the ordinate axes are of very different magnitudes; the short-range component clearly converges much more slowly than the long-range component, and the variation in the osmotic coefficient is dominated by the short-range component's slower convergence and larger-magnitude changes.

The osmotic coefficient is estimated by taking snapshots of the RDFs every 50 attempted MC steps; sampling as frequently as every 8 attempted steps, so as to increase the sampling (the number of snapshots) by a factor of 6, did not improve the stability of the estimated osmotic coefficients (by stability, we mean an empirical observation of convergence rather than the definition of stability employed in numerical analysis). The fact that more frequent snapshots do not reduce the number of snapshots required to converge the osmotic coefficient suggests that the limiting factor in convergence is not necessarily the number of samples but rather the rate at which the MC procedure explores phase space by moving a single ion at each attempted step. The slow convergence of the osmotic coefficient therefore immediately presents a practical difficulty, that of computational expense, and suggests that a worthwhile development in future work would be a more rapidly convergent MC procedure.

4.2. Inverse Monte Carlo Parameters

We now illustrate the basic convergence behavior of the IMC procedure. Figure 2 shows the osmotic coefficients estimated after each of 100 IMC iterations for two IMC realizations, one of which used 18 ion pairs and the other 75 pairs. The RDFs used as input for these IMC realizations are from Lyubartsev and Laaksonen⁶⁸, which are from an MD simulation of 4 M NaCl using Smith–Dang ions^{43,44} and flexible SPC water⁹⁶. Three important details

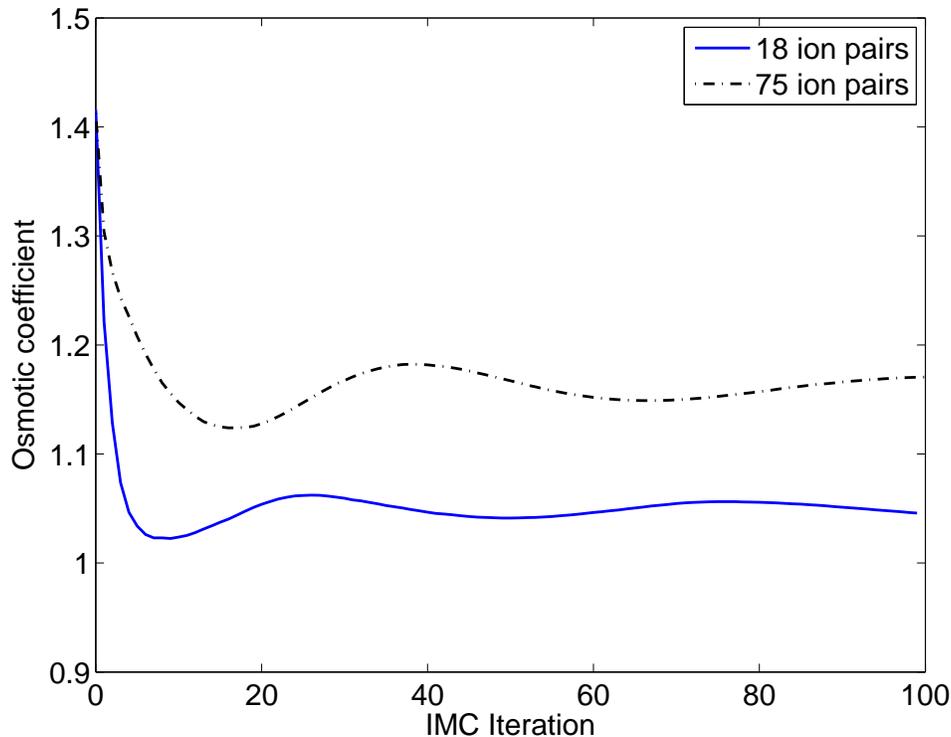


FIG. 2: Convergence of two realizations of IMC, one using 18 ion pairs and one using 75 ion pairs. Both osmotic coefficients vary nonmonotonically with increasing iteration number, and converge to different final estimates, which suggests a size dependence in the IMC procedure.

are evident: both realizations converge; the realizations converge to different values; the convergence is nonmonotonic. For both realizations, the number of production MC iterations at each step was 6 million, and the number of equilibration steps was 10^5 . Because the estimation of the osmotic coefficient actually uses all the snapshots taken during the IMC realization, estimates of the osmotic coefficient accurately reflect the current set of EPPs after the first six or seven iterations.

Figure 3 shows the size dependence of the IMC process, visible in Figure 2 as the IMC simulations appear to be converging to EPPs with different estimates of the osmotic coefficient. Since the osmotic coefficient is an intensive quantity, it is problematic that the number of ions used in the MC simulation at each IMC iteration can have a significant impact.

Table 1 illustrates these dependencies in more detail. The rightmost column indicates the IMC iteration at which the IMC realization reached its first minimum; quantities tabulated in the other columns, such as osmotic estimates $\phi_{SD}(N)$ and the average energy per ion,

TABLE I: Dependence of IMC-computed quantities on the number of ions in the MC simulations.

Ion Pairs	Osmotic Coeff.	Energy / NOP	RDF Error ($\times 10^{-3}$)	-VIRS/NOP	-VIRE/NOP	Iteration
18	1.034850	-2.2638	2.17	-0.0283	0.0632	7
25	1.071426	-1.9234	1.96	0.0098	0.0616	7
36	1.133333	-1.7235	2.44	0.0734	0.0599	7
50	1.138834	-1.6410	3.56	0.0810	0.0579	8
75	1.165253	-1.5656	6.07	0.1089	0.0563	9
100	1.166433	-1.5254	9.26	0.1109	0.0555	9
150	1.178856	-1.5111	14.15	0.1238	0.0550	13
200	1.182512	-1.4626	23.22	0.1269	0.0556	9
225	1.176065	-1.5036	18.74	0.1214	0.0547	17
250	1.188518	-1.4702	24.70	0.1330	0.0555	9
275	1.199162	-1.4429	32.51	0.1437	0.0554	9
300	1.183036	-1.4800	15.99	0.1272	0.0558	9
325	1.177499	-1.4789	17.33	0.1219	0.0556	9
350	1.168518	-1.4776	20.35	0.1127	0.0558	9

have been calculated from the effective pair potentials at that particular iteration. Figure 3 is a plot of the osmotic coefficient estimate as a function of the number of ion pairs. Clearly, in estimating osmotic coefficients, it is important to thoroughly assess size-dependence effects. The number of ions appears to be a principal factor in the estimation of the osmotic coefficient—that is, regardless of the number of MC steps or the number of IMC iterations, the EPPs calculated by using 18 or 200 ion pairs will give rise to substantially different osmotic coefficients.

Tables 2, 3, and 4 contain structural information on the EPPs at the IMC iterations listed in Table 1 and illustrate that structural details of the EPPs, such as the location and depth of wells, are well converged even for relatively small numbers of ion pairs, even though the estimate of the osmotic coefficient is still changing by experimentally significant amounts.

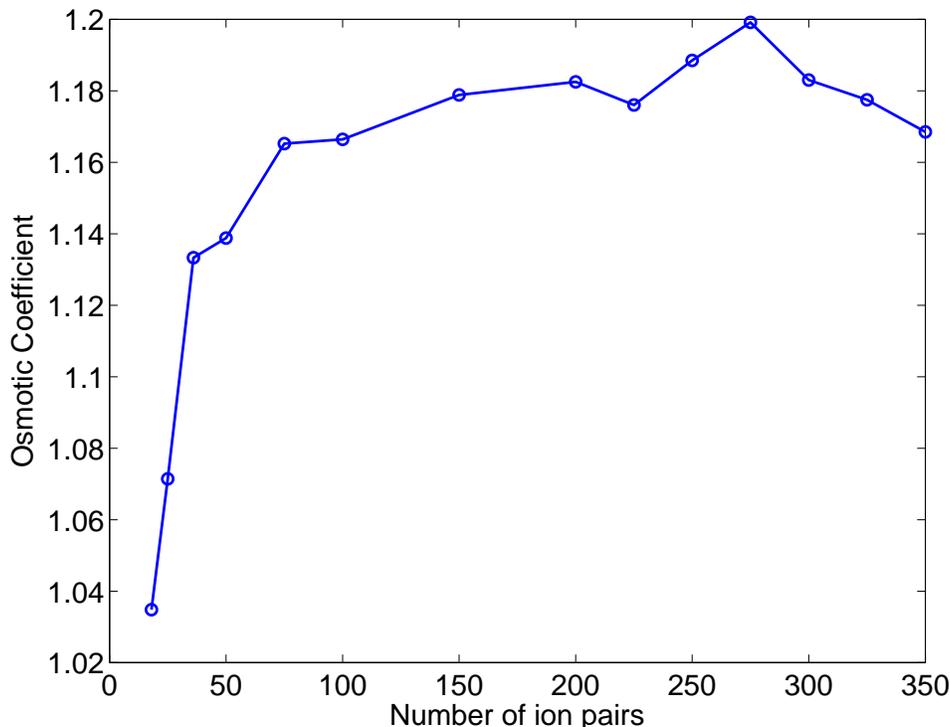


FIG. 3: Variation of the osmotic coefficient estimated from IMC using different numbers of ion pairs. For small numbers of particles the answer deviates substantially from the estimated 1.16–1.20 observed in larger simulations. See text for details of the IMC procedures used.

4.3. Reproducibility Study with Identical Input

The IMC process, like other reverse Monte Carlo methods^{97,98}, is essentially a stochastically driven optimization algorithm, with the input RDFs taken from a necessarily finite sample (i.e., the MD simulation). Consequently, one must establish the magnitude of errors introduced by stochasticity. In a reproducibility experiment, ten independent realizations of IMC were computed. Each used exactly the input RDFs from Lyubartsev and Laaksonen, with only the seed value for the random number generator varied between IMC realizations. The IMC realizations were terminated at 100 iterations and performed by using an MC box of length 19.4 Å and a cutoff of 9.7 Å. The MC simulations in these IMC processes were conducted by using 18 ion pairs and 6×10^6 production Monte Carlo steps. The default IMC dielectric constant of 78 was left unchanged.

The average osmotic coefficient over the 10 realizations was 0.726, with a standard deviation of 5.34×10^{-3} . The average 2-norm error from the reference RDFs was 1.38×10^{-3} ,

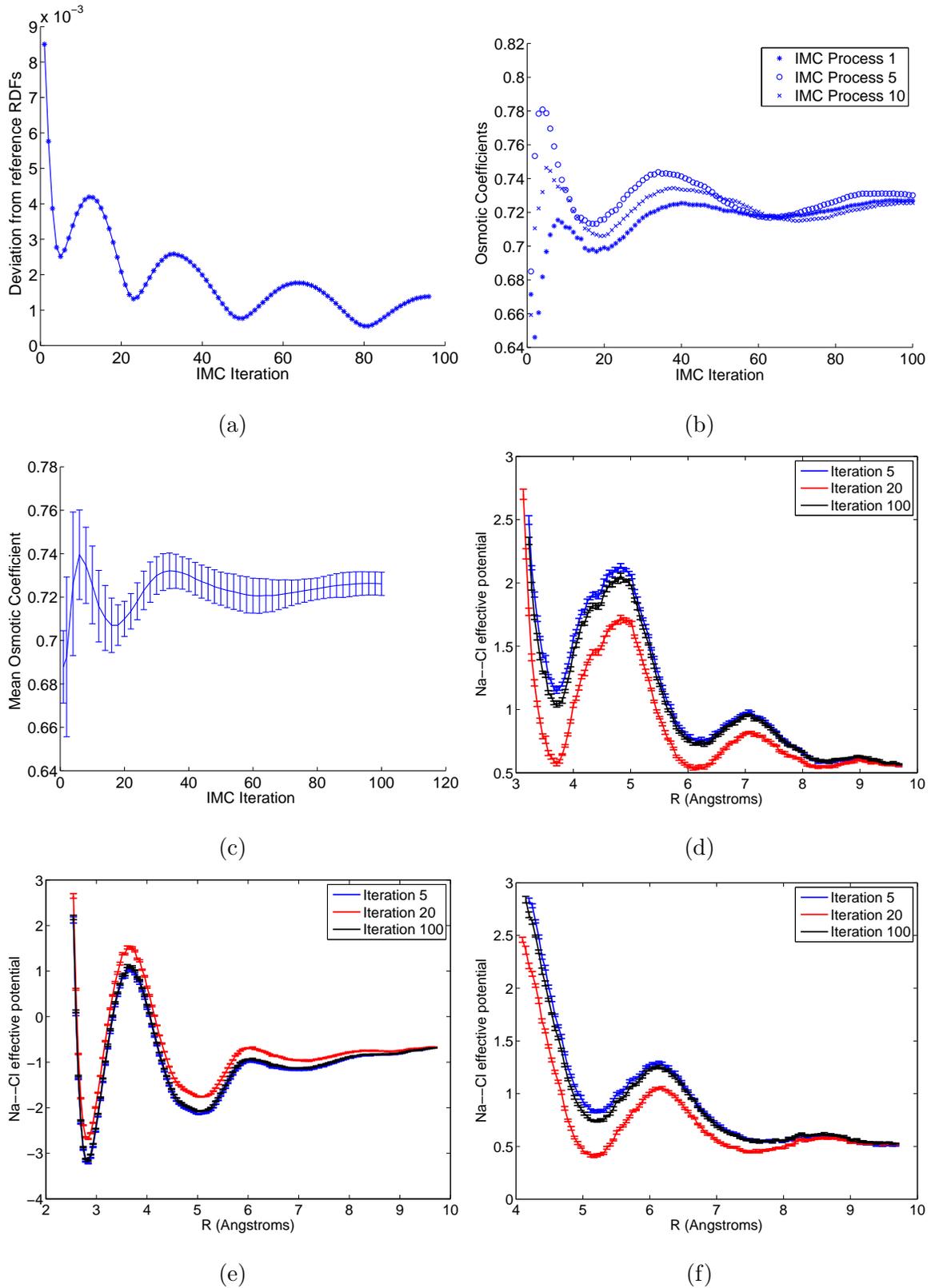


FIG. 4: Reproducibility of the IMC procedure given identical RDF inputs from Lyubartsev and Laaksonen⁶⁸. See text for details about the IMC parameters used. (a) Typical convergence behavior for a single IMC realization, measured according to the squared-norm difference between the reference (MD) RDFs and the MC-determined RDFs. (b) Osmotic coefficients as a function of IMC iteration number for three representative realizations. All exhibit qualitatively the same behavior, although different oscillation periods and amplitudes, which will

TABLE II: Structural details of the Na-Na effective pair potentials, determined at the first minimum in the deviation between reference radial distribution functions (RDFs) and the RDFs generated from the EPP-based Monte Carlo simulations. The first column contains the number of ion pairs. The second and third columns represent the location and potential at the first potential well; the fourth and fifth columns represent the location and potential at the first potential barrier. The sixth column is the depth of the first well as measured from the well-potential value (second column) to the barrier potential (fourth column). All locations are in angstroms, and all potentials are in $k_B T$.

No. Ion Pairs	1st Well (Loc.)	1st Well (Val.)	1st Barrier (Loc.)	1st Barrier (Val.)	Well Depth
18	3.710	1.1802	4.826	2.1426	0.9653
25	3.710	1.1429	4.826	2.1447	1.0018
36	3.710	1.1071	4.826	2.1126	1.0055
50	3.710	1.0687	4.826	2.1216	1.0529
75	3.710	1.0555	4.826	2.0824	1.0269
100	3.710	1.0243	4.826	2.0633	1.0390
150	3.710	1.0047	4.826	2.0479	1.0431
200	3.710	1.0232	4.826	2.0575	1.0432
225	3.710	0.9751	4.826	2.0137	1.0386
250	3.710	0.9925	4.826	2.0404	1.0479
275	3.710	0.9942	4.826	2.0364	1.0422
300	3.710	1.0351	4.826	2.0455	1.0103
325	3.710	1.0168	4.826	2.0453	1.0285
350	3.710	0.9963	4.826	2.0634	1.0671

with standard deviation 9.66×10^{-5} . The scale of the standard deviation indicates that when independent IMC realizations are given exactly the same RDFs as input (here, the set of RDFs computed by Lyubartsev and Laaksonen⁷⁵), the IMC procedure generates highly reproducible results despite the use of different random seeds. Figure 4(a) contain plots of the 2-norm of the difference between the MC-generated RDFs and the reference (input) RDFs as a function of iteration, for one realization of the IMC procedure. All realizations exhibited the same qualitative behavior; the nonmonotone convergence is analyzed in more

TABLE III: Structural details of the Na-Cl effective pair potentials, determined at the first minimum in the deviation between reference radial distribution functions (RDFs) and the RDFs generated from the EPP-based Monte Carlo simulations. The first column contains the number of ion pairs. The second and third columns represent the location and potential at the first potential well; the fourth and fifth columns represent the location and potential at the first potential barrier. The sixth column is the depth of the first well as measured from the well-potential value (second column) to the barrier potential (fourth column). All locations are in angstroms, and all potentials are in $k_B T$.

No. Ion Pairs	1st Well (Loc.)	1st Well (Val.)	1st Barrier (Loc.)	1st Barrier (Val.)	Well Depth
18	2.837	-3.2758	3.613	1.0029	4.2787
25	2.837	-3.2237	3.613	1.0478	4.2715
36	2.837	-3.1620	3.710	1.1393	4.3013
50	2.837	-3.1396	3.613	1.1653	4.3049
75	2.837	-3.1142	3.662	1.1869	4.3011
100	2.837	-3.0888	3.662	1.2135	4.3023
150	2.837	-3.0234	3.662	1.2459	4.2693
200	2.837	-3.0404	3.662	1.2649	4.3053
225	2.837	-2.9866	3.662	1.2966	4.2832
250	2.837	-3.0269	3.662	1.2538	4.2807
275	2.837	-3.0100	3.662	1.2532	4.2632
300	2.837	-3.0599	3.662	1.2331	4.2929
325	2.837	-3.0206	3.662	1.2246	4.2452
350	2.837	-3.0307	3.662	1.2543	4.2850

detail in the following section. Figure 4(b) contains plots of the osmotic coefficients of three of the IMC realizations as a function of iteration. Figure 4(c) is a plot of the mean osmotic coefficient as a function of iteration, as well as the standard deviation. Thus, although the different IMC realizations may follow different paths, they do converge to the same EPPs (though not within a small number of iterations). Figures 4(d), (e), and (f) are plots of the mean effective pair potentials at iterations 5, 20, and 100, along with the standard deviations. We note that the standard deviations do not seem to decrease significantly with

TABLE IV: Structural details of the Cl-Cl effective pair potentials, determined at the first minimum in the deviation between reference radial distribution functions (RDFs) and the RDFs generated from the EPP-based Monte Carlo simulations. The first column contains the number of ion pairs. The second and third columns represent the location and potential at the first potential well; the fourth and fifth columns represent the location and potential at the first potential barrier. The sixth column is the depth of the first well as measured from the well-potential value (second column) to the barrier potential (fourth column). All locations are in angstroms, and all potentials are in $k_B T$.

No. Ion Pairs	1st Well (Loc.)	1st Well (Val.)	1st Barrier (Loc.)	1st Barrier (Val.)	Well Depth
18	5.165	0.8423	6.135	1.3321	0.4898
25	5.311	0.8541	6.135	1.3394	0.4853
36	5.165	0.8455	6.135	1.3397	0.4942
50	5.165	0.8588	6.135	1.3477	0.4889
75	5.165	0.8409	6.135	1.3439	0.5030
100	5.165	0.8186	6.135	1.3488	0.5302
150	5.165	0.8113	6.135	1.3265	0.5152
200	5.165	0.7811	6.135	1.3333	0.5521
225	5.165	0.7655	6.135	1.3020	0.5365
250	5.165	0.8057	6.135	1.3252	0.5195
275	5.165	0.8232	6.135	1.3345	0.5113
300	5.165	0.8106	6.135	1.3445	0.5339
325	5.165	0.8055	6.135	1.3369	0.5314
350	5.165	0.8214	6.135	1.3436	0.5222

increase in the number of IMC iterations.

The oscillations in the estimates of the osmotic coefficient and the deviation from the reference RDFs have different periods. The osmotic coefficient varies on a longer period than the deviation; and, for both metrics, the periods get longer with increasing iteration count and as one increases the number of ion pairs in the Monte Carlo simulations.

4.4. Analysis of the Nonmonotone Convergence

To analyze the convergence behavior more carefully, and in particular the presence of multiple minima in the RMS deviation between the reference and MC-calculated RDFs as a function of iteration, we examine the two IMC experiments of 100 iterations described previously, one using 18 ion pairs and the other using 75 ion pairs. Kalcher and Dzubiella⁶⁹ and others³⁸ have noted that the overall thermodynamics are not easily interpretable in terms of particular features of the EPP such as well depths, and our results support that view. Figure 5 is a plot of the RMS deviations as a function of iteration, and Figure 6 contains plots of the 18-pair and 75-pair EPPs at the first four minima (iterations 5, 20, 42, and 69 for the 18-pair problem and iterations 8, 29, 55, and 87 for the 75-pair problem). Note that running the IMC procedure until four minima in the deviation have been observed gives two intervals for bounding the osmotic estimates $\phi_{SD}(N)$. In all our experiments the bounds become tighter with more iterations.

The nonmonotone (oscillatory) convergence observed in Figure 5 suggests that the normal IMC procedure, which takes full Newton–Raphson steps at every IMC iteration, may not be ideal. In particular, nonmonotonicity suggests that the Jacobian changes significantly over the iterations, with some search directions (in the space of effective pair potentials) possessing small curvature. These directions with small curvature lead to the calculated step being larger than it would be if the full nonlinearity of the optimization problem were taken into account. Line-search methods that ensure monotonicity⁹⁹ may accelerate convergence; they present an interesting subject for future work.

We have examined the EPPs associated with the IMC iterations that are local minima in the 2-norm deviation between the reference RDF and the IMC RDF. The four anion–cation EPPs from the 18-ion-pair calculation are plotted in Figure 6(a). The EPPs at iterations 5 and 42 clearly are similar, as are those at iterations 20 and 69, with the differences between pairs much larger in magnitude than the differences between the members of each pair. The four anion-cation EPPs from the 75-pair iterations are plotted in Figure 6(b) and show a similar grouping.

The difference between the 75-pair EPPs at iterations 55 and 87 is plotted in Figure 6(c), as is the difference between the EPPs of iterations 20 and 69. Except at short distances, the search directions of small curvature appear to be associated with a slowly decaying

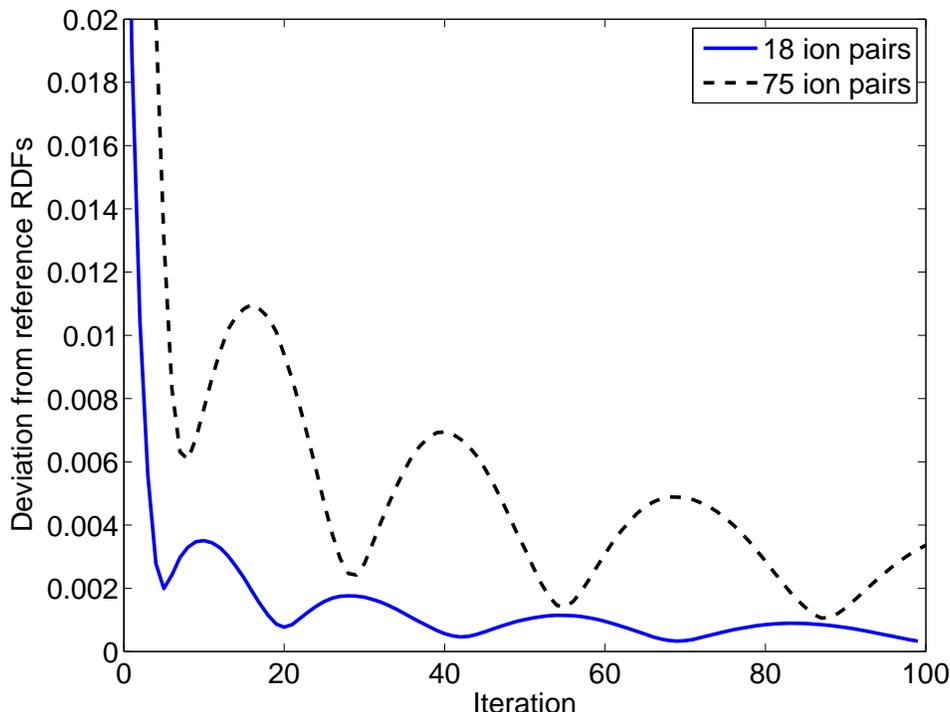


FIG. 5: The IMC process converges nonmonotonically. The squared-norm deviation between the reference RDFs (from MD) and the EPP-driven MC-based RDFs are plotted as a function of IMC iteration number. Results are plotted only for 18 and 75 ion pairs, but the nonmonotonicity was observed in all calculations. Note that the period of the oscillation increases with increasing number of ion pairs.

smooth function (albeit with some added noise). This result is not surprising in light of the interpretation of nonmonotonicity given above, considering that such a smooth function contributes little to the RDF and thereby little to the virial and the osmotic coefficient. For example, from the 18-ion-pair calculation, the osmotic coefficients at iterations 6, 21, 43, and 70 are 1.0465, 1.0508, 1.0455, and 1.0530, respectively; similarly, the osmotic coefficients of the 75-pair calculation at iterations 9, 30, 56, and 88 are 1.178, 1.160, 1.160, and 1.163. Thus, the variations are within the targeted uncertainty tolerance of about 0.02.

In Figures 7(a) and (b) are plotted the EPPs and the resulting forces for the 18-pair and 75-pair IMC calculations at the first minima (iterations 6 and 9, respectively). All plots were generated by first fitting the EPPs to 22nd-order polynomials, and the forces were calculated by taking the analytical derivatives of the polynomial fits. These plots indicate how sensitively the osmotic coefficient depends on fine details in the pair potentials and the

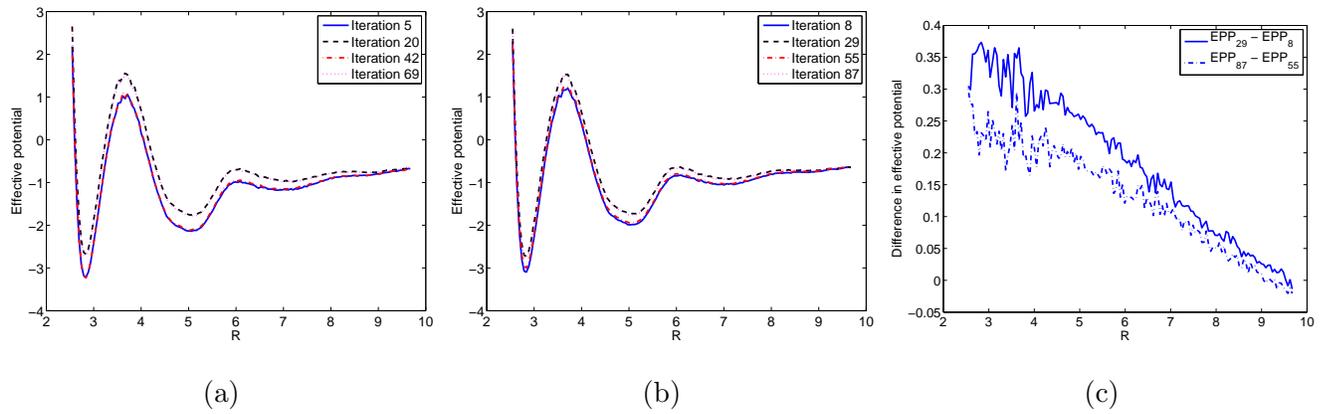


FIG. 6: The Na-Cl effective pair potentials at the first four minima in the IMC iterations employing 18 and 75 ion pairs, illustrating how the potentials are grouped. (a) 18 ion pairs. (b) 75 ion pairs. (c) The difference between the groups is dominated by a smooth, slowly decaying function (increased noise at short distances is believed to be due to poorer statistics). All potentials are in $k_B T$, and the forces are in $k_B T/\text{\AA}$.

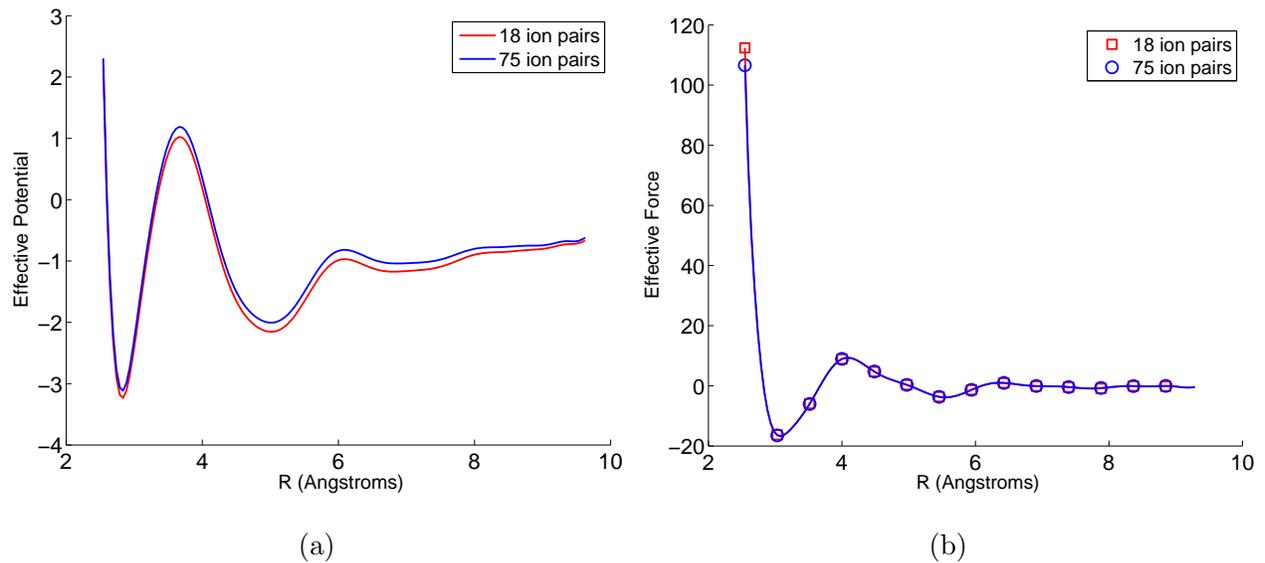


FIG. 7: Comparison of effective Na-Cl potentials (a), and the forces derived from them (b), for EPPs taken from 18-pair and 75-pair realizations of IMC. The EPPs are taken at the iterations corresponding to the first minima in the 2-norm error between the reference and MC-generated RDFs. The reference RDFs are from Lyubartsev and Laaksonen⁷⁵. For clarity in plotting, the potentials have been fit to polynomials of order 22, and the forces have been calculated analytically from the polynomial representation. All potentials are in $k_B T$, and all forces are in $k_B T/\text{\AA}$.

RDFs: despite the substantial differences in the estimates of the osmotic coefficients, the forces in Figure 7(b) are nearly indistinguishable to the eye, and both EPPs give excellent fits to the reference RDFs.

For the remainder of the paper, we use 75 ion pairs except where stated otherwise.

4.5. Reproducibility Study with Different Inputs

Having shown that the stochastic IMC procedure converges robustly given the same input and having analyzed the convergence behavior, we can study the performance of IMC when provided with inputs from different MD simulations. We find that MD simulation length required to achieve acceptable statistics on the estimated osmotic coefficient is, for our 40 Å simulation cell, on the order of 10 ns. Such a time scale is not adequate to obtain converged radial distribution functions. However, it is sufficient to reach a level of variation in the reproducibility results that are comparable to 0.02, which is the variability observed when simply changing the number of ions in the MC simulation (see Figure 3).

In this section, we use IMC to estimate the osmotic coefficients of sodium chloride solutions of 1 M concentration. The force field was the same as described by Bouazizi et al.³⁵, who used the SPC water model and the Smith–Dang parameters for sodium and chloride^{43,44}. Seven independent 12 ns MD simulations were conducted, of which the first 2 ns were used for equilibration and the remainders for calculating RDFs. All IMC calculations were conducted by using the static dielectric constant of 61, which is that estimated for SPC water¹⁰⁰. We discuss the use of concentration-dependent dielectric constants in Section 5.

The molecular dynamics simulations were prepared and analyzed by using VMD¹⁰¹, with the actual MD trajectories calculated with NAMD³⁹. For each simulation, the ionic solution of interest was created by randomly placing the appropriate number of ions in a 40 Å simulation box filled with 1,899 water molecules. The resulting systems were equilibrated by using 10^4 steps of energy minimization and then 0.5 ns of dynamics in the NPT ensemble, prior to the 2-ns equilibration. Langevin dynamics were used in all simulations to maintain constant temperature, with 2 fs timesteps and nonbonded forces evaluated every second timestep. The rigidity of the water molecules was enforced by using SETTLE¹⁰². Long-range electrostatic interactions were calculated by using the particle-mesh Ewald¹⁰³ facility of NAMD, using a grid resolution of 1 Å. All production calculations used the NPT ensemble.

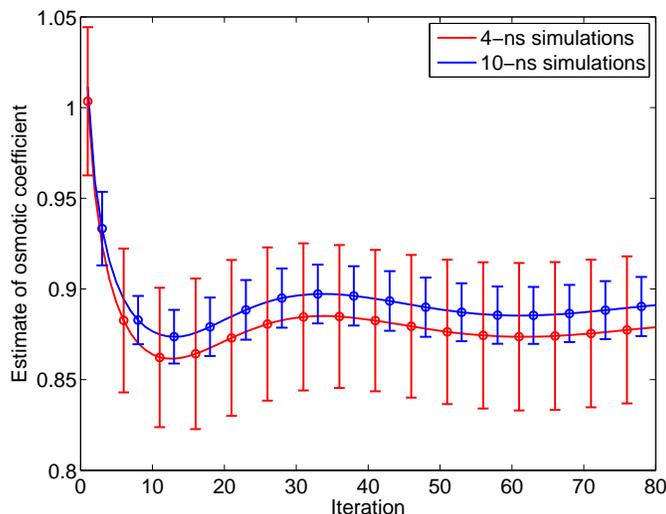


FIG. 8: Reproducibility of the IMC method given different RDFs from seven independent MD simulations. The mean estimated osmotic coefficients are plotted, with error bars denoting the standard deviation at each iteration.

ble; pressure was maintained at one atmosphere by using a Langevin-dynamics-controlled Nosé–Hoover barostat^{104,105}. The Langevin piston period was 100 fs, and the decay was set to 50 fs. Snapshots were taken every 100 fs. Radial distribution functions were calculated by using VMD¹⁰¹.

Figure 8 contains plots of the mean and standard deviation of the osmotic coefficients from two sets of IMC computations. The first set of IMC realizations used RDFs that sampled the first 4 ns of the production MD, and the second set used all 10 ns. For each set, the mean osmotic coefficient at each iteration is plotted along with the sample standard deviation. The figure makes clear that the standard deviations are smaller in the longer simulations, as would be expected. A similar phenomenon is seen in Figure 8 as was noted for Figure 4: the standard deviation does not vary over the course of the IMC iterations.

We also computed the mean osmotic coefficients from the third minimum in the deviations, because not all realizations gave four minima: averaging the osmotic coefficients at the third minimum, the 4 ns simulations had a mean osmotic coefficient of 0.8659, with a standard deviation of 0.036; the 10 ns simulations gave rise to a mean osmotic coefficient of 0.8857, with a standard deviation of 0.013. Thus, the 10 ns simulations reduced the standard deviation to comparable magnitude as the deviation between the most recent extrema in the osmotic coefficient, which was between 0.010 and 0.015 for all observed cases.

5. DISCUSSION

In this paper we have detailed how to obtain reliable estimates of osmotic coefficients of solutions by using the inverse Monte Carlo method introduced by Lyubartsev and Laaksonen⁷⁵. Although the estimates are at this point size dependent, in contradiction to the intensive nature of the osmotic coefficient, a simple correction eliminates this dependence and will be detailed in further work. The IMC software takes as input radial distribution functions calculated from molecular dynamics simulations and determines effective pair potentials for use in implicit-solvent Monte Carlo simulations. The inverse Monte Carlo method, one of several methods that estimate these quantities^{66,106}, is a straightforward approach for connecting explicit-water MD simulations to experimental thermodynamics and can be used for coarse graining much larger systems, which otherwise would be computationally intractable^{107,108}. Our use of IMC has the same goals as those apparent in the work of Hess et al. and Kalcher and Dzubiella⁶⁹, namely, to evaluate the viability of coarse graining MD simulations and to improve our models of solute–solvent interactions, particularly for simulations of biomolecules such as proteins and nucleic acids. Joung and Cheatham pointed out that the chemical potential of an aqueous solution containing sodium chloride changes by much less than 1 kcal/mol between infinite dilution and approximately 1 M concentration⁶⁷, an energetic difference that represents a high standard for computation. It is remarkable that biological systems have evolved such sophisticated mechanisms of controlling ionic concentration that these small differences can be exploited so robustly.

The results presented in the previous sections indicate that osmotic coefficients estimated by using the IMC procedure are robust to approximately 0.02–0.03, whereas experimental values for biological solutions are commonly between 0.80 and 1.20. Although this qualitative error assessment falls far below the accuracy of experimental data, which are usually reported to three digits beyond the decimal, such a confidence level ought to still have its uses in verifying force fields. The confidence level is predicated on the following details of the IMC calculation: that the MC simulations use at least 75 ion pairs and include at least 6 million production steps at each iteration and that at least 25 IMC iterations are performed, although using more iterations allows the determination of an interval within which the osmotic estimate will almost certainly lie. The reproducibility studies suggest that the RDFs require a minimum of 10 ns of MD simulation in order for the estimated osmotic coefficient

to be reproducible to the same level of accuracy as observed varying the number of ions in the IMC procedure. With the current implementation of the IMC software, improving the convergence of the estimates beyond this level requires significantly greater computational work. Nevertheless, even an approximate osmotic coefficient may be useful for verifying that force fields for MD simulations reproduce the measured thermodynamics of solutions of differing concentrations and compositions. We have found that the IMC approach is in fact very stable, although it tends to take many iterations for the EPP and osmotic coefficient to be reasonably converged. In the work reported here, we did not adjust the cut-off distance at which the long-range Coulomb potential was used, nor did we adjust the number of terms retained in the Ewald sum; however, an exploration of the stability of osmotic coefficient with respect to these parameters is warranted.

Because we have been most interested in the parameters required to obtain substantially converged estimates of the osmotic coefficient, we have neglected some important aspects of the inverse-potential problem. In particular, Lyubartsev et al. noted that thermodynamic consistency, and thus the size-independence of the calculated osmotic coefficient, requires that the IMC process should use the same number of ion pairs as the MD simulation used as input^{68,75}; as described by Murtola et al., the effective (coarse-grained) Hamiltonian defined by the effective pair potentials is a function of the thermodynamic state at which it was determined¹⁰⁸. However, results demonstrating which effect dominates error, the ensemble inconsistency or MC convergence, do not seem to have been published, and this represents one area for future work. Also, the observed variations in calculated quantities suggest that single-point calculations are inadequate and that multiple EPPs should be computed independently, with the statistics of these computations reported.

Lyubartsev and Laaksonen also emphasize the viability of the IMC algorithm for estimating osmotic coefficients at salt concentrations other than the one at which the IMC process was conducted⁶⁸. We have found in our own work that when different numbers of salts are used in the MD and MC calculations, extrapolation (that is, evaluating the osmotic coefficient at a concentration other than at the concentration of parameterization) gives estimates of the osmotic coefficient that are inconsistent with the estimates from EPPs parameterized at the new concentration. These findings do not contradict the earlier results because, as mentioned above, they emphasized the importance of using the same number of salt pairs in the MC iterations as in the MD simulations^{68,75,77}, whereas our emphasis on detailed

empirical convergence analysis led us to explore using a wide range of different numbers of salts in the MC iterations. However, the ability to perform such extrapolations would be valuable for estimating the quality of force fields (by reducing the required MD simulation time at different concentrations); a serious effort should be made to establish the precise conditions under which the EPPs can be used at multiple concentrations with confidence.

As described in Section 2, the IMC software approximates the EPPs using piecewise-constant basis functions. This approach is straightforward to conceptualize and implement, but may be computationally inefficient. In particular, the relatively smooth nature of the potentials away from the limits $r \rightarrow 0$ and $r \rightarrow \infty$ suggests that linear or polynomial basis functions may allow significant reductions in the dimensionality of the Newton–Raphson problem. Furthermore, the use of global basis functions (say, orthogonal polynomials) rather than the highly localized piecewise-constant basis functions may be a means to accelerate convergence. The demonstrated performance and robustness of the IMC approach^{107,109} establish the merit of going to the effort to actually implement these more sophisticated techniques. Such implementations may be important for extensions to multicomponent solutions or for studying three-body and higher-order correlations⁶⁸. Other advances in the inverse method also seem to argue for the use of as efficient a basis set as possible, as recent work by Murtola and Karttunen *et al.* has incorporated constraints¹¹⁰ as well as internal states¹⁰⁸ in the inverse problem. Furthermore, convergence of the IMC process can also be improved by using a potential-splitting approach similar to that of Kalcher and Dzubiella⁶⁹. The original IMC method solves for a total effective potential at a distance less than the cutoff distance, with the short-ranged solvent-packing component of the total effective potential converging much more quickly than the smoother component of the total effective potential. Kalcher and Dzubiella, in contrast, perform the splitting directly between the short- and long-range components, and the apparent concentration-independence of the short range potential suggests a possible path to efficient computation of extrapolated osmotic coefficients⁶⁹.

It is also worth noting from the theoretical viewpoint, rather than the numerical viewpoint, the differing treatments of the long-range interactions in IMC⁷⁵, in the work of Kalcher and Dzubiella⁶⁹, and in that of Hess *et al.*^{56,111}. The IMC procedure described by Lyubartsev and Laaksonen employs the bulk static dielectric constant of the solvent at all concentrations⁶⁸. Hess *et al.* use concentration-dependent dielectric constants^{56,111}. This difference

is similar to the the difference between the original primitive model, which used the bulk-solvent dielectric constant in all solutions, and the modified primitive model proposed by Simonin et al.⁶⁴, who demonstrated that incorporating a concentration-dependent dielectric constant increased the concentration range over which the model made reasonable predictions for osmotic coefficients. Kalcher and Dzubiella, in contrast, used a Debye–Hückel potential at long range. These approaches are not all equivalent, and a detailed comparison of their strengths and weaknesses remains to be performed. It may be easier to identify inconsistencies between these models by performing such a comparison in a medium with lower dielectric, owing to the rapid decay of Coulombic forces in high dielectrics, which may mask important differences between models of long-range interactions. A wealth of experimental thermodynamic data exists to help parameterize and refine force fields and to resolve questions of how to most efficiently calibrate models against reality.

ACKNOWLEDGMENTS

The authors thank A. P. Lyubartsev, B. M. Pettitt, B. Roux, and D. Henderson for valuable discussions and suggestions. The authors are grateful to A. P. Lyubartsev and colleagues for making the IMC software available. The authors gratefully acknowledge the use of the Jazz computing cluster operated by the Mathematics and Computer Science Division at Argonne National Laboratory as part of its Laboratory Computing Resource Center. J. P. B. acknowledges partial support from a Wilkinson Fellowship in Scientific Computing funded by the Office of Advanced Scientific Computing Research, Office of Science, U. S. Dept. of Energy, under Contract DE-AC02-06CH11357. B. E. acknowledges support from the National Institutes of Health, grant GM076013.

¹ G. Hummer, L. R. Pratt, and A. E. Garcia, *Journal of Physical Chemistry A* **102**, 7885 (1998).

² J. M. G. Barthel, H. Krienke, and W. Kunz, *Physical chemistry of electrolyte solutions: modern aspects* (Springer, 1998).

³ P. C. Jordan, *IEEE Transactions on Nanobioscience* **4**, 3 (2005).

⁴ M. K. F. Wikstrom, *Nature* **266**, 271 (1977).

- ⁵ B. E. Schultz and S. I. Chan, *Annual Review of Biophysics and Biomolecular Structure* **30**, 23 (2001).
- ⁶ J. Barber and B. Andersson, *Nature* **370**, 31 (1994).
- ⁷ E. N. Dedkova and L. A. Blatter, *Cell Calcium* **44**, 77 (2008).
- ⁸ M. J. Page and E. Di Cera, *Physiological Reviews* **86**, 1049 (2006).
- ⁹ D. E. Clapham, *Cell* **131**, 1047 (2007).
- ¹⁰ L. J. Henderson, *The fitness of the environment: An inquiry into the biological significance of the properties of matter* (Macmillan, New York, 1913).
- ¹¹ K. Linderstrom-Lang, *Compt. Rend. Trav. Lab. Carlsberg (ser chimie)* **15**, 1 (1924).
- ¹² L. J. Henderson, *Blood. A study in general physiology* (Yale University Press, New Haven, CT, 1928).
- ¹³ E. J. Cohen and J. Edsall, *Proteins, amino acids, and peptides* (Reinhold, New York, 1943).
- ¹⁴ J. Edsall and J. Wyman, *Biophysical chemistry* (Academic Press, 1958).
- ¹⁵ C. Tanford, *Physical chemistry of macromolecules* (Wiley, New York, 1961).
- ¹⁶ A. Warshel and S. T. Russell, *Quarterly Review of Biophysics* **17**, 283 (1984).
- ¹⁷ C. Tanford and J. Reynolds, *Nature's robots: a history of proteins* (Oxford, New York, 2001).
- ¹⁸ S. Durand-Vidal, J.-P. Simonin, and P. Turq, *Electrolytes at interfaces* (Kluwer, Boston, 2000).
- ¹⁹ S. Durand-Vidal, P. Turq, O. Bernard, C. Treiner, and L. Blum, *Physica A* **231**, 123 (1996).
- ²⁰ W. R. Fawcett, *Liquids, solutions, and interfaces: from classical macroscopic descriptions to modern microscopic details* (Oxford University Press, New York, 2004).
- ²¹ H. S. Harned and B. B. Owen, *The physical chemistry of electrolytic solutions* (Reinhold Publishing Corporation, New York, 1958), 3rd ed.
- ²² L. L. Lee, *Molecular thermodynamics of nonideal fluids* (Butterworth-Heinemann, New York, 1988).
- ²³ L. L. Lee, *Molecular thermodynamics of electrolyte solutions* (World Scientific, Singapore, 2008).
- ²⁴ K. S. Pitzer, *Activity coefficients in electrolyte solutions* (CRC Press, Boca Raton, FL, 1991).
- ²⁵ K. S. Pitzer, *Thermodynamics* (McGraw Hill, New York, 1995), 3rd ed.
- ²⁶ R. A. Robinson and R. H. Stokes, *Electrolyte solutions* (Butterworths Scientific Publications, London, 1959), 2nd ed.
- ²⁷ W. J. Hamer and Y.-C. Wu, *J. Phys. Chem. Ref. Data* **1**, 1047 (1972).

- ²⁸ R. N. Goldberg and R. L. Nuttall, *J. Phys. Chem. Ref. Data* **7**, 263 (1978).
- ²⁹ K. S. Pitzer and G. Mayorga, *J. Phys. Chem.* **77**, 2300 (1973).
- ³⁰ E. Colin, W. Clarke, and D. N. Glew, *J. Phys. Chem. Ref. Data* **14**, 489 (1985).
- ³¹ K. S. Pitzer, J. C. Pelper, and R. H. Busey, *J. Phys. Chem. Ref. Data* **13**, 1 (1984).
- ³² G. M. Blackburn, T. H. Lilley, and E. Walmsley, *J. Chem. Soc., Faraday Trans. 1* **78**, 1641 (1982).
- ³³ K. Miyajima, M. Sawada, and M. Nakagaki, *Bull. Chem. Soc. Jpn.* **56**, 1620 (1983).
- ³⁴ B. Hille, *Ionic channels of excitable membranes* (Sinauer Associates Inc., 2001), 3rd ed.
- ³⁵ S. Bouazizi, S. Nasr, N. Jaïdane, and M.-C. Bellissent-Funel, *Journal of Physical Chemistry B* **110**, 23515 (2006).
- ³⁶ J. E. Enderby, *Annual Review of Physical Chemistry* **34**, 155 (1983).
- ³⁷ E. Matsubara and Y. Waseda, *J. Phys.: Condens. Matter* **1**, 8575 (1989).
- ³⁸ B. M. Pettitt and P. J. Rossky, *Journal of Chemical Physics* **84**, 5836 (1986).
- ³⁹ J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, *Journal of Computational Chemistry* **26**, 1781 (2005).
- ⁴⁰ B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *Journal of Computational Chemistry* **4**, 187 (1983).
- ⁴¹ D. A. Case, T. E. C. III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *Journal of Computational Chemistry* **26**, 1668 (2005).
- ⁴² A. D. M. Jr., D. Bashford, M. Bellott, R. L. D. Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., *Journal of Physical Chemistry B* **102**, 3586 (1998).
- ⁴³ L. X. Dang and D. E. Smith, *Journal of Chemical Physics* **99**, 6950 (1993).
- ⁴⁴ D. E. Smith and L. X. Dang, *Journal of Chemical Physics* **100**, 3757 (1994).
- ⁴⁵ J. Aqvist, *Journal of Physical Chemistry* **94**, 8021 (1990).
- ⁴⁶ J. W. Ponder and D. A. Case, *Advances in Protein Chemistry* **66**, 27 (2003).
- ⁴⁷ A. Warshel and M. Levitt, *Journal of Molecular Biology* **103**, 227 (1976).
- ⁴⁸ F. S. Lee, Z. T. Chu, and A. Warshel, *Journal of Computational Chemistry* **14**, 161 (1993).
- ⁴⁹ J. L. Banks, G. A. Kaminski, R. H. Zhou, D. T. Mainz, B. J. Berne, and R. A. Friesner, *Journal of Chemical Physics* **110**, 741 (1999).
- ⁵⁰ G. Lamoureux and B. Roux, *Journal of Physical Chemistry B* **110**, 3308 (2006).
- ⁵¹ G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, *Journal of Physical Chemistry*

- A **108**, 621 (2004).
- ⁵² B. C. Kim, T. Young, E. Harder, R. A. Friesner, and B. J. Berne, *Journal of Physical Chemistry B* **109**, 16529 (2005).
- ⁵³ T. W. Whitfield, S. Varma, E. Harder, G. Lamoureux, S. B. Rempe, and B. Roux, *Journal of Chemical Theory and Computation* **3**, 2068 (2007).
- ⁵⁴ C. A. Wick and S. S. Xantheas, *Journal of Physical Chemistry B* **113**, 4141 (2009).
- ⁵⁵ D. Horinek, S. I. Mamatkulov, and R. R. Netz, *Journal of Chemical Physics* **130**, 124507 (2009).
- ⁵⁶ B. Hess, C. Holm, and N. van der Vegt, *Journal of Chemical Physics* **124** (2006).
- ⁵⁷ P. Auffinger, T. E. Cheatham III, and A. C. Vaiana, *Journal of Chemical Theory and Computation* **3**, 1851 (2007).
- ⁵⁸ A. A. Chen and R. V. Pappu, *Journal of Physical Chemistry B* **111**, 11884 (2007).
- ⁵⁹ P. E. Smith, *Journal of Chemical Physics* **129**, 124509 (2008).
- ⁶⁰ S. Weerasinghe and P. E. Smith, *Journal of Physical Chemistry B* **107**, 3891 (2003).
- ⁶¹ J. C. Rasaiah and H. L. Friedman, *Journal of Chemical Physics* **48**, 2742 (1968).
- ⁶² D. N. Card and J. P. Valleau, *Journal of Chemical Physics* **52**, 6232 (1970).
- ⁶³ J. P. Valleau and L. K. Cohen, *Journal of Chemical Physics* **72**, 5935 (1980).
- ⁶⁴ J.-P. Simonin, L. Blum, and P. Turq, *Journal of Physical Chemistry* **100**, 7704 (1996).
- ⁶⁵ Z. Abbas, E. Ahlberg, and S. Nordholm, *Journal of Physical Chemistry B* **113**, 5905 (2009).
- ⁶⁶ T. Lazaridis and M. E. Paulaitis, *A. I. Ch. E. Journal* **39**, 1051 (1993).
- ⁶⁷ I. S. Joung and T. E. Cheatham III, *Journal of Physical Chemistry B* **113**, 13279 (2009).
- ⁶⁸ A. P. Lyubartsev and A. Laaksonen, *Physical Review E* **55**, 5689 (1997).
- ⁶⁹ I. Kalcher and J. Dzubiella, *Journal of Chemical Physics* **130**, 134507 (2009).
- ⁷⁰ M. Druchok, Y. Kalyuzhnyi, J. Rešičič, and V. Vlachy, *Journal of Chemical Physics* **124**, 114902 (2006).
- ⁷¹ S. N. Yu, W. H. Geng, and G. W. Wei, *Journal of Chemical Physics* **126**, 244108 (2007).
- ⁷² Z. Li and J. Wu, *Physical Review E* **70**, 031109 (2004).
- ⁷³ P. L. Hansen, R. Podgornik, and V. A. Parsegian, *Physical Review E* **64**, 021907 (2001).
- ⁷⁴ M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, *Biophysical Journal* **72**, 1047 (1997).
- ⁷⁵ A. P. Lyubartsev and A. Laaksonen, *Physical Review E* **52**, 3730 (1995).

- ⁷⁶ T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, *Physical Chemistry Chemical Physics* **11**, 1869 (2009).
- ⁷⁷ A. P. Lyubartsev and A. Laaksonen, *Journal of Physical Chemistry* **100**, 16410 (1996).
- ⁷⁸ A. P. Lyubartsev and A. Laaksonen, *Computer Physics Communications* **121-122**, 57 (1999).
- ⁷⁹ A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen, *Soft Materials* **1**, 121 (2003).
- ⁸⁰ A. A. Louis, P. G. Bolhuis, J. P. Hansen, and E. J. Meijer, *Physical Review Letters* **85**, 2522 (2000).
- ⁸¹ P. G. Bolhuis, A. A. Louis, J. P. Hansen, and E. J. Meijer, *Journal of Chemical Physics* **114**, 4296 (2001).
- ⁸² S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, *Journal of Chemical Physics* **120**, 10896 (2004).
- ⁸³ S. Izvekov and G. A. Voth, *Journal of Chemical Physics* **123**, 134105 (2005).
- ⁸⁴ Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfandter, and G. A. Voth, *Biophysical Journal* **95**, 5073 (2008).
- ⁸⁵ D. Boda, D. Busath, B. Eisenberg, D. Henderson, and W. Nonner, *Physical Chemistry Chemical Physics* **4**, 5154 (2002).
- ⁸⁶ D. Boda, M. Valiskó, B. Eisenberg, W. Nonner, D. Henderson, and D. Gillespie, *Journal of Chemical Physics* **125**, 034901 (2006).
- ⁸⁷ D. Boda, W. Nonner, M. Valiskó, D. Henderson, B. Eisenberg, and D. Gillespie, *Biophysical Journal* **93**, 1960 (2007).
- ⁸⁸ D. Boda, M. Valiskó, B. Eisenberg, W. Nonner, D. Henderson, and D. Gillespie, *Physical Review Letters* **92**, 168102 (2007).
- ⁸⁹ D. Boda, W. Nonner, D. Henderson, B. Eisenberg, and D. Gillespie, *Biophysical Journal* **94**, 3486 (2008).
- ⁹⁰ R. L. Henderson, *Physics Letters A* **A49**, 197 (1974).
- ⁹¹ J. T. Chayes, L. Chayes, and E. H. Lieb, *Communications in Mathematical Physics* **93**, 57 (1984).
- ⁹² J. T. Chayes and L. Chayes, *Journal of Statistical Physics* **36**, 471 (1984).
- ⁹³ H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems* (Kluwer, Dordrecht, Netherlands, 2000).

- ⁹⁴ J. Kaipio and E. Somersalo, *Statistical and computational inverse problems* (Springer, New York, NY, 2005).
- ⁹⁵ S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, *J. Phys.: Condens. Matter* **16**, R481 (2004).
- ⁹⁶ K. Toukan and A. Rahman, *Physical Review B* **31**, 2643 (1985).
- ⁹⁷ R. L. McGreevy, *J. Phys.: Condens. Matter* **13**, R877 (2001).
- ⁹⁸ F. L. B. da Silva, B. Svensson, T. Åkesson, and B. Jönsson, *Journal of Chemical Physics* **109**, 2624 (1998).
- ⁹⁹ D. P. Bertsekas, *Nonlinear Programming* (Athena Scientific, Belmont, MA, 1999), 2nd ed.
- ¹⁰⁰ D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen, *Journal of Chemical Physics* **108**, 10220 (1998).
- ¹⁰¹ W. Humphrey, A. Dalke, and K. Schulten, *Journal of Molecular Graphics* **14**, 33 (1996).
- ¹⁰² S. Miyamoto and P. A. Kollman, *Journal of Computational Chemistry* **13**, 952 (1992).
- ¹⁰³ T. Darden, D. York, and L. Pedersen, *Journal of Chemical Physics* **98**, 10089 (1993).
- ¹⁰⁴ G. J. Martyna, D. J. Tobias, and M. L. Klein, *Journal of Chemical Physics* **101**, 4177 (1994).
- ¹⁰⁵ S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, *Journal of Chemical Physics* **103** (1995).
- ¹⁰⁶ P. J. Lenart, A. Jusufi, and A. Z. Panagiotopoulos, *Journal of Chemical Physics* **126** (2007).
- ¹⁰⁷ A. P. Lyubartsev, *Eur. Biophys. J.* **35**, 53 (2005).
- ¹⁰⁸ T. Murtola, M. Karttunen, and I. Vattulainen, *Journal of Chemical Physics* **131**, 055101 (2009).
- ¹⁰⁹ A. P. Lyubartsev and S. Marčelja, *Physical Review E* **65** (2002).
- ¹¹⁰ T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen, *Journal of Chemical Physics* **126**, 075101 (2007).
- ¹¹¹ B. Hess, C. Holm, and N. van der Vegt, *Physical Review Letters* **96**, 147801 (2006).

NOTICE

The submitted manuscript has been created by the University of Chicago, LLC as Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.