FEATURE ARTICLE

Molecular Theories and Simulation of Ions and Polar Molecules in Water

Gerhard Hummer, Lawrence R. Pratt,* and Angel E. García

Theoretical Division, Mail Stop B268, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received: May 11, 1998; In Final Form: July 29, 1998

Recent developments in molecular theories and simulation of ions and polar molecules in water are reviewed. The hydration of imidazole and imidazolium is used to exemplify the theoretical issues. The treatment of long-ranged electrostatic interactions in simulations is discussed extensively. It is argued that the Ewald approach is an easy way to get correct hydration free energies corresponding to thermodynamic limit from molecular calculations. Molecular simulations with Ewald interactions and periodic boundary conditions can also be more efficient than many common alternatives. The Ewald treatment permits a conclusive extrapolation to infinite system size. Accurate results for well-defined models have permitted careful testing of simple theories of electrostatic hydration free energies, such as dielectric continuum models. The picture that emerges from such testing is that the most prominent failings of the simplest theories are associated with solvent proton conformations that lead to non-Gaussian fluctuations of electrostatic potentials. Thus, the most favorable cases for second-order perturbation theories are monoatomic positive ions. For polar and anionic solutes, continuum or Gaussian theories are less accurate. The appreciation of the specific deficiencies of those simple models have led to new concepts, multistate Gaussian and quasi-chemical theories, which address the cases for which the simpler theories fail. It is argued that, relative to direct dielectric continuum treatments, the quasi-chemical theories provide a better theoretical organization for the computational study of the electronic structure of solution species.

1. Introduction

Water, the most commonly encountered liquid, exerts both chemical and physical influences on aqueous molecular processes. Hydration effects are often divided into hydrophobic and hydrophilic categories. Hydrophilic solutes are typically ionic or polar species and may participate in chemical interactions with the water solvent. Because of the long range of the electrostatic interactions and their strength relative to k_BT, hydrophilic hydration presents distinctive conceptual and practical issues for understanding and predicting the influence of hydration on chemical and biochemical events in water.

A principal and long-standing technical issue is the treatment of infinitely long-ranged interactions in the context of a sample of finite size.[1] Recent work has helped to resolve this problem. One algorithmic approach to treatment of long-ranged interactions is the use of Ewald interactions within the conventional periodic boundary conditions.[2] We argue here that *the Ewald approach is an easy way to get correct hydration free energies* from molecular calculations, that is, to achieve well-characterized results appropriate to the thermodynamic limit in which the system size tends to infinity for given densities and temperature. In addition, molecular simulations with Ewald interactions and periodic boundary conditions can also be more efficient than rougher approximations that are often employed to compute hydration free energies for molecularly well-defined problems. We anticipate results below by noting that we obtain accurate, thermodynamic limiting results for the hydration free energy of imidazole with as few as 16 water molecules included in the simulation. The price to be paid for this accuracy and efficiency is additional effort in understanding Ewald calculations from a physical viewpoint and in implementing Ewald interactions,[2] its equivalents,[3−7] and alternatives.[8−10]

The physical issues motivating simulation calculations of this type revolve around dielectric continuum models of hydration of ionic and polar solutes.[11] It is natural and common for a simple approximation to provide a conceptual baseline for considering more accurate theoretical results. But the converse comparison is foremost for this work. The theoretical efforts over recent years have provided sharper tests of the validity of the continuum approach than merely the question: is an empirically correct hydration free energy obtained? Recent work has clarified that the dielectric models are simple implementations of thermodynamic perturbation theory through second order in electrostatic coupling parameters such as solute charges;[12−14] dielectric models can also be considered a simple implementation of an ansatz that electrostatic potential fluctuations are distributed according to a Gaussian probability density,[15] or they can be considered a simplified linear response theory.[16−18]

Second-order perturbation theory was found to be satisfactory for some solutes such as alkali ions,[14] but unsatisfactory for water[13,19] and anions.[14] In the latter cases, of course, an a posteriori adjustment of cavity radii could still produce the correct hydration free energies.[20] However, the more ambitious molecular theory ties the values of radii parameters to molecular properties that depend on the thermodynamic state of the system

S1089-5639(98)02195-1 CCC: $15.00
Published on Web 09/22/1998

(temperature, pressure, and composition of the solvent) and to nonelectrostatic characteristics of the solute−solvent interactions. The radii are not separately adjustable when viewed from that deeper level of molecular theory. However, the radii can be well-defined and are *not* properties of the solutes alone but incorporate information about the solvent and thermodynamic state.

For water as a solvent, the case of exclusive concern here, the most prominent failings of second-order perturbation theory are associated with solvent proton conformations that lead to non-Gaussian fluctuations of electrostatic potentials.[14,21] Thus, the most favorable cases for the second-order perturbation theories are monovalent atomic positive ions. In such cases, oxygen−hydrogen bonds of water are oriented away from the ion. Neutral, polar molecules that may form specific hydrogen bonds with the solvent are more challenging for these theories, though the hydration free energies sought are smaller in magnitude than for typical ions. Negative molecular ions are expected to offer further complications because now the problematic proton interactions with the solute will be strong. However, we have less experience with realistic negative ions partly because the molecular models used for simulation are less well-developed than for other cases.

The appreciation of the different possibilities for fluctuations has led to new theories of electrostatic hydration free energies.[21] These theories analyze electrostatic distributions more broadly, still using Gaussian models at crucial steps, but now several Gaussian distributions are derived from an analysis of the first-shell environment of the solute. For the important case of hydration of a water molecule, this extension repairs the breakdown of a single Gaussian theory. Negative ions can still be problematic, but the multiple Gaussian approach has also motivated development of quasi-chemical theories[22] that are based, in principle, on full information about the thermal motion of the first hydration shell. Though experience with the quasi-chemical theories is limited,[23] we anticipate that they should provide better descriptions of the hydration free energies, in addition to providing a reasonable pathway to carry out solution phase electronic structure calculations on hydrated negative ions—calculations that would be difficult particularly in the absence of hydration effects.

In the following section, we will first introduce the model solute imidazole, which was chosen as a molecular solute to exemplify, combine, and extend aspects of ionic and polar solutes studied previously.[12−14,21,24] Results for imidazole and imidazolium will be used throughout the manuscript to illustrate the theoretical issues. We will discuss the Ewald treatment of electrostatic interactions, motivating it in various ways. Subsequently, finite-size effects will be studied. The correction for the typically large finite-size effects is essential for accurate calculations of solvation free energies of polar and charged solutes. We distinguish between electrostatic finite-size effects that are independent of the thermodynamic state and the characteristics of the solute, and the remaining thermodynamic finite-size effects. We will then introduce perturbative methods for calculating solvation free energies that are based on the approximately Gaussian character of the electrostatic potential fluctuations. Non-Gaussian behavior and its accurate treatment using multistate Gaussian and quasi-chemical models will be the focus of the last section.

2. Example: Imidazole and Imidazolium in Water

To illustrate the various issues arising in calculations of solvation free energies of charged and polar molecules, we present new calculations of the hydration of imidazole and imidazolium. We choose this example because recent interest in these problems has focused on predicting acid−base equilibria of biochemical relevance.[25−31,33] We will calculate the charging free energies of the protonated imidazolium and the neutral, polar imidazole (Figure 1). Imidazole in water provides a rich example: the polar imidazole molecule can be protonated at the N_3 position to form a molecular cation, imidazolium. This protonation reaction has a pK_a of about 7.[32] It provides a basis for pK_a calculations of ionizable residues of proteins.[25−31] Imidazole is the building block of histidine, one of the most active amino acids enzymatically and ubiquitous in the active sites of enzymes that operate at room temperature and neutral pH.[33]

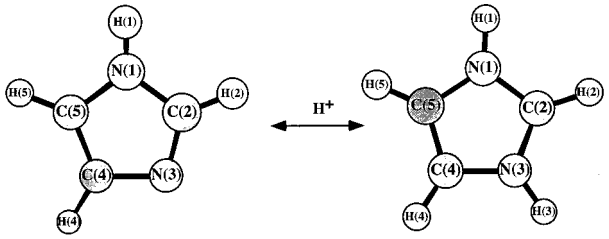

Figure 1. Protonation equilibrium between imidazole and imidazolium.

The protonation of imidazole has been studied previously using combinations of dielectric, quantum mechanical, and computer simulation methods.[33−38] Here, we will focus on the solvation contribution to the protonation equilibrium and reserve quantum mechanical intramolecular effects for subsequent treatment.

The imidazole/imidazolium system is more complex than the systems we have studied before using explicit solvent models: mono- and divalent ions,[14,21] water in water,[13] and tetramethylammonium.[24] The analyses below should also illustrate how the calculation of solvation free energies using the Ewald method and equilibrium fluctuations of electrostatic potentials can be extended to proteins, in particular the calculation of pK_a's of amino acids.

Monte Carlo Computer Simulations of Imidazole in Water. We studied the solvation of imidazole [Im(p)] and imidazolium [Im(+)] in water using Monte Carlo (MC) simulations in the canonical ensemble. For the Im(p) and Im(+) molecules, we used the partial charges and geometry of Topol et al.,[36] as compiled in Table 1. For water, we used the SPC/E model.[40] The temperature was 298 K. The Ewald method was used for the long-range electrostatic interactions with a real-space screening factor of $\eta = 5.6/L$, where L is the length of the periodically replicated cubic box. A cutoff of $k^2 \le 38(2\pi/L)^2$ was applied in Fourier space, resulting in 2×510 **k** vectors being considered. A cutoff of $L/2$ was applied to the Lennard-Jones and real-space electrostatic interactions based on atoms. The background dielectric constant in the Ewald method was corrected from infinity to 80.[41] The partial molar volume of the imidazole was chosen as 2 times that of bulk water at a density of 997.07 kg/m^3, such that the pressure[42] was about 1 atm.

The Metropolis MC method was used to sample configurational space in the simulations.[43] The translational and rotational move widths of water were chosen to give about 40% acceptance ratios. The solute was allowed to move as well. Simulations started from random configurations or configurations of previous runs with different charges, equilibrated for at least 100 000 and 50 000 MC passes, respectively, where one pass is one attempted move for each of the particles. Electrostatic potentials

TABLE 1: Coordinates *x* and *y* (in nm) and Charges *q* (in Elementary Charge Units *e*) of the Atoms in the Planar Imidazole and Imidazolium from the Quantum Mechanical Calculations of Topol et al.[36]

atom	*x*	*y*	*q*	σ	ϵ
			Imidazole		
N1	0.000 0	0.110 5	−0.090 285	0.325 000	0.711 28
C2	−0.109 1	0.028 2	0.232 373	0.339 967	0.359 82
N3	−0.074 1	−0.098 3	−0.715 903	0.325 000	0.711 28
C4	0.063 6	−0.098 4	0.217 356	0.339 967	0.359 82
C5	0.112 0	0.029 8	−0.374 687	0.339 967	0.359 82
H1	−0.000 9	0.211 2	0.318 027	0.106 908	0.065 69
H2	−0.210 2	0.066 1	0.102 391	0.242 146	0.062 76
H4	0.119 7	−0.190 5	0.082 346	0.242 146	0.062 76
H5	0.211 9	0.070 0	0.228 383	0.242 146	0.062 76
			Imidazolium		
N1	0.000 0	0.112 8	−0.115 106	0.325 000	0.711 28
C2	−0.108 6	0.035 3	0.010 825	0.339 967	0.359 82
N3	−0.066 3	−0.091 2	−0.122 786	0.325 000	0.711 28
C4	0.071 9	−0.094 9	−0.139 642	0.339 967	0.359 82
C5	0.114 0	0.034 4	−0.122 097	0.339 967	0.359 82
H1	−0.001 8	0.214 1	0.398 875	0.106 908	0.065 69
H2	−0.211 0	0.068 6	0.230 198	0.242 146	0.062 76
H3	−0.127 4	−0.172 1	0.402 905	0.106 908	0.065 69
H4	0.127 3	−0.187 2	0.232 002	0.242 146	0.062 76
H5	0.213 1	0.076 6	0.224 826	0.242 146	0.062 76

[a] The Lennard-Jones-parameters σ and ϵ (in nm and kJ/mol) are taken from the AMBER force field.[39] Lorentz-Berthelot mixing rules were applied to combine the Lennard-Jones parameters of the solute atoms with those of SPC/E water.[40]

at solute atom positions and binding energies of the solute were calculated after every pass using the Ewald method in simulations extending over 200 000 MC passes each. Simulations were performed for the uncharged, half-charged, and fully charged Im(p) and Im(+) in their respective geometries. The equilibrium simulations were performed with 16, 32, 64, 128, 256, and 512 water molecules to study finite size effects.

To complete the thermodynamic cycle and check for consistency, two runs of slow growth thermodynamic integration were used to calculate the free energy of converting the geometry from the uncharged Im(p) to the uncharged Im(+) conformation within 150 000 MC passes. Six runs of 200 000 MC passes were used to calculate the free energy of converting the polar Im(p) into an Im(+) cation, starting from different equilibrated configurations and averaging three charging and three uncharging runs. The thermodynamic integrations were carried out with 256 water molecules.

We will discuss the results of these calculations as they come up in the theoretical narrative. However, before considering more subtle issues, we can make a direct comparison of the average electrostatic potential exerted by the solvent observed during the simulation with the corresponding predictions of dielectric models. Figure 2 shows that comparison for several sets of radii in current use. Such a comparison illustrates the basic issue of sensitivity of thermodynamic results to the radii parameters and whether extant empirically adjusted radii are transferable to slightly different cases.

3. Noteworthy Aspects of the Ewald Treatment of Electrostatic Interactions

Viewed from an historical perspective, the most appropriate treatment of electrostatic interactions for simulation calculations has been a contentious issue. In this setting it is helpful to note some broad, and nontechnical, characteristics of Ewald treatments that might be typically overlooked. We preface these observations by noting that simulation calculations treat finite

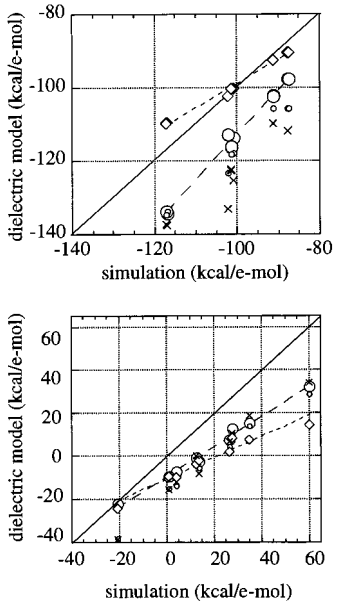

Figure 2. Comparison of dielectric models (ordinate) with molecular simulations (abscissa) for the induced electrostatic potentials due to the solvent at the atom centers for Im(+) (upper panel) and Im(p) (lower panel). Dielectric model results were obtained for several sets of radii in current use: (diamonds) $R_C = 0.267$ nm, $R_N = 0.231$ nm;[44] (large circles) $R_{H(N)} = 0.1160$ nm, $R_{H(C)} = 0.1710$ nm, $R_C = 0.230$ nm, $R_N = 0.150$ nm;[36] (small circles) $R_H = 0.1172$ nm, $R_C = 0.2096$ nm, $R_N = 0.1738$ nm;[45] (crosses) $R_H = 0.1172$ nm, $R_C = 0.1635$ nm, $R_N = 0.1738$ nm.[45] A boundary element method was used for the dielectric model calculations, $\epsilon = 77.4$.[46] Notice that a radii set that happens to be qualitatively satisfactory for the cation (diamonds) can be significantly less satisfactory for the slightly different circumstance of the neutral polar molecule.

systems. Most commonly, periodic (or Born−von Karman) boundary conditions[47,48] are utilized for the exterior boundary of the finite system considered. The theoretical issues engendered by these boundary conditions with finite-ranged[49,50] and long-ranged interactions are reasonably well understood. Of course, simulation calculations need not address issues of what is happening outside the simulation cell. We note that it is possible to compute the Ewald potential, and in more than one way, without consideration of image charges outside the simulation cell. So intuitive arguments based upon image charges can be avoided completely.

It is convenient to express the Ewald electrostatic energy of a system of partial charges $q_{i\alpha}$ at positions $\mathbf{r}_{i\alpha}$ on molecules i as a sum of effective pair interactions and self terms,

$$U = \sum_{\substack{i,j \\ i<j}} \sum_{\alpha,\beta} q_{i\alpha} q_{j\beta} \varphi(\mathbf{r}_{i\alpha j\beta}) + \sum_{i} \sum_{\substack{\alpha,\beta \\ \alpha<\beta}} q_{i\alpha} q_{i\beta} \left[\varphi(\mathbf{r}_{i\alpha i\beta}) - \frac{1}{|\mathbf{r}_{i\alpha i\beta}|} \right] + \frac{1}{2} \sum_{i} \sum_{\alpha} q_{i\alpha}{}^2 \lim_{\mathbf{r}\to 0} \left[\varphi(\mathbf{r}) - \frac{1}{|\mathbf{r}|} \right] \quad (1)$$

where $\mathbf{r}_{i\alpha j\beta} = \mathbf{r}_{j\beta} - \mathbf{r}_{i\alpha}$. The Coulomb energy U was split into intermolecular, intramolecular and self interaction contributions. The Fourier representation of $\varphi(\mathbf{r})$ reveals the periodicity of this potential:[51]

$$\varphi(\mathbf{r}) = \frac{1}{V} \sum_{\substack{\mathbf{k} \\ k\neq 0}} \frac{4\pi}{\mathbf{k}^2} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (2)$$

The $\mathbf{k}$ sum extends over the reciprocal lattice derived from the

real-space lattice **n** of periodically replicated simulation boxes. For a cubic lattice of length $L = V^{1/3}$, we have $\mathbf{n} = L\,(i, j, k)$ and $\mathbf{k} = 2\pi L^{-1}\,(i, j, k)$, where i, j, and k are integers. For numerical convenience, $\varphi(\mathbf{r})$ is partly transformed into real space, which leads to its Ewald lattice sum representation:

$$\varphi(\mathbf{r}) = \sum_{\mathbf{n}} \frac{\mathrm{erfc}(\eta|\,\mathbf{r}+\mathbf{n}|)}{|\,\mathbf{r}+\mathbf{n}|} + \sum_{\substack{\mathbf{k}\\ k\neq 0}} \frac{4\pi}{Vk^2}\, \mathrm{e}^{-k^2/4\eta^2 + i\mathbf{k}\cdot\mathbf{r}} - \frac{\pi}{V\eta^2} \quad (3)$$

η is a convergence parameter that is chosen to accelerate numerical convergence. Note that the value of $\varphi(\mathbf{r})$ is independent of η,[52]

$$\frac{\partial\varphi(\mathbf{r})}{\partial\eta} \equiv 0 \quad (4)$$

and that the average potential in the box is zero,[14,51,53,54]

$$\tilde{\varphi}(\mathbf{k}=0) = \int_V \mathrm{d}\mathbf{r}\varphi(\mathbf{r}) = 0 \quad (5)$$

In the following, we will separate the Coulomb energy U_{el} of the solute from the total Coulomb energy eq 1. When the system contains a solute with partial charges q_β at positions $\boldsymbol{r}_\beta$, its electrostatic interaction energy can be split into a solvent term and self-interactions,

$$U_{\mathrm{el}} = \sum_\beta \sum_i \sum_\alpha q_\beta q_{i\alpha} \varphi(\mathbf{r}_{\beta i\alpha}) + \sum_{\substack{\beta,\gamma\\ \beta<\gamma}} q_\beta q_\gamma \left[\varphi(\mathbf{r}_{\beta\gamma}) - \frac{1}{|\mathbf{r}_{\beta\gamma}|}\right] + \frac{1}{2}\sum_\beta q_\beta^2 \lim_{\mathbf{r}\to 0}\left[\varphi(\mathbf{r}) - \frac{1}{|\mathbf{r}|}\right] \quad (6)$$

The first, second, and third sum are the direct interactions with water, the interactions of charges on the solute with other solute charges, and the self-interactions of solute charges, respectively.

Ewald Potential $\varphi(\mathbf{r})$ Is the Solution of the Poisson Equation That Is Periodic with the Fundamental Period of the Simulation Cell.[14,51,53,55] A periodic solution of the Poisson equation requires that the surface integral of the electric field normal to the surface of the simulation cell be zero. This means that the material in the simulation cell must have zero net electric charge. If the physical system of interest is nonneutral, a uniform background distribution of charge is included to neutralize the nonzero charge of the physical system.

Consider an elementary charge. Because the Ewald potential is periodic, we can consider the Ewald electrostatic potential implied by centering the simulation cell on this elementary charge. By symmetry the Ewald normal electric field is zero on the cell boundary. The Ewald potential can thus be considered to be that of a cutoff on the cell boundary—the cutoff at the maximum distances achievable—with zero normal derivative analogous to a shifted-force correction.

Ewald Potential Pushes the Electrostatic Boundary outward as far as Possible but Still Retains Smoothness on the Boundary. The minimum image cutoff shares with the Ewald treatment the property that the electrostatic potential is not cutoff in any region of the simulation volume, the largest volume that must be physically considered. However, as demonstrated above, the Ewald potential is smooth on that boundary since it is the periodic solution of the Poisson equation. This is an important technical advantage that facilitates investigation of system size dependence of computed properties, i.e., the variation of system properties with variations of the cell boundaries. In those cases where the physical system of interest is nonneutral, it is a helpful point of view that the background charge density is a simple device that permits smoothness of the computed potential on the system boundary. We emphasize that the effect of the neutralizing background charge disappears as the thermodynamic limit is approached, in which the background charge density disappears.

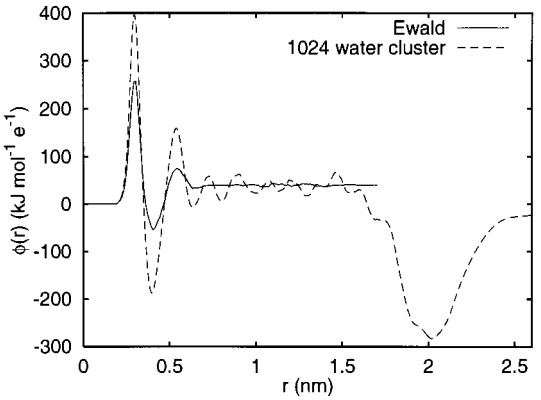

Figure 3. Electrostatic potential at the center of a neutral Lennard-Jones solute in SPC water[57] from simulations of a solute at the center of a cluster with 1024 water molecules (dashed line) and with periodic boundary conditions and Ewald summation (solid line). Shown is the potential obtained by integrating the charge density around the solute up to a distance R using $1/r$ (cluster) and $\varphi(\mathbf{r})$ (Ewald) for the Coulomb interactions. The results are those of ref 58.

Reaction Field Potentials Are Ewald Potentials for Different Background Charges. Reaction field methods[8−10,56] are computationally efficient alternatives to Ewald summation. The effective potentials of the site−site reaction field (SSRF) method[9] and the generalized reaction field (GRF) method[10] can be viewed as Ewald potentials for a nonhomogeneous background charge. The SSRF and GRF potentials are the solutions to the Poisson equation for a source charge and a compensating background. The background charge densities in the SSRF and GRF method are that of a homogeneous sphere and a radially symmetric charge distribution centered around the source charge, respectively. The SSRF and GRF method have a finite range because of the radially symmetric background charge densities that exactly compensate the source charge when the cutoff distance is reached.

Comparison of Ewald Potentials from Simulation of Water to Electrostatic Potentials in Isolated Water Droplets. It is interesting to make some simple numerical comparisons between the Ewald potentials that are experienced in simulation of water with periodic boundary conditions to the corresponding electrostatic potentials in water droplets. Such a comparison is simplified if we locate a distinguished solute at the origin of our Cartesian coordinate system. For a spherical Lennard-Jones solute, Figure 3 shows a typical variation of the electrostatic potential at the solute center with inclusion of the charge density in progressively larger spherical volumes of radius R around the solute. Notice the substantial variation of the electrostatic potential with inclusion of the solvation shells near the solute. However, after about three shells the net electrostatic potential oscillates about the Ewald asymptotic value before the ball penetrates the physical interface of the droplet. Thereafter, the net electrostatic potential displays the effects of the surface polarization of the droplet as it makes a transition to the very different value that characterizes the whole droplet. We observe that the Ewald potential faithfully captures this interior potential while avoiding detailed considerations of the droplet interface.

Single Ion Hydration Free Energies Are Well-Defined within Molecular Simulations. Electrostatic potentials can be defined unambiguously as solutions of the Poisson equation with specified charge densities and boundary conditions. Electrostatic potentials are computed throughout simulations of aqueous solutions even if charge densities and boundary conditions may not be specified explicitly. If the system of interest is nonneutral, these issues deserve emphasis because single ion free energies are typically not measured experimentally.

On a molecular scale, dependences on specifics of the boundary conditions in the definition of single ion hydration free energies can be avoided. This is accomplished by spherically integrating the electrostatic potential over the charge density around a charge site up to a distance where that potential saturates. Using Ewald interactions corresponds to such a spherical integration.[41,54] Figure 3 shows results for electrostatic potentials obtained using two different choices for the boundary conditions—a solute−water cluster and a periodic system. The observed agreement is a nontrivial computer experimental observation.

4. System Size Extrapolation

Computer simulations are performed for a finite system of molecules. In most applications, the properties of the thermodynamic-limit, infinite system are sought. In Coulombic systems, pronounced finite-size effects are ubiquitous due to the long range of the interactions. We can separate finite-size effects in Coulombic systems into two categories: (1) those caused purely by the long-range electrostatics independent of the thermodynamic state, and (2) finite size effects that depend on the thermodynamic state (temperature, pressure, etc.).

The electrostatic finite-size effects in the Ewald treatment of Coulomb interactions arise from self-interactions and interactions with the neutralizing background essential for nonneutral systems under periodic boundary conditions. Electrostatic finite-size effects can be treated exactly by including the second and third sum in the electrostatic energy U_{el} of the solute eq 6, which account for self-interactions of the solute.[13,14,53,59] For an ion of charge q, the resulting correction to the solvation chemical potential is[14,53]

$$\mu_{elec} = \mu_{sim} + \frac{q^2\xi}{2} \tag{7}$$

where μ_{sim} is the chemical potential for charging the ion from zero charge to net charge q calculated from the Ewald interactions with the solvent excluding self-interactions (i.e., including only the first sum in eq 6). μ_{elec} includes the self-interactions, and ξ is the ionic self-term. For a cubic box of length L, we have $\xi = \lim_{\mathbf{r}\to 0}[\varphi(\mathbf{r}) - 1/|\mathbf{r}|] \approx -2.837297/L$. Electrostatic finite size corrections for polar molecules are developed in ref 13. The corresponding free energy of changing partial charges located at positions $\mathbf{r}_\alpha$ on a molecule from q_α to q'_α is

$$\Delta\mu_{elec} = \Delta\mu_{sim} + \left\langle \frac{1}{2}\sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}} (q'_\alpha q'_\beta - q_\alpha q_\beta)\left[\varphi(\mathbf{r}_{\alpha\beta}) - \frac{1}{|\mathbf{r}_{\alpha\beta}|}\right]\right\rangle + \frac{1}{2}\sum_\alpha (q'^{\,2}_\alpha - q^2_\alpha)\xi \tag{8}$$

where $\langle ...\rangle$ denotes a canonical average. $\Delta\mu_{sim}$ includes only the energy difference corresponding to the first sum in eq 6, excluding self-interactions. Note that in constant pressure simulation with varying box volume, ξ must also be averaged.

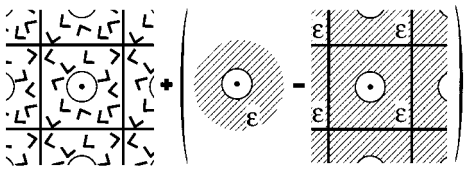

Figure 4. Schematic representation of the thermodynamic finite-size correction. The thermodynamic finite-size correction is the difference between an infinite Born model and a Born model under periodic boundary conditions. A spherical ion of charge q and radius R_B is embedded in a medium with a dielectric constant ϵ inside the simulation box. In addition, the box is filled with the neutralizing background charge. Periodic boundary conditions are applied. The corresponding electrostatic potential is determined from the Poisson equation with appropriate boundary conditions on the box boundary and ion surface.[58]

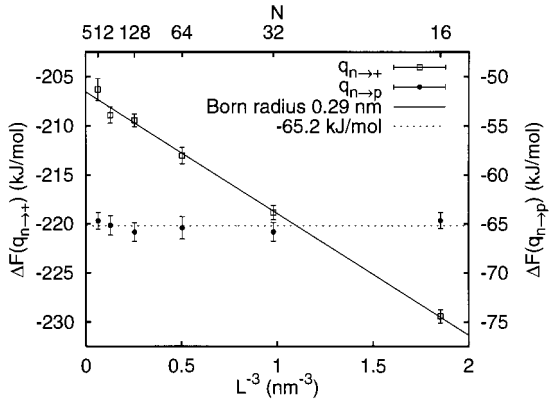

Figure 5. Finite-size dependence of the free energies of charging imidazole and imidazolium (filled circles and open squares on the right- and left-hand scale, respectively), as a function of the inverse volume of the simulation box, $1/L^3$. The top scale gives the number of water molecules.

Thermodynamic finite-size effects on the other hand can only be corrected approximately within a model. For instance, we can use the difference between a truly infinite version of a model and its finite periodic version to correct for thermodynamic finite size effects,[60−62] as schematically shown in Figure 4. For spherical ions, a Born model[63] and its periodic equivalent leads to finite size corrections that depend on the dielectric constant ϵ of the solvent, an effective Born radius R_B of the ion, and the net charge q of the ion:[60,64]

$$\mu_{therm} \approx \mu_{elec} + \frac{1}{2}q^2\left[\frac{-\xi}{\epsilon} + \frac{4\pi(\epsilon-1)R_B^{\,2}}{3\epsilon L^3}\right] \tag{9}$$

where μ_{therm} is the chemical potential for charging that includes the thermodynamic and electrostatic finite-size corrections.

Figure 5 illustrates the finite size effects for the free energies of charging Im(p) and Im(+) after correction for Ewald self-interactions. Free energies were calculated from sixth-order integration formulas with corrected means and variances from Table 2 that include the electrostatic[14] but not the thermodynamic finite-size correction. The free energy is plotted as a function of $1/L^3$ where L is the box length. We find that the free energy of charging the polar Im(p) is independent of the system size within the statistical errors of about 1 kJ/mol for $N = 16$ to $N = 512$ water molecules. However, the free energy of charging the Im(+) cation shows a system size dependence proportional to $1/L^3$, as would be expected from our finite size

TABLE 2: Averages C_1 (in kJ/mol) and Variances C_2 [in (kJ/mol)^2] of the Electrostatic Energy of Imidazole and Imidazolium in the Uncharged, Half-Charged, and Fully Charged States. Finite-Size Corrections Have Been Applied

	uncharged		half-charged		charged	
N	ave	var	ave	var	ave	var
			Imidazole			
16	1.2 ± 0.5	209.6 ± 7	−56.7 ± 1.3	436.8 ± 25	−156.9 ± 1.5	495.6 ± 32
32	0.5 ± 1.0	224.4 ± 7	−57. ± 1.6	427.4 ± 22	−158.7 ± 1.5	456.4 ± 26
64	1.8 ± 0.6	208.2 ± 7	−58.8 ± 2.0	510.0 ± 25	−155.5 ± 1.5	480.9 ± 15
128	1.0 ± 0.7	205.1 ± 9	−59.2 ± 1.6	429.5 ± 16	−156.1 ± 1.8	491.9 ± 20
256	1.3 ± 0.5	204.8 ± 7	−59.2 ± 1.6	429.1 ± 22	−152.9 ± 2.0	473.4 ± 24
512	1.1 ± 0.5	209.6 ± 9	−57.7 ± 1.3	410.8 ± 17	−155.9 ± 2.0	539.2 ± 30
			Imidazolium			
16	33.4 ± 1.3	1326.7 ± 10	−226.5 ± 1.0	1305.5 ± 7	−500.8 ± 1.0	1391.0 ± 14
32	33.9 ± 1.0	1267.6 ± 12	−215.0 ± 0.9	1267.0 ± 10	−483.4 ± 2.0	1375.0 ± 20
64	35.9 ± 1.1	1234.0 ± 12	−208.7 ± 1.3	1252.4 ± 13	−473.9 ± 1.5	1307.9 ± 18
128	34.9 ± 1.0	1226.0 ± 14	−203.5 ± 0.9	1215.9 ± 17	−469.1 ± 1.3	1291.2 ± 22
256	34.1 ± 1.1	1222.5 ± 22	−204.4 ± 1.2	1256.3 ± 24	−465.2 ± 1.5	1330.2 ± 33
512	35.9 ± 1.5	1248.8 ± 18	−201.2 ± 1.8	1248.3 ± 26	−461.7 ± 1.5	1305.3 ± 25

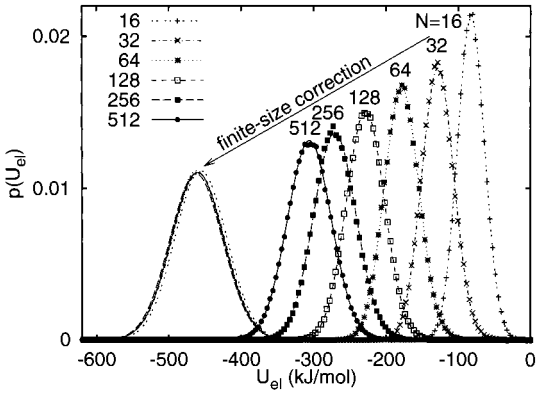

Figure 6. Finite-size correction of the probability densities $p(U_{el})$ of the electrostatic energies U_{el} of Im(+). The uncorrected U_{el} histograms are shown with symbols, together with corresponding Gaussian distributions. After correction for electrostatic and thermodynamic finite-size effects, the corresponding Gaussian distributions "collapse" and agree closely for all system sizes of $16 \le N \le 512$ water molecules.

analysis.[60] Rather than using a more realistic shape of the molecule in the dielectric model of ref 60, we fit the observed free energies to the spherical Born model eq 9 with an effective radius $R_B = 0.207$ nm and an infinite dielectric constant which reproduces the data over the whole range of system sizes of $16 \le N \le 512$ water molecules.

To further illustrate the power of the finite-size corrections, Figure 6 shows the probability distributions of electrostatic interaction energies U_{el} of the imidazolium cation Im(+) with the N water molecules, the first term of eq 6. Also shown in Figure 6 are the corresponding Gaussian distributions, which nicely reproduce the calculated histograms of U_{el}. However, we observe a strong system size dependence: small systems have narrower distributions of U_{el} with less negative averages compared to large systems. When we apply the electrostatic and thermodynamic finite size corrections for the mean and average, the Gaussian distributions "collapse" to a single distribution corresponding to the limit of an infinite system size. We note that the finite-size correction is large: for $N = 16$ water molecules, the average U_{el} changes by about −370 kJ/mol (150 k_BT).

5. Perturbation Theory

A fundamental view of the thermodynamics due to electrostatic interactions may be obtained from consideration of the distribution $p(u; \lambda = 0)$ of electrostatic energies u in the reference charge state $\lambda = 0$. The part of the chemical potential due to electrostatic interactions $\Delta\mu(\lambda)$, the thermodynamic parameter sought, is then expressed by the fundamental result

$$e^{-\beta\Delta\mu(\lambda)} = \langle e^{-\beta\lambda u}\rangle_{\lambda=0} = \int du\, p(u; \lambda = 0)\, e^{-\beta\lambda u} \quad (10)$$

Here, $\beta^{-1} = k_BT$ is Boltzmann's constant times the temperature and $\langle ...\rangle_{\lambda=0}$ denotes a thermal average with the solute in reference state $\lambda = 0$. This formula requires the consideration of the electrostatic potential even though the electrostatic potential of a phase is an operationally subtle property.[58,65] Despite that subtlety, the potential sought is conceptually well-defined as the solution of Poisson's equation with specified charge density and boundary conditions.[14,53,58,59,66]

Direct use of eq 10 can present difficulties. Though $p(u)$ is often substantially Gaussian, the fundamental formula eq 10 is sensitive to the tails of $p(u)$. That limits the applicability of eq 10 for calculations of even small changes in the charge state λ. In addition, the simple estimator $\ln\langle e^{-\beta\lambda u}\rangle_{\lambda=0} \approx \ln[M^{-1}\sum_{i=1}^{M} e^{-\beta\lambda u_i}]$ from M energies u_i observed in a simulation is biased and large sample sizes M are required for this bias to be negligible.[67]

Perturbation or cumulant expansions provide a technique to analyze these distributions.[12−16,68−73] A cumulant expansion[74] with respect to λ of eq 10 provides

$$\langle \exp(-\beta\lambda u)\rangle_{\lambda=0} = \exp\left[\sum_{n=0}^{\infty}(-\beta\lambda)^n\frac{C_n}{n!}\right] \quad (11)$$

This defines the cumulants C_n of order $n = 0, 1, 2$ as

$$C_0 = 0 \quad (12a)$$

$$C_1 = \langle u\rangle_{\lambda=0} \quad (12b)$$

$$C_2 = \langle (u - \langle u\rangle_{\lambda=0})^2\rangle_{\lambda=0} \quad (12c)$$

We interpret eq 11 as a Taylor expansion in λ but augment $\Delta\mu(\lambda)$ to include the self-contribution $(\lambda q)^2\xi/2$, where ξ vanishes in the thermodynamic limit but accounts for finite-size effects as discussed above. Then for the charging of an ion from a neutral reference condition we have

$$\Delta\mu(\lambda) = \lambda q\langle u\rangle_{\lambda=0} - \frac{(\lambda q)^2}{2}[\beta\langle (u - \langle u\rangle_{\lambda=0})^2\rangle_{\lambda=0} - \xi] + ... \quad (13)$$

This result should be compared to the Born[63] formula for the

hydration free energy due to electrostatic interactions of a spherical ion of radius R and charge λq:

$$\Delta\mu_B(\lambda) = -\frac{(\lambda q)^2}{2R}\left(\frac{\epsilon - 1}{\epsilon}\right) \quad (14)$$

where ϵ is the dielectric constant of the solvent. The matching of the second-order terms between the cumulant expansion and the continuum formula provides a determination of the radius R of a spherical ion. The distributions $p(u; \lambda = 0)$ required to evaluate the cumulant averages involve the nonelectrostatic interactions $\lambda = 0$ between solvent molecules and the solute. The indicated average thus generally depends upon full characterization of the solvent. Note that the continuum model neglects the molecular contribution linear in λ. This linear term contributes to the asymmetry between anion and cation solvation, making the solvation of anions more favorable for a given ion size.[14]

In principle, higher order cumulants could be used to obtain information about the other Taylor coefficients. However, as was observed by Smith and van Gunsteren,[69] higher-order cumulants are increasingly difficult to extract from computer simulations of limited duration. Though direct extension of perturbation theory beyond fourth order has been impractical, interpolative approximations polynomial in λ have been more successful. For the charging of water and ions, polynomials of order six and higher were necessary to account for the simulation data.[13,14,71] Thus, perturbation theory was found to be unsatisfactory in such cases. For atomic[14] and molecular ions,[24] a kink is typically observed for $d\Delta\mu(\lambda)/d\lambda$ as a function of charge λ at modest values of this parameter when the solvation shell changes from a cationic to an anionic structure. Additional nonlinearities were observed at high values of the ionic charge λ.[17,75]

Table 2 contains the averages and variances corrected for electrostatic finite-size effects of the electrostatic energies of Im(p) and Im(+) for system sizes between 16 and 512 water molecules. Errors of one standard deviation of the mean were estimated by plotting the block error as a function of the number of blocks. The estimated error reaches a plateau when block values are uncorrelated. From the averages and variances, we can calculate the chemical potentials of charging using integration formulas (*ijk*) exact to various orders that involve *i*, *j*, and *k* derivatives of the free energy with respect to the coupling parameter at the uncharged, half-charged, and fully charged state:[71]

$$\Delta\mu(010) \approx C_1(\lambda = 0.5) \quad (15a)$$

$$\Delta\mu(101) \approx \frac{1}{2}[C_1(\lambda = 0) + C_1(\lambda = 1)] \quad (15b)$$

$$\Delta\mu(111) \approx \frac{1}{6}[C_1(\lambda = 0) + 4C_1(\lambda = 0.5) + C_1(\lambda = 1)] \quad (15c)$$

$$\Delta\mu(202) \approx \Delta\mu(101) - \frac{\beta}{12}[C_2(\lambda = 0) - C_2(\lambda = 1)] \quad (15d)$$

$$\Delta\mu(212) \approx \frac{1}{30}[7C_1(\lambda = 0) + 16C_1(\lambda = 0.5) + 7C_1(\lambda = 1)] - \frac{\beta}{60}[C_2(\lambda = 0) - C_2(\lambda = 1)] \quad (15e)$$

where the cumulants C_1 and C_2 contain the electrostatic finite-size corrections. Formulas involving higher cumulants and different nodes λ_i are discussed in ref 71. The integration formulas eqs 15a−e are exact to order 2, 2, 4, 4, and 6 in a perturbation expansion, respectively.[70,71] Figure 7 shows the free energy difference between Im(+) and Im(p) as a function of the integration order for the 512 water molecule system. We find that as the order of the integration formula increases, the free energy difference converges, with the sixth order formula bracketed by the two fourth-order formulas. The statistical error of the free energy difference is about 1.5 kJ/mol. Notice that the discrepancy between the two second-order results of Figure 7 is significant on the scale of the statistical uncertainties. This emphasizes that the charging free energy is not a quadratic function of the coupling parameter. Note that the centered second-order formula[76] has a smaller systematic error.

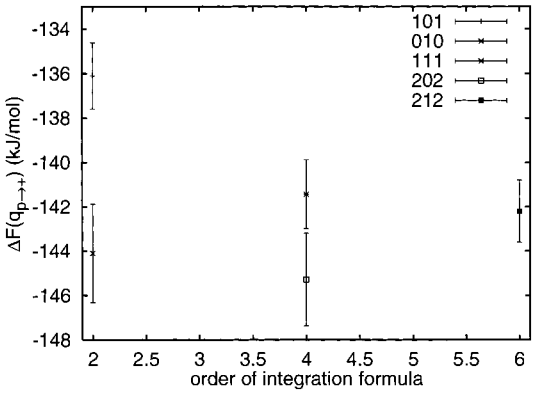

Figure 7. Free energy of charging the polar imidazole Im(p) to the imidazolium cation Im(+) as a function of the order of the integration formula. (*ijk*) indicates the number *i*, *j*, and *k* of derivatives used at the uncharged, half-charged, and fully charged state.[71]

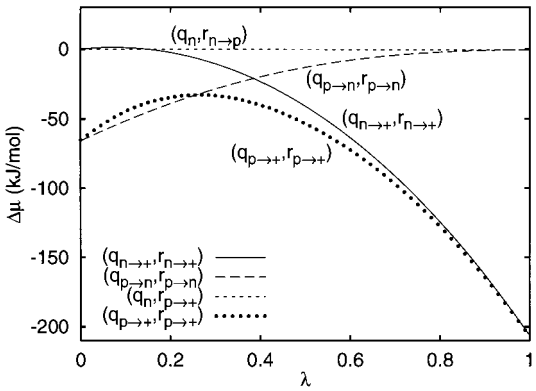

Figure 8. Thermodynamic cycle with the free energies connecting the four states of the uncharged and charged imidazole and imidazolium as a function of a linear coupling parameter. Shown are results for charging of imidazolium (solid line), uncharging of imidazole (long dashed line), conversion of imidazole to imidazolium (dotted line) and conversion of uncharged imidazole to uncharged imidazolium (short dashed line).

Figure 8 illustrates the complete four-node thermodynamic cycle, where λ is a coupling parameter changing the partial charges on the molecule linearly from state zero to one. The four nodes of the cycle are the uncharged and charged imidazole and imidazolium. We find that the free energies of charging and conformational changes are consistent within the statistical errors. Interestingly, the free energy of charging the polar Im(p) to the Im(+) cation has a maximum for the linear charging path chosen here. This increase reflects the linear terms of eq 13, i.e., increasing the net charge on the imidazole initially costs free energy.

Dielectric continuum models predict a quadratic proportionality of the free energy of charging on the linear coupling

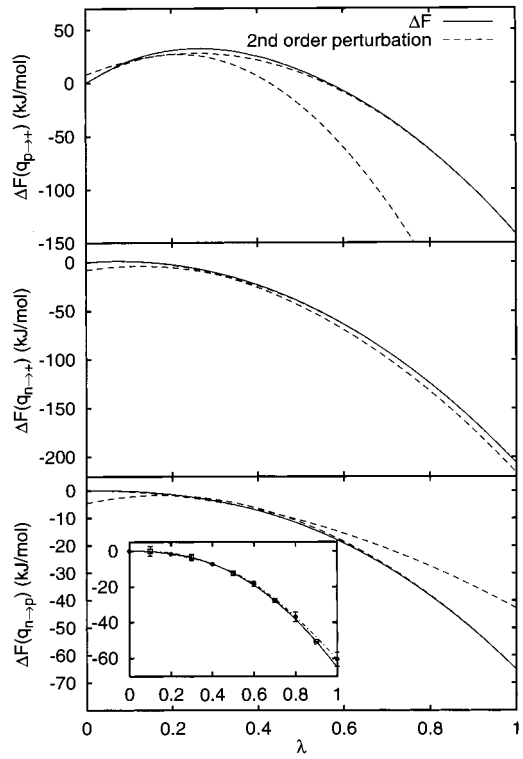

Figure 9. Comparison of the second-order perturbation expansion (dashed lines) with the reference free energies of charging Im(p) to Im(+) (top), uncharged Im(+) to the cationic Im(+) (middle), and uncharged Im(p) to the polar Im(p) (bottom). Also included as an inset in the bottom panel is a comparison with multistate Gaussian models (symbols and dot−dash lines) shown with estimated statistical errors. The multistate expansions about the charged and uncharged states are shown with open squares and filled circles, respectively.

parameter λ. In a molecular theory, such a quadratic charging free energy arises when the probability density of electrostatic potential fluctuations is Gaussian. Second-order perturbation theory would then be exact. Figure 9 compares second-order perturbation theory with the reference sixth-order free energy polynomial calculated from the averages and variances in Table 2. We find that the perturbation expansions about the charged state ($\lambda = 1$) are accurate over a relatively wide range from $\lambda = 1$ to almost $\lambda = 0.2$. The expansion about the uncharged state $\lambda = 0$ on the other hand breaks down rapidly at about $\lambda = 0.2$.

6. Non-Gaussian Fluctuations

Multistate Gaussian Models. One idea for improvement of dielectric models is based upon a physical description of the structure of the first hydration shell. That first hydration shell can be viewed from the perspective of Stillinger−Weber inherent structures or substates.[77] These are potential energy basins of attraction for steepest descent energy minimization. If those first hydration shell molecules stayed always in one basin, then a Gaussian model for thermal fluctuations would be reasonable. Empirical radii parameters reflect the characteristics of that single basin. However, changing conditions may result in reweighting of slightly accessible basins or the opening of new basins. The Gaussian or dielectric models may fail to describe these possibilities well. This picture is physically better defined than the commonly nonspecific discussions of electrostriction and dielectric saturation.

A corresponding "multistate Gaussian model" was developed in ref 21. Attention is directed to the thermal probability distribution of electrostatic potential energies of the solute. Rather than approximating this distribution as a single Gaussian distribution, perhaps with perturbative corrections, we discriminate hydration structure on the basis of simple parameters diagnostic of hydration substates. We assume that the probability distribution of electrostatic potential energies is Gaussian for each substate. Therefore, the full distribution is a superposition of Gaussian distributions for the various substates.

Thus we attempt to represent the observed complicating features of $p(u)$ by a combination of simpler states:

$$p(u) = \sum_n w_n p_n(u) \tag{16}$$

with weights $w_n \geq 0$, $\sum_n w_n = 1$ and normalized densities $p_n(u) \geq 0$, $\int du\, p_n(u) = 1$. We will seek $p_n(u)$'s of Gaussian form, representing the overall system as a linear combination of Gaussian subsystems, each showing linear response to electrostatic interactions. Representing $p(u)$ by a sum of Gaussian densities can give nontrivial results for the chemical potential, as can be seen by substituting eq 16 into eq 10,

$$\Delta\mu(\lambda) = -k_B T \ln \sum_n w_n e^{-\beta\lambda m_n + \beta^2\lambda^2\sigma_n^2/2} \tag{17}$$

where m_n and σ_n^2 are the mean and variance of the Gaussian p_n, respectively.

The non-Gaussian fluctuations of the electrostatic potential in liquid water are associated with changes in the conformations of protons that make hydrogen bonds to the solute. If those fluctuations could be tempered, a Gaussian model might become more accurate. Thus, suitable substate diagnostic parameters are the number of hydrogen bonds made to the solute.

Explicit calculations have shown that this approach eliminates most of the detailed numerical inaccuracies of the Gaussian fluctuation models for hydration of a water molecule in liquid water.[21] The markedly non-Gaussian $p(u)$ was accurately represented as the sum of Gaussian distributions implied by this definition of a hydration substate. We found $w_n > 10^{-3}$ for $1 \leq n \leq 6$ with 3.64 being the average number of neighbors and $n = 4$ the most probable number of neighbors. The calculated change of the chemical potential upon change of the charge state of a solute water molecule is correct to within 5%. This is a remarkable result because $\Delta\mu(\lambda)$ is nonquadratic, requiring an eighth-order polynomial to fit the simulation data for chemical-potential derivatives.[13,14,71] This shows that sufficient information can be extracted from the simulation to describe the distribution $p(u)$ helpfully and that such an approach can be successful even for perturbations involving changes of the chemical potential as large as 14 $k_B T$.

Similar behavior can be anticipated for hydration of other neutral, polar solutes such as the imidazole example studied here. Figure 9 (inset) shows the results of the multistate Gaussian model applied to charging and uncharging the polar imidazole Im(p). Fluctuation data were collected from a simulation of the uncharged and charged Im(p) in $N = 128$ water molecules, extending over 10^6 MC passes to allow for error estimates. Instead of determining the overall mean and variance of the electrostatic potential for a second-order perturbation expansion, we calculate the means and variances for several Gaussian distributions from structures sorted according to the number of hydrogen bonds. Inspection of the radial distribution functions of water oxygen and hydrogen around imidazole sites shows one strong hydrogen bond donor, H1, and one acceptor, N3. As a criterion for the formation of a water−imidazole

hydrogen bond, we used that the distance between the acceptor (nitrogen N3 or water oxygen) and the donor hydrogen has to be smaller than 0.23 nm. We can then sort structures according to the numbers of hydrogen bonds accepted and donated by the solute. With this simple criterion, we find that six and two Gaussian distributions contribute to the expansions about the charged and uncharged state of Im(p), respectively. This multistate Gaussian model greatly improves the quality of the expansions, both about the charged and uncharged state. The reference free energy is now within the statistical errors of the two multistate Gaussian models over the whole range $0 \leq \lambda \leq 1$.

Quasi-Chemical Theories. Those difficult anionic cases mentioned above can be attacked more directly. The local neighborhood is again used to discriminate structural possibilities. But, in addition, the consequences for the hydration free energy of the molecular interactions within that neighborhood are treated fully. This reserves the longer ranged interactions for simple approximations, e.g., with Gaussian models.

These theoretical developments arose from recent molecular calculations[23] that suggested how a chemical perspective can be helpful in computing thermodynamic properties of water and aqueous solutions. That calculation used electronic structure results on the $Fe(H_2O)_6^{3+}$ cluster and simple, physical estimates of further solvation effects. The results were organized according to the pattern of a simple chemical reaction and a surprisingly accurate evaluation of the hydration free energy was obtained. Despite this recent motivation, the theories developed are akin to good approximations of historical and pedagogical importance in the areas of cooperative phenomena and phase transitions.[78] In those areas, similar approximations are called Guggenheim, quasi-chemical, or Bethe approximations.

These quasi-chemical theories[22] are constructed by considering a geometric volume fixed on the solute molecule and performing a calculation analogous to the evaluation of a grand canonical partition function for that volume. All the possibilities for occupancy of that volume by solvent or other solution species must be considered eventually. The final result can be described by reference to simple patterns of chemical equilibria such as

$$X^{q-} + nH_2O \rightleftharpoons X(H_2O)_n^{q-} \tag{18}$$

The solute of interest is denoted here generically as X^{q-}; a star-type cluster of that solute with n water (W) molecules is considered as the product. For such a cluster, call it an M-cluster, we could calculate the equilibrium ratio K_M for a dilute gas phase. We adopt the convention that $K_M = 1$ for the $n = 0$ case that zero ligands are involved. Note that $K_M\rho_W{}^n$ is dimensionless and this observation resolves standard state issues. The factors denoted by $\langle \exp\{-\beta\Delta u\}\rangle_{0,C}$, where C indicates either a water molecule or the M-cluster, carry information about the solvation free energy of the species involved. For the species other than the cluster this is the familiar Widom factor. For the cluster, this factor requires a slight additional restriction. It is the average of the Boltzmann factor for cluster−solution interactions over the thermal motion of the cluster and solution under the condition that the only interactions between these subsystems rigidly exclude additional solvent molecules from the cluster volume for the complex. This restriction enforces a constraint required to preserve simple enumerations that underlie these results. The theoretical structure is designed so that simple approximations such as dielectric models might be used for the factors $\langle \exp\{-\beta\Delta u\}\rangle_{0,C}$. But more detailed techniques might be applied to the calculation of K_M. Finally, we compile

$$\tilde{K}_M \equiv K_M \frac{\langle \exp\{-\beta\Delta u\}\rangle_{0,M}}{[\langle \exp\{-\beta\Delta u\}\rangle_{0,W}]^n} \tag{19}$$

Thus, this $\tilde{K}_M$ is built on the pattern of the chemical equilibrium eq 18 but without a "solvation factor" for the reactant solute indicated on the left.

Now consider all possibilities for clusters. A thermodynamic implication of this information is

$$\mu_{X^{q-}} = kT \ln[\rho_{X^{q-}}V/q_{X^{q-}}] - kT \ln[p_0 \sum_M \tilde{K}_M \rho_W{}^n] \tag{20}$$

p_0 is the probability that the clustering volume would be observed to be *empty* in the equilibrium solution; thus, $-kT \ln p_0$ is the free energy for formation of a cavity for the clustering volume in the solution. The sum is over all clusters with zero or more ligands. The product of the densities involved with each term includes a density factor for each ligand. This formula makes the conventional separation between the contributions of intermolecular interactions and the noninteraction (ideal) terms; $q_{X^{q-}}$ is the partition function of the bare solute in the absence of interactions with any other species and $\rho_{X^{q-}}$ is the density of the solute. As an example, for an atomic ion such as the chloride ion Cl^- we would put $q_{X^{q-}} = V/\Lambda^3$ with V the volume of the system and Λ the thermal deBroglie wavelength of the chloride ion. This formula becomes approximate when approximate models are adopted for p_0, for K_M, and for the solvation factors. Those quantities depend on definition of the clustering volume. But, since the physical problem is independent of those parameters, the theory should be insensitive to them.

The motivation for this approach is that a substantial but intricate part of the free energy sought is to be found in K_M. The number of possibilities for ligand populations will be small for molecular scale clustering volumes. So a limited number of terms must be considered. Because the clusters will be small systems, elaborate computational methods can be applied to the prediction of the K_M, including current electronic structure techniques. With the complicated chemical interactions separated for individual treatment the remaining hydration contributions should be simpler and the required theories better controlled.

Equation 20 should be compared with eq 17. One difference is that eq 20 attempts to provide the whole hydration free energy, not just the part due to electrostatic interactions. That explains the presence of p_0 in eq 20. Beyond that, the structures of these formulas are similar. The presence of more than one term in the sum of eq 20 is an expression of an entropy contribution associated with the possibilities for different ligand populations. Finally, the complete calculation of the K_M includes non-Gaussian statistical possibilities not anticipated by eq 17.

7. Conclusions

Recent calculations of the hydration free energy due to electrostatic interactions between charged and polar solutes in water have obtained high accuracy results for the simple molecular models that are the basis of most simulation calculations.[13,14,21,62,79,80] An important step in securing those high accuracy results has been a careful consideration of treatment of long-ranged interactions. That work suggests that *the Ewald method is an easy way to get correct hydration free energies* from molecular calculations, that is, to achieve well-characterized results appropriate to the thermodynamic limit in which the system size tends to infinity for given densities and temperature. Additionally, it suggests that molecular simulations

with Ewald potentials and periodic boundary conditions can have efficiencies comparable to rougher approximations that are often employed to compute hydration free energies for molecularly well-defined problems.[81] And furthermore, this has produced a simple, effective, and clear understanding of how to extrapolate electrostatic hydration free energies to the thermodynamic limit; an accurate evaluation of the hydration free energies of imidazole and imidazolium can be obtained with as few as 16 water molecules included in the simulation.

These high accuracy results for well-defined models permitted careful testing of simple theories of electrostatic hydration free energies. The simplest theories, dielectric continuum models, have been found to be rough despite the fact that they can always be adjusted to reproduce an empirical answer given a priori. Such a conclusion has surely been widely expected. However, the testing has led to new theories, the multistate Gaussian and quasi-chemical theories, that should permit more revealing molecular scale calculations. The quasi-chemical approaches seem to provide the most natural way to utilize current electronic structure packages to study electronic structure issues for solution species. This should be particularly helpful for treatment of basic, molecular anions that are ubiquitous in aqueous solution chemistry.

Physical conclusions more specifically are that the most prominent failings of the simplest theories are associated with solvent proton conformations that lead to non-Gaussian fluctuations of electrostatic potentials. Thus, the most favorable cases for the second-order perturbation theories are monoatomic positive ions. In such cases, oxygen−hydrogen bonds are oriented away from the ion, placing those protons as far out as possible. Neutral, polar molecules that may form specific hydrogen bonds with the solvent are more difficult for these theories, though the hydration free energies sought are smaller in magnitude. Negative molecular ions are expected to offer further complications because now the problematic proton motions occur close to the solute and the hydration effects will be larger for anionic species.

Acknowledgment. This work was supported by the LDRD program at Los Alamos. We thank G. J. Tawa for providing unpublished results of his calculations on imidazole and imidazolium. G.H. thanks Dr. Attila Szabo and Prof. Martin Neumann for many useful discussions and collaborations.

References and Notes

(1) Stell, G.; Friedman, H. L., Technical Report No. LBL-10634; University of California: Lawrence Berkeley Laboratory, Livermore, CA, 1980 (unpublished)..
(2) Ewald, P. P. *Ann. Phys. (Leipzig)* **1921**, *64*, 253.
(3) Slattery, W. L.; Doolen, G. D.; DeWitt, H. E. *Phys. Rev. A* **1980**, *21*, 2087.
(4) Lekner, J. *Phys. A* **1991**, *176*, 485.
(5) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(6) Schmidt, K. E.; Lee, M. A. *J. Stat. Phys.* **1991**, *63*, 1223.
(7) Pollock, E. L.; Glosli, J. *Comput. Phys. Commun.* **1996**, *95*, 93.
(8) Barker, J. A.; Watts, R. O. *Mol. Phys.* **1973**, *26*, 789.
(9) Hummer, G.; Soumpasis, D. M.; Neumann, M. *Mol. Phys.* **1992**, *77*, 769.
(10) Hummer, G.; Soumpasis, D. M.; Neumann, M. *J. Phys.: Condens. Matter* **1994**, *6*, A141.
(11) Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216.
(12) Pratt, L. R.; Hummer, G.; García, A. E. *Biophys. Chem.* **1994**, *51*, 147.
(13) Hummer, G.; Pratt, L. R.; García, A. E. *J. Phys. Chem.* **1995**, *99*, 14188.
(14) Hummer, G.; Pratt, L. R.; García, A. E. *J. Phys. Chem.* **1996**, *100*, 1206.
(15) Levy, R. M.; Belhadj, M.; Kitchen, D. B. *J. Chem. Phys.* **1991**, *95*, 3627.
(16) Hwang, J. K.; Warshel, A. *J. Am. Chem. Soc.* **1987**, *109*, 715.
(17) Jayaram, B.; Fine, R.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1989**, *93*, 4320.
(18) Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512.
(19) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1994**, *116*, 3949.
(20) Tawa, G. J.; Pratt, L. R. *J. Am. Chem. Soc.* **1995**, *117*, 1625.
(21) Hummer, G.; Pratt, L. R.; García, A. E. *J. Am. Chem. Soc.* **1997**, *119*, 8523.
(22) Pratt, L. R.; LaViolette, R. A. *Mol. Phys.* **1998**, *94*, 909.
(23) Martin, R. L.; Hay, P. J.; Pratt, L. R. *J. Phys. Chem. A* **1998**, *102*, 3565.
(24) Garde, S.; Hummer, G.; Paulaitis, M. E. *J. Chem. Phys.* **1998**, *108*, 1552.
(25) Warshel, A.; Sussman, F.; King, G. *Biochem.* **1986**, *25*, 8368.
(26) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219.
(27) Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 3572.
(28) Delbuono, G. S.; Figueirido, F. E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 85.
(29) Saito, M. *J. Phys. Chem.* **1995**, *99*, 17043.
(30) You, T. J.; Bashford, D. *Biophys. J.* **1995**, *69*, 1721.
(31) Figueirido, F.; Del Buono, G. S.; Levy, R. M. *J. Phys. Chem.* **1996**, *100*, 6389.
(32) Ganellin, C. R. In *Molecular and Quantum Pharmacology*, *Proceedings of the Seventh Jerusalem Symposium on Quantum Chemistry and Biochemistry*; Bergmann, E. D., Pullman, B., Eds.; Reidel: Dordrecht, 1974; pp 43−53.
(33) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley: New York, 1991.
(34) Rashin, A. A.; Rabinowitz, J. R.; Banfelder, J. R. *J. Am. Chem. Soc.* **1990**, *112*, 4133.
(35) Nagy, P. I.; Durant, G. J.; Smith, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 2912.
(36) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Phys. Chem. A* **1997**, *101*, 10075.
(37) Bash, P. A.; Ho, L. L.; Mackerell, A. D.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698.
(38) Ho, L. L.; Mackerell, A. D.; Bash, P. A. *J. Phys. Chem.* **1996**, *100*, 4466.
(39) Cornell, W. D.; Cieplak, P.; Bayley, C. I.; Gould, I. R.; Merz, Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(40) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.
(41) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. London A* **1980**, *373*, 27.
(42) Hummer, G.; Grønbech-Jensen, N.; Neumann, M. *J. Chem. Phys.* **1998**, *109*, 2791.
(43) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(44) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
(45) Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253.
(46) Pratt, L. R.; Tawa, G. J.; Hummer, G.; García, A. E.; Corcelli, S. A. *Int. J. Quantum. Chem.* **1997**, *64*, 121.
(47) Ashcroft, N. W.; Mermin, N. D. *Solid State Physics*; Saunders College: Philadelphia, 1976.
(48) Peierls, R. *Surprises in Theoretical Physics*; Princeton University Press: Princeton, NJ, 1979.
(49) Pratt, L. R.; Haan, S. W. *J. Chem. Phys.* **1981**, *74*, 1864.
(50) Pratt, L. R.; Haan, S. W. *J. Chem. Phys.* **1981**, *74*, 1873.
(51) Brush, S. G.; Sahlin, H. L.; Teller, E. *J. Chem. Phys.* **1966**, *45*, 2102.
(52) Hummer, G. *Chem. Phys. Lett.* **1995**, *235*, 297.
(53) Hummer, G.; Soumpasis, D. M. *J. Chem. Phys.* **1993**, *98*, 581.
(54) Nijboer, B. R. A.; Ruijgrok, T. W. *J. Stat. Phys.* **1988**, *53*, 361.
(55) Cichocki, B.; Felderhof, B. U.; Hinsen, K. *Phys. Rev. A* **1989**, *39*, 5350.
(56) Neumann, M. *Mol. Phys.* **1983**, *50*, 841.
(57) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces: Proceedings of the 14th Jerusalem Symposium on Quantum Chemistry and Biochemistry*; In Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1981; pp 331−342.
(58) Hummer, G.; Pratt, L. R.; García, A. E.; Berne, B. J.; Rick, S. W. *J. Phys. Chem. B* **1997**, *101*, 3017.
(59) Figueirido, F.; Del Buono, G. S.; Levy, R. M. *J. Chem. Phys.* **1995**, *103*, 6133.
(60) Hummer, G.; Pratt, L. R.; García, A. E. *J. Chem. Phys.* **1997**, *107*, 9275.
(61) Figueirido, F.; Del Buono, G. S.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101*, 5622.
(62) Lynden-Bell, R. M.; Rasaiah, J. C. *J. Chem. Phys.* **1997**, *107*, 1981.
(63) Born, M. *Z. Phys.* **1920**, *1*, 45.
(64) Note the factor 4π in eq 9, corrected from 2π in the original publication.[60] A consequence of this correction is that, for sodium ions,[60]

the ion radius entering Born's free energy formula and the effective radius R_B entering the correction formula eq 9 are not identical. Accordingly, the solute-size dependent term in the thermodynamic finite-size correction eq 9 is best seen as giving the dimensional behavior proportional to $1/L^3$, with the solute size R_B as a parameter.

(65) Ashbaugh, H. S.; Wood, R. H. *J. Chem. Phys.* **1997**, *106*, 8135.

(66) Hummer, G.; Pratt, L. R.; García, A. E.; Garde, S.; Berne, B. J.; Rick, S. W. *J. Phys. Chem. B* **1998**. In press.

(67) Wood, R. H.; Mühlbauer, W. C. F.; Thompson, P. T. *J. Phys. Chem.* **1991**, *95*, 6670.

(68) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(69) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 577.

(70) Zhou, H.-X.; Szabo, A. *J. Chem. Phys.* **1995**, *103*, 3481.

(71) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004.

(72) Liu, H.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1996**, *100*, 9485.

(73) Archontis, G.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 11246.

(74) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100.

(75) Figueirido, F.; Del Buono, G. S.; Levy, R. M. *Biophys. Chem.* **1994**, *51*, 235.

(76) King, G.; Barford, R. A. *J. Phys. Chem.* **1993**, *97*, 8798.

(77) Stillinger, F. H.; Weber, T. A. *Phys. Rev. A* **1982**, *25*, 978.

(78) Brush, S. G.; Kikuchi, R. Technical Report No. UCRL-14287; Lawrence Radiation Laboratory, University of California: Livermore, CA, 1965 (unpublished).

(79) Kalko, S. G.; Sese, G.; Padro, J. A. *J. Chem. Phys.* **1996**, *104*, 9578.

(80) Sakane, S.; Ashbaugh, H. S.; Wood, R. H. *J. Phys. Chem. B* **1998**, *102*, 5673.

(81) The total CPU time of all imidazole and imidazolium simulations with $N = 16$ and 32 water molecules was, including equilibrations, 4 and 8 h on a single SGI R10000 chip, respectively. By extrapolating the $N = 16$ and 32 results to $N \rightarrow \infty$, we obtain an effective Born radius of 0.30 nm and a solvation free energy difference between Im(+) and Im(p) that is within 1.2 kJ/mol of the fit to all data for $16 \leq N \leq 512$.

(82) Several related works have become available since the writing of this paper. Hünenberger, P. H.; van Gunsteren, W. F. *J. Chem. Phys.* **1998**, *108*, 6117. Bogusz, S.; Cheatham, T. E., III; Brooks, B. R. *J. Chem. Phys.* **1998**, *108*, 7070. Levy, R. M.; Gallicchio, E. *Ann. Rev. Phys. Chem.* **1998**. Submitted for publication. Darden, T.; Pearlman, D.; Pedersen, L. **1998**. Preprint. Sham, Y. Y.; Warshel, A. **1998**. Preprint.