PHYSICAL REVIEW E VOLUME 52, NUMBER 4 OCTOBER 1995

Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach

Alexander P. Lyubartsev* and Aatto Laaksonen
Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S-109 61, Stockholm, Sweden

(Received 15 March 1995; revised manuscript received 5 July 1995)

An approach is presented to solve the reverse problem of statistical mechanics: reconstruction of interaction potentials from radial distribution functions. The method consists of the iterative adjustment of the interaction potential to known radial distribution functions using a Monte Carlo simulation technique and statistical-mechanics relations to connect deviations of canonical averages with Hamiltonian parameters. The method is applied to calculate the effective interaction potentials between the ions in aqueous NaCl solutions at two different concentrations. The reference ion-ion radial distribution functions, calculated in separate molecular dynamics simulations with water molecules, are reproduced in Monte Carlo simulations, using the effective interaction potentials for the hydrated ions. Application of the present method should provide an effective and economical way to simulate equilibrium properties for very large molecular systems (e.g., polyelectrolytes) in the presence of hydrated ions, as well as to offer an approach to reduce a complexity in studies of various associated and aggregated systems in solution.

PACS number(s): 61.20.Ja, 05.20.Gg, 02.70.−c, 61.20.Qg

I. INTRODUCTION

Computer simulations, such as Monte Carlo (MC) or molecular dynamics (MD) methods, are important tools, and routinely used to study molecular systems in condensed phases and solutions [1,2]. In fact, for a given Hamiltonian, these methods can supply numerical solutions to the main problem in equilibrium statistical mechanics by calculating statistical ensemble averages for experimentally observable quantities. While the approximations inherent in the simulation procedure itself (e.g., limited sampling of the configurational space, finite size of the simulation box, etc.) are usually insignificant and under control, the choice of the Hamiltonian or the interaction potentials is often questionable. Today in most computer simulations so called effective site-site pair potentials are used, which are parametrized to reproduce some set of experimental data. The interaction potentials obtained from quantum mechanical calculations are in general more detailed and possibly also more reliable than the empirically parametrized simple pair potentials. However, these *ab initio* potentials are more expensive to use due to their complicated functional form and can be, so far, calculated only for small molecular systems. An effective computational scheme for reconstruction of the Hamiltonian from an available set of canonical averages, the reverse problem, would provide an additional choice and should be of great importance to the field of molecular simulations.

Recently, a reverse MC procedure has been suggested [3], in which the starting point was a radial distribution function (RDF) obtained from experimental structure factor data. No input potential was required in this method, and the simulation was carried out to minimize differences between calculated and reference averages. The reverse MC simulation produced a set of configurations for a further analysis. Because the Hamiltonian was not reconstructed during the simulation, the reverse problem was not completely solved.

A similar problem of adjusting of potential (force field) parameters to experimentally observed properties was considered in a very recent paper [4]. It was suggested that the method of weak coupling would allow one to fit one or several potential parameters if an analytical form of the potential was selected.

The main objective of this paper is to present a method to reconstruct the Hamiltonian from radial distribution functions. Clearly, in a general case the solution of this problem is not unique. However, if we consider a limited class of Hamiltonians (e.g., those represented by a sum of pair interactions), the solution appears fairly well defined. Below we present a method of automatical adjustment of the pair interaction potential, irrespective of its analytical form, to known radial distribution functions.

Besides presenting a way to solve the reverse problem, we also suggest a few important applications for effective potentials as interaction potentials between atom groups and molecules. In spite of the fact that computers are becoming faster and faster each year and will continue to do so for some time to come through parallel architectures, it will only be practical to simulate systems consisting of 10^3-10^4 particles during a few hundred picoseconds in the near future. Complex biomolecules and organic molecules (e.g., proteins, nucleic acids, membranes, carbohydrates, etc.) should be simulated by ex-

---

[*] On leave from Scientific Research Institute of Physics, St. Petersburg State University, 198904 St. Petersburg, Russia.

plicitly including the solvent molecules, which increases the number of interactions by several orders of magnitude. The larger the molecular systems grow, the longer the simulations needed to follow low-amplitude motions and slow conformational transitions. In other words, it is too demanding a task today to carry out such simulations with an explicit account of all the atoms, and therefore simplified models may have to be used in order to compromise. For example, in the primitive electrolyte model, which is often used for simulations of ion environment around biomolecules, the ions are modeled as hydrated with several water molecules in the hydration shell. The solvent in the primitive electrolyte model is considered as a dielectric medium with a permittivity ε. Of course, the primitive model potential $1/(\varepsilon r)$ is a rather crude approximation, especially at short distances between the ions. The effective interionic potential, accounting for the molecular nature of the solvent, would be a good replacement, instead of the primitive model potential in studies of these systems.

It is, in principle, possible to derive the effective potential for a simulation of a given (reduced) system from the simulation of a more detailed (full) system. The reduced system has fewer degrees of freedom. A simulation of the full system yields a representative sample of the statistical ensemble. After removing the noninteresting degrees of freedom (e.g., the coordinates of solvent molecules), we obtain a set of configurations of the reduced system. The task is to find an effective potential which reproduces this set of configurations and corresponding canonical averages as close as possible. Once again, we need to solve the reverse problem: reconstruct the interaction potential from canonical averages. Although different canonical averages may be chosen for the purpose, the RDF is a good choice. The RDF's give a detailed definition of the structural properties of the system. They can be routinely determined in computer simulations of the full system, and experimental RDF's are available for a large number of liquids of small molecules from neutron and x-ray diffraction studies. The basic idea is that the solution of the reverse problem gives an effective potential which can subsequently be used for simulations of reduced systems on a larger scale. For example to build up large biomolecular assemblies in solutions.

The idea of the method, presented in this paper, goes back to the renormalization group Monte Carlo method for phase transition studies in the Ising model by Swendsen and co-workers [5,6]. The algorithm was used to compute the renormalizing constants (an analog to the potential) for the blocked spins. It now appears that the applications of this method are more general than the lattice systems near the phase transition point. It will be shown below that it is possible to renormalize the Hamiltonian of a molecular system in such a way that only the degrees of freedom of primary interest are kept.

A general algorithm for finding the effective interaction potential from RDF will be presented in Sec. II. Section III contains an example of some successive direct and reverse MC simulations for the one-component model, in which the initial interaction potential is restored. In Sec. IV, this technique is applied to a calculation of the effective potential for aqueous NaCl solutions at two concentrations. The reference ion-ion RDF curves are calculated in MD simulations of the NaCl solution, with water molecules explicitly included. A brief discussion of the method is finally given in Sec. V.

II. THEORETICAL BACKGROUND

Consider a system with a Hamiltonian (potential energy) given as

$$H\{q_i\}=\sum_{\alpha} K_\alpha S_\alpha\{q_i\} , \tag{1}$$

where $S_\alpha\{q_i\}$ are functions of particle coordinates q_i, and K_α are constants defining the interaction potential. The summation in Eq. (1) may also be replaced by an integral.

The Hamiltonian of a system with pair interactions can be given in the same fashion as Eq. (1):

$$\begin{aligned} H=\sum_{i,k}\Psi(|q_i-q_k|)&=\sum_{i,k}\int_0^\infty \Psi(r)\delta(r-|q_i-q_k|)dr \\ &=\int_0^\infty \Psi(r)\sum_{i,k}\delta(r-|q_i-q_k|)dr . \end{aligned} \tag{2}$$

In comparison with Eq. (1), the sum is now replaced by an integral, α by r, K_α by $\Psi(r)$, and $S_\alpha\{q_i\}$ by $\sum_{i,k}\delta(r-|q_i-q_k|)$.

Generalization to systems with several particle types is straightforward. It is possible to extend the Hamiltonian, given in Eq. (1), for systems with three-particle interactions in an analogous way.

The Hamiltonian in Eq. (1) is defined by a set of constants K_α. These constants span a space of Hamiltonians of a given class, determined by the form of $S_\alpha\{q_i\}$. These Hamiltonians may be considered as equivalent if they have the same canonical averages $\langle S_\alpha\{q_i\}\rangle$ for each α. For systems defined by pair interactions [Eq. (2)] this means the coincidence of the radial distribution functions $\rho(r)$, since $\langle S_r\rangle=4\pi r^2\rho(r)$. The averages $\langle S_\alpha\rangle$ are functions of constants $\{K_\alpha\}$. In order to solve the reverse problem, we have to obtain a set of constants $\{K_\alpha\}$ from some averages $\langle S_\alpha\rangle=S_\alpha^*$. The averages can be calculated from experimental structure data (structure factors) or from computer simulations of the full system.

In the vicinity of an arbitrary point in the space of Hamiltonians, $\{K_\alpha\}$, we can write

$$\Delta\langle S_\alpha\rangle=\sum_{\gamma}\frac{\partial\langle S_\alpha\rangle}{\partial K_\gamma}\Delta K_\gamma+O(\Delta K^2) , \tag{3}$$

where the derivatives $\partial\langle S_\alpha\rangle/\partial K_\gamma$ are expressed as

$$\begin{aligned} \frac{\partial\langle S_\alpha\rangle}{\partial K_\gamma}&=\frac{\partial}{\partial K_\gamma}\left[\frac{\int dq\, S_\alpha(q)\exp\left[-\beta\sum_{\lambda}K_\lambda S_\lambda(q)\right]}{\int dq\,\exp\left[-\beta\sum_{\lambda}K_\lambda S_\lambda(q)\right]}\right] \\ &=-\beta(\langle S_\alpha S_\gamma\rangle-\langle S_\alpha\rangle\langle S_\gamma\rangle) , \end{aligned} \tag{4}$$

and q is the set of degrees of freedom of the reduced system.

Let $K_\alpha^{(0)}$ denote a set of trial values for the constants K_α. By carrying out a MC simulation using these values $K_\alpha^{(0)}$, a set of averages $\langle S_\alpha^{(0)} \rangle$ is obtained. The differences between the trial values and the reference values are $\Delta\langle S_\alpha \rangle^{(0)} = \langle S_\alpha^{(0)} \rangle - S_\alpha^*$. Then, by solving the system of linear equations, given in Eq. (3), with appropriate coefficients calculated from Eq. (4), and by omitting terms of order $O(\Delta K^2)$, we obtain $\Delta K_\alpha^{(0)}$ and use them as corrections to the potential according to Eq. (5):

$$K_\alpha^{(1)} = K_\alpha^{(0)} + \Delta K_\alpha^{(0)} \ . \tag{5}$$

The MC simulation is then repeated with another potential $K_\alpha^{(1)}$ to determine a set of corrections $\Delta K_\alpha^{(1)}$. The procedure is repeated until a convergence is reached, e.g., until $\Delta\langle S_\alpha \rangle$ becomes vanishingly small for each α within the accuracy of the statistical error of the simulation. The algorithm is similar to that for a solution of multidimensional nonlinear equations using the Newton-Raphson method [7].

A similar method was applied in a study of the critical point region in the Ising model in Ref. [6]. In that particular case the number of constants K_α was finite. In fact, it was in the range from 1 to 7. For molecular systems, described with pair interaction potentials, the formal number of constants is infinite because of the integral in Eq. (2). In numerical solutions we can use a finite grid to approximate a continuous function.

Let R_{cut} be the cutoff radius for the interaction potential in the computer simulation. For example, R_{cut} can be chosen as half of the cubic box length. The interval $[0, R_{\text{cut}}]$ can be divided into M small slices, each slice corresponding to a distance $r_\alpha = \alpha R_{\text{cut}}/M$, $\alpha = 1, \ldots, M$. Then the Hamiltonian of the system of N particles can be written approximately as

$$H = \sum_{\alpha=1}^{M} K(r_\alpha) S_\alpha \ , \tag{6}$$

where $K(r_\alpha) = \Psi(r_\alpha)$ is the potential at a distance r_α, while S_α is the number of pairs between the particles with distances found inside the αth slice. In computer simulations, S_α is normally used as an estimator of the radial distribution function $\rho(r)$:

$$\rho(r_\alpha) = \langle S_\alpha \rangle \frac{V}{2\pi r_\alpha^2 N(N-1)} \ . \tag{7}$$

It is apparent from Eq. (7) that if we know the radial distribution function we also know the averages $\langle S_\alpha \rangle$. As a trial function or an initial approximation to the effective potential function, we can use, for example, the potential of the mean force:

$$K_\alpha^{(0)} = -kT \ln \rho^*(r_\alpha) \ . \tag{8}$$

III. TEST EXAMPLE: ONE-COMPONENT SYSTEM

To demonstrate how the method presented works, and to reveal some possible hidden difficulties, a few test simulations were carried out for a simple one-component system. The system consists of 50 particles in a cubic box of length $L = 20$ with periodic boundary conditions. All quantities are given in dimensionless units. We started with a model piecewise potential $\Psi^*(r)$. Its exact values in units of kT are given in Table I. A rather coarse grid (the number of slices is 20) was chosen to facilitate the representation of the results in a table form. Potential $\Psi^*(r)$ resembles a Lennard-Jones potential, with $\sigma = 5$ and $\varepsilon = 2$ corresponding to the reduced Lennard-Jones

TABLE I. Restoration of the potential from a radial distribution function for a simple one-component model. Ψ^* is a model potential (in units of kT) and ρ^* is the corresponding RDF from the direct MC simulation; the rest of the table consists of results of the iterational procedure (Sec. II) which reproduce potential Ψ^* from RDF ρ^*. The initial approximation of the potential Ψ_0 was a potential of mean force (8). In the last line, values of the internal energy are shown.

	Direct MC				Reverse MC			
			First iteration		Second iteration		Third iteration	
r	Ψ^*	ρ^*	Ψ_0 (p.m.f.)	ρ_0	Ψ_1	ρ_1	Ψ_2	ρ_2
3.25	∞	0	∞	0	∞	0	∞	0
3.75	10	0.00012	9.028	0.00015	8.788	0.00013	9.239	0.000 11
4.25	6	0.00231	6.07	0.00243	5.679	0.00198	5.736	0.002 23
4.75	2	0.0824	2.50	0.0862	1.95	0.0756	1.987	0.0844
5.25	−2	4.047	−1.398	4.113	−2.113	4.125	−1.990	4.055
5.75	−1	1.467	−0.383	1.496	−1.098	1.475	−1.007	1.468
6.25	−1	1.483	−0.394	1.450	−1.077	1.482	−1.009	1.484
6.75	0	0.5685	0.565	0.523	−0.061	0.559	−0.022	0.560
7.25	0	0.5922	0.524	0.508	−0.044	0.592	−0.015	0.596
7.75	0	0.640	0.446	0.531	−0.042	0.652	−0.012	0.640
8.25	0	0.7507	0.287	0.674	0.025	0.731	−0.005	0.753
8.75	0	0.9446	0.057	0.978	0.034	0.913	−0.015	0.943
9.25	0	1.122	−0.115	1.273	0.005	1.135	−0.001	1.117
9.75	0	1.215	−0.195	1.399	0.006	1.229	−0.002	1.223
U	−313		−167		−339		−315±5	

parameters $T^*=0.5$ and $\rho^*=0.75$.

The goal of this test was to ensure that a successive application of the direct and reverse MC procedures yields the initial interaction potential. The radial distribution function $\rho^*(r)$ was calculated from a direct MC simulation, and its values are given in the third column of Table I. Thereafter the original potential $\Psi^*(r)$ was put aside, and the algorithm, given in Sec. II, was applied to restore the potential from the radial distribution functions $\rho^*(r)$.

Each simulation consisted of 10^7 MC moves. At distances of 3.5 or closer, the potential Ψ^* was set to infinity (hard core), leading to zero RDF intensities at these distances and to zero coefficients in the corresponding lines of the linear equation system in Eq. (3). The zero lines can be simply omitted from the equation system by assuming that if $\rho(r)$ is zero for some interparticle distances, the probability should be zero for any pair of particles to approach the distance and the corresponding potential can be set to infinity. The true potential can still assume a finite value, but the absolute value is not important.

The results are given in Table I. The initial potential $\Psi_0(r)$ in the reverse MC procedure was set equal to the potential of mean force [Eq. (8)]. The iterative procedure converged after three iterations. Only statistical fluctuations were observed during the further iterations. The resulting potential after the third iteration coincides with the original potential Ψ^*. The deviations are less than 0.02, with the exception of small distances between particles. The deviations and statistical errors at small distances are larger. The essential point here is not the small distances, but the high values of the potential at these distances. Because of the high potential value, the corresponding states have a low probability in the canonical ensemble. On the one hand, this is one reason for the uncertainties in the effective potential. On the other hand, this means that the perturbation in the potential does not lead to a noticeable perturbation in RDF and other canonical averages. It can be seen in Table I that RDF, obtained from the effective potential (ρ_2), coincides in the limit of statistical error with original RDF ρ^* in the whole range of interparticle distances. One can also observe clear differences between effective (second column) and mean force (fourth column) potentials.

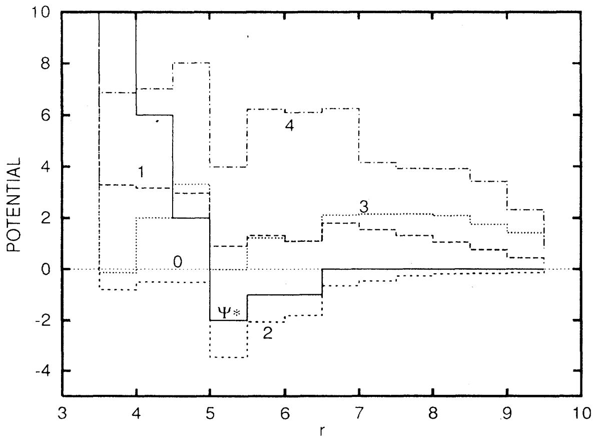

FIG. 1. An example of a nonconvergent iterative procedure starting from zero potential (0); other lines (1–4) are potentials obtained after corresponding iteration. Ψ^* is the original potential in kT units as in Table I.

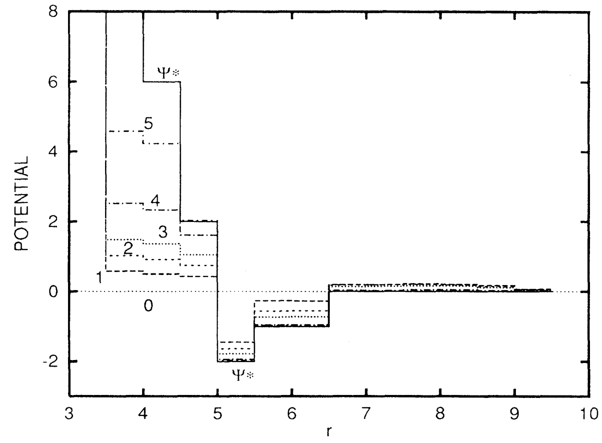

FIG. 2. An example of the convergent iterative procedure starting from zero potential with a damping parameter $\lambda=0.5$; notations as in Fig. 1.

It is not necessary to start the iterative procedure by using the potential of the mean force as a trial function. The obvious reason for this choice is the fact that in many cases the potential of the mean force is quite close to the interparticle potential, therefore the iterative procedure converges rapidly. If the initial approximation is very poor, however, the procedure may never converge. An example of this is the iterative procedure starting from a zero potential [$\Psi_0(r)=0$]. This is shown in Fig. 1, where the potential begins to fluctuate heavily. However, even in the case where the convergence is not achieved, it is possible to apply a damping procedure, in which the left side of Eq. (3) is given as $\lambda\Delta\langle S\alpha\rangle$, where $0<\lambda<1$. This leads to smaller corrections at each iteration step, but also reduces the contribution from the second order term $O(\Delta K^2)$ in Eq. (3). The convergence will be reached, although more iterations will be needed. Figure 2 shows an iterative procedure starting from a zero potential and using $\lambda=0.5$ which converged to the original potential Ψ^*, as did the procedure starting from the potential of mean force in Table I.

IV. CALCULATIONS OF THE EFFECTIVE POTENTIAL FOR Na$^+$ AND Cl$^-$ IONS IN WATER SOLUTION

During the last few years, several works have reported on simulations of ions in water [8–14], in which the calculation of the potential of the mean force has been one of the primary objectives. For electrolyte systems, the potential of the mean force represents the solvent- and ion-mediated averaged potential between ions, and therefore is a valuable quantity for the understanding of many chemical and biochemical processes in solution. However, if we would simply try to exclude water and carry out simulations with the potential of the mean force, the results obtained at finite concentrations (e.g., RDF) would differ from results obtained in ordinary simulations of the detailed system with explicit water Only at infinite dilution (i.e., calculated for a single ion pair in solution), the potential of the mean force coincides with the effective potential. At finite concentrations, screening effects due to ion-ion interactions lead to different behaviors of the effective and mean force potentials. Also, accurate calculations of the potential of the mean force at infinite dilution is quite a difficult task because of poor statistics due to the fact that only one ion pair is included.

The present method can be used to calculate effective interaction potentials which can be directly used for simulations of ionic systems without explicit account of water. These potentials can later be used in simulations of large ionic and polyionic systems for which the explicit water simulations would be too expensive in terms of computing power.

In the present work, ion-ion RDF's were obtained from MD simulations of ions in water. Thereafter, the algorithm described in Sec. II was applied to provide effective potentials between the ions. We used the constant temperature constant pressure (*NPT*) molecular dynamics algorithm [15]. The simulations were carried out in a cubic box with periodic boundary conditions. The Ewald method [1] was applied for treatment of long range electrostatic interactions. The flexible simple-point-charge (SPC) water model [16] was used, and Lennard-Jones parameters for the ions were taken from Ref. [12]: $\sigma = 2.35$ Å and $\varepsilon = 0.544$ kJ/mol for Na^+ and $\sigma = 4.4$ Å and $\varepsilon = 0.419$ kJ/mol for Cl^-. Two separate simulations were carried out: one with four ion pairs and 248 water molecules (0.87-M ion concentration) and another with ten ion pairs and 236 water molecules (2.2-M ion concentration). The RDF's were collected during 1.0 ns, after 200 ps of equilibration. The simulation temperature was 300 K and the pressure 10^5 Pa. The multiple time step algorithm by Tuckerman, Berne, and Martyna [17] was applied, with a small time step of 0.2 fs for intramolecular and short range (less than 4.5 Å) intermolecular interactions, and a large time step of 1.0 fs for the other interactions.

Calculation of an ion-ion RDF appeared to be a tedious task because of the relatively small number of ions in the system and their low mobility, especially in the case of a direct contact between opposite charged ions. Such configurations (ion pairs without intermediate water molecules) were found to exist in time periods of about 50 ps, leading to a poor sampling of the configurational space. The total time period of about 1.0 ns is still to be considered as a minimum to obtain a fairly reliable estimation of ion-ion RDF's. Even in this case, the statistical error in the first maximum of Na^+-Cl^- RDF was about one unit in the intensity scale. At larger distances the RDF's are determined more accurately, with a relative statistical error of about 5% or less.

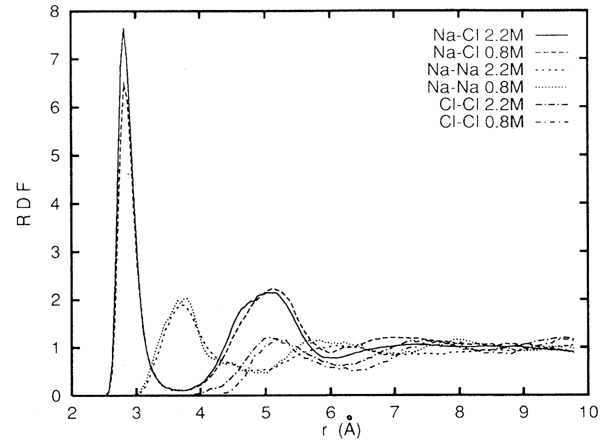

FIG. 3. Ion-ion radial distribution functions obtained from MD simulations.

RDF's obtained from the MD simulation of 1 ns are shown in Fig. 3. These curves show typical features for ion-ion RDF's, and are similar to those calculated in other reported works [8,9,13,14] using a variety of water and ion potential models. The anion-cation RDF has a first contact maximum at $r \sim 2.8$ Å, and a second maximum at $r = 5.1$ Å corresponding to the configuration with a single water molecule between the ions. The Na-Na and Cl-Cl RDF's have their first maxima at distances of 3.72 and 5.3 Å, respectively. The differences, when compared with other researches, are mainly in the intensities, especially for the first maximum of the Na-Cl RDF. Differences between RDF's at the two different concentrations studied in this work are rather insignificant. The density and the internal energy values obtained in the simulations are given in Table II [18].

The next step is the calculation of the effective potential between ions from RDF, on the basis of the algorithm given in Sec. II. The RDF's were calculated at distances ranging from 0 to $R_{\mathrm{cut}} = L/2$ (L is the box side length), and therefore the effective potential can be determined only for these distances. Because of the ionic

TABLE II. Thermodynamical data of constant-temperature and constant-pressure MD simulation of NaCl in a water solution. The simulation time is 1 ns after 0.2 ns equilibration.

Quantity Run No.	1	2
Number of H_2O	248	236
Number of NaCl pairs	4	10
Temperature (K)	300	300
Pressure (bar)	1	1
Salt concentration (M)	0.87	2.2
Average box length (Å)	19.76	19.44
Density (g/cm^3)	1.042(3)	1.093(4)
Internal energy (kJ/mol)	−53.54(5)	−69.57(5)

charges, some care should be taken in the treatment of long-range electrostatic forces. One can assume that, at distances larger than $R_{\rm cut}$, the interactions between the ions are purely Coulombic: $q_iq_j/\varepsilon r$ with a dielectric constant corresponding to water: $\varepsilon=78$. To avoid discontinuity at $R_{\rm cut}$, one can add a constant term $q_iq_j/\varepsilon R_{\rm cut}$ to the potential, inside the cutoff sphere. Adding a constant potential to the Hamiltonian in Eq. (1) does not influence Eqs. (4)–(6), and the algorithm can be applied in the same way as for noncharged system (Sec. III). In the simulations, the long range potential was taken into account using the Ewald method:

$$\left.\frac{1}{\varepsilon r}\right|_{r>R_{\rm cut}}=\left.\sum_{\substack{\text{periodic}\\ \text{cells}}}\left[\frac{1}{\varepsilon r}\right]-\frac{1}{\varepsilon r}\right|_{r<R_{\rm cut}}$$

$$=\left.\frac{\operatorname{erfc}(\alpha r)}{\varepsilon r}+\frac{2\pi}{\varepsilon L^3}\sum_k\frac{1}{k^2}\exp\left[-\frac{k^2}{4\alpha^2}\right]\cos(kr)-\frac{\alpha}{\varepsilon\sqrt{\pi}}-\frac{1}{\varepsilon r}\right|_{r<R_{\rm cut}}, \tag{9}$$

where k shows reciprocal space vectors $[k=(2\pi/L)(l,m,n)$, l, m, and n are integers], and α is the Ewald convergence parameter. Usually α is chosen in such a way that $\operatorname{erfc}(\alpha r)$ can be considered as zero for $r>L/2$. In this work a value $\alpha=5.5/L$ was used.

The MC simulations for calculations of the effective potentials (without explicit water molecules) were carried out in the *NVT* ensemble for the same number of ions as in the MD simulations. The box size was equal to the average box size in the constant-pressure MD simulations (see Table II). The interval $[0,R_{\rm cut}]$ was divided into 200 slices, thus the total number of linear equations in Eq. (3) for three ion-ion RDF's was 600. After removing the zero lines, corresponding to the small, nonprobable distances with zero RDF values, the number of equations was reduced to 418. Solution of such a linear equation system takes about one second of CPU time, thus further refinement of the grid or consideration of a larger number of RDF's does not cause a problem. The initial potential was the potential of the mean force [Eq. (8)] calculated from the MD simulated RDF's with the addition of the term given in Eq. (9) in order to take into account the long range part of the Coulombic interactions. The iterative procedure converged after three iterations. The final effective potentials are shown in Figs. 4–6 together with the corresponding potentials of the mean force. These effective potentials, being used in a MC simulation of the ionic solution without water molecules, yield the same radial distribution functions as were obtained from the full MD simulations in the limit of statistical error.

It may seem at first sight that the effective potential is only a shifted potential of the mean force. However, this is not the case, as can be seen, for example, by examining Fig. 7, in which the differences between the potential of the mean force and the effective potentials are displayed. At the cutoff distance, the difference is about $0.7kT$, which is the constant term $q_iq_j/\varepsilon R_{\rm cut}$ added to the potential. At smaller distances the difference increases, reaching a value of $(1.5-1.8)\,kT$. The fact that the potential of the mean force decays faster than the effective po-

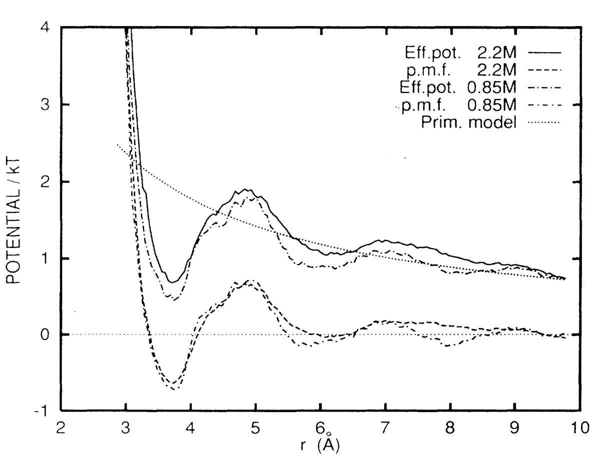

FIG. 4. Na-Na effective potential, and the potential of the mean force.

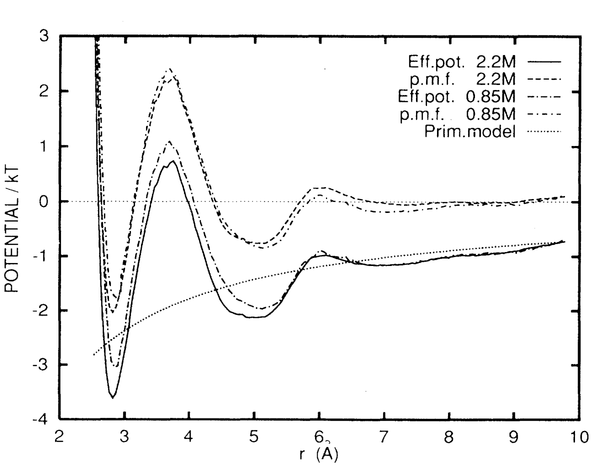

FIG. 5. Na-Cl effective potential, and the potential of the mean force.

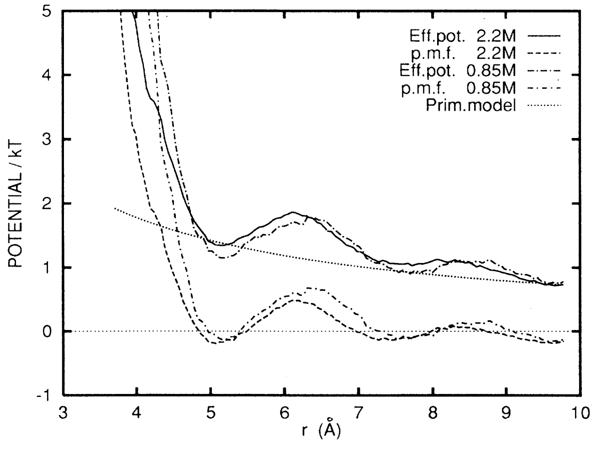

FIG. 6. Cl-Cl effective potential, and the potential of the mean force.

tential is quite natural behavior. This could be expected because at large distances the potential of the mean force decays as a screen Coulomb potential $\sim\exp(-r/r_D)$, where r_D is the Debye radius, whereas the effective potential should approach the potential of the primitive model $q_iq_j/\varepsilon r$. This potential, with a dielectric constant $\varepsilon=78$, is also displayed in Figs. 4–6. One can see that the effective potential makes 1–2 oscillations, thereby reflecting the true molecular nature of the solvent, and then approaches the primitive model potential, already at distances 7 Å and larger. The absolute value of the effective potential is somewhat higher than the primitive model potential, and corresponds to a dielectric constant value of about 70. It can be compared with the result of Ref. [19], $\varepsilon=82.5$, for the same flexible SPC water as used in this paper. Thus calculations of effective potentials can be used for a reasonable estimation of the dielectric constant for specific potential models for solvents. Of course, larger simulation cells and estimations of RDF at larger distances are desirable in order to ensure a true asymptotic behavior.

The calculation of effective potentials may initially require relatively long computer simulation runs for the full system to obtain the reference radial distribution functions, but, once calculated, the effective potentials can be used for simulations of ionic and polyionic systems (e.g., DNA solutions as an example) on a much larger scale. A simulation of ionic systems using effective potentials consumes about equal computer time as a simulation of ions using the primitive model, not forgetting that the effective potentials reproduce structural properties (RDF) corresponding to the full account of the solvent on the molecular level. They do not suffer from the drawbacks of the primitive electrolyte model and can therefore be considered as a reliable approximation between explicit atom solvents and the continuum models for solvents.

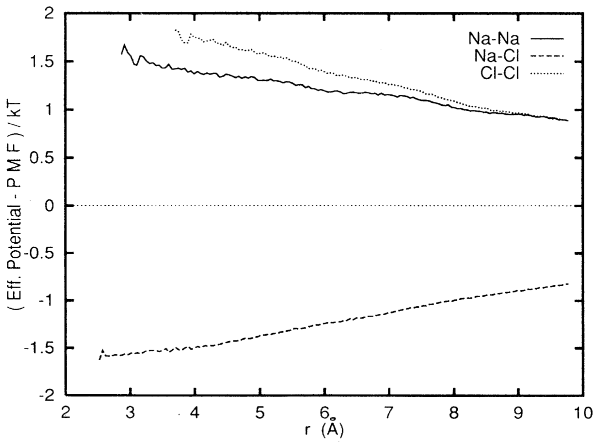

FIG. 7. Differences between the effective potential and the potential of the mean force in the 2.2 M ion solution.

V. CONCLUSION

In this paper we have described a consistent method for constructing effective interaction potentials, and demonstrated it on a real molecular system—the ionic water solution. It is possible, in the same way, to determine the effective potentials between small molecules, atom groups of molecules, and especially macromolecules. These effective potentials can be used in MC simulations on a larger scale without explicit consideration of solvent molecules. For large distances, where the RDF and the effective potential are not calculated in a simulation of the full system, we can assume that only an electrostatic potential with a proper dielectric constant is acting, as was done in Sec. IV. All equilibrium effects of the solvent are taken into account in the effective potentials, in the sense that they reproduce the correct pair distribution functions and provide adequate structural properties. This is especially important for studies of conformational properties of macromolecules. For nonspherical molecules one can include, in the same way, effects of orientational correlation functions, and calculate the orientational part of the effective potentials. The method can also be used for a derivation of effective atom-atom potentials in the case when the atom-atom pair distribution functions are known from the experiment, e.g., from neutron scattering or from *ab initio* (quantum) MD simulations (e.g., Car-Parinello molecular dynamics [20]).

How reliable are the effective potentials? The exact Hamiltonian of the reduced system, which reproduces all canonical averages for interesting degrees of freedom $\{q_i, i=1,\dots,n\}$, is obtained by averaging over the other degrees of freedom $\{q_i, i=n+1,\dots,N\}$:

$$-\beta H^*\{q_i\}_{i=1,\dots,n} = \ln\int\prod_{j=n+1}^{N} dq_j \exp(-\beta H\{q_i,q_j\})_{\substack{i=1,\dots,n\\ j=n+1,\dots,N}} . \tag{10}$$

We use the pair approximation, Eqs. (1) and (2), for the

Hamiltonian H^*, neglecting three- and higher-order particle interactions. Their effect, raising due to interactions with other (rejected) degrees of freedom, may appear in concentration and temperature dependences of the effective potentials. In principle it is possible to include three-body interactions in the method, determining three-body correlation functions from simulations of the full system and then applying the same algorithm. The problem is only technical. While the number of lines in a system of linear equations (3) when pair potentials are used may be several hundred, for three-body potentials it would be millions. In most cases the contribution from the three-body interactions should be small. At least this can be concluded since the overwhelming majority of molecular simulations are done using pair potentials [1]. Thus we may also expect that the effective potentials, obtained by the present algorithm, will provide a good description of equilibrium properties.

Since the effective potentials are constructed from radial distribution functions, they are, first of all, intended to reproduce structural properties. As to thermodynamical properties, one should take into account possible concentration and temperature dependences of the effective potentials and possibly also the three-particle contributions to the entropy and its derivatives. Although in some specific cases these dependences may be insignificant (e.g., in our example of ion solution the effective potentials for two different concentrations are very close to each other), the possible dependences should be studied in each specific case. Notice also that in simulations with effective potentials the thermodynamics quantities will be calculated in terms of the Hamiltonian of the reduced system (1), e.g., for an ion solution the pressure will be the osmotic pressure, etc.

Another question is how do the effective potentials depend on the system size? In our example of ion solution (Sec. IV) the reduced system consisted of a few (8 or 20) particles, thus the effect of a finite number of particles and a finite box size on RDF may be noticeable. It is important from this point to carry out the effective potential calculations (i.e., the MC simulations of the reduced system) with the same number of ions, and in a box of the same geometry as was used in the MD simulation of the full system. In this case we can imply that the finite box size equally affects both full system and reduced system RDF's. Thus we assume that the size effects are mostly canceled, and we can use the effective potentials for simulations on a larger scale. Still this problem needs further study.

Concerning the dynamical properties, we refer to work [21] in which a similar problem of removal of noninteresting degrees of freedom in molecular dynamics simulations was considered. In the molecular dynamics simulation, in addition to the effective potentials a memory function appears. There are no clear criteria for the choice of the memory function. It should be fitted in trial simulations to reproduce correctly the dynamical properties of the full system. In principle one could neglect the memory function and carry out a Brownian dynamics simulation of the reduced system with effective potentials.

ACKNOWLEDGMENTS

This work has been supported by the Swedish Natural Science Foundation (NFR) and by the Wenner-Gren Foundation.

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).

[2] K. Binder and D. W. Heerman, *Monte Carlo Simulation in Statistical Physics* (Springer, Berlin, 1988).

[3] R. L. McGreevy and L. Pusztai, Mol. Sim. **1**, 359 (1988).

[4] C. D. Berweger, W. F. van Gunsteren, and F. Muller-Plathe, Chem. Phys. Lett. **232**, 429 (1995).

[5] R. H. Swendsen, Phys. Rev. Lett. **42**, 859 (1979).

[6] G. S. Pawley, R. H. Swendsen, D. J. Wallace, and K. G. Wilson, Phys. Rev. B **29**, 4030 (1984).

[7] G. Dalquist and A. Björck, *Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1974).

[8] E. Guardia, R. Rey, and J. A. Padro, J. Chem. Phys. **95**, 2823 (1991).

[9] D. E. Smith and A. D. J. Haymet, Fluid Phase Equilibria **88**, 79 (1993).

[10] D.E. Smith and L. X. Dang, Chem. Phys. Lett. **230**, 209 (1994).

[11] R. A. Friedman and M. Mezei, J. Chem. Phys. **102**, 419 (1995).

[12] D. E. Smith and L. X. Dang, J. Chem. Phys. **100**, 3757 (1994).

[13] G. Hummer, D. M. Soumpasis, and M. Neumann, Mol. Phys. **81**, 1155 (1993).

[14] L. R. Pratt, G. Hummer, and A. E. Garcia, Biophys. Chem. **51**, 147 (1994).

[15] S. Melchionna, G. Ciccotti, and B. L. Holian, Mol. Phys. **78**, 533 (1993).

[16] K. Toukan and A. Rahman, Phys. Rev. B **31**, 2643 (1985).

[17] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).

[18] More results on these molecular dynamics simulations of NaCl water solution will be given in our forthcoming paper.

[19] J. Anderson, J. J. Ullo, and S. Yip, J. Chem. Phys. **87**, 1726 (1987).

[20] R. Car and M. Parinello, Phys. Rev. Lett. **55**, 2471 (1985).

[21] P. Romiszovski and R. H. Yaris, J. Chem. Phys. **94**, 6751 (1991).