Potential Truncation Effects in Molecular Simulations

Vom Fachbereich Chemie

der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigte

Dissertation

vorgelegt von

Diplom-Ingenieur Bernd Schilling

aus Köln

Berichterstatter:	Prof. Dr. Jürgen Brickmann
Mitberichterstatter:	Prof. Dr. Hans Jörg Lindner
Tag der Einreichung:	08. August 2005
Tag der mündlichen Prüfung:	24. Oktober 2005

Darmstadt 2005

D17

Für Claudimaus
und meine Eltern

Die vorliegende Arbeit wurde am Institut für Physikalische Chemie der Technischen Universität Darmstadt unter der Leitung von Prof. Dr. J. Brickmann in der Zeit von Januar 1997 bis Juni 2005 durchgeführt.

Mein besonderer Dank gilt an dieser Stelle:

Prof. Dr. J. Brickmann für die gewährte finanzielle und inhaltliche Unterstützung.

PD Dr. S. M. Kast für die herausfordernde Themenstellung, die ständige Diskussionsbereitschaft und die vielen hilfreichen Hinweise bei jeder Art von theoretischen und praktischen Fragen.

Dr. Friedemann Schmidt für die enge und erfolgreiche Kooperation bei der Kombination von Simulation und RISM Integralgleichungstheorie.

Dr. Dirk Zahn für die gute Zusammenarbeit bei der Erweiterung des Wolf Potenzials.

Dr. Robert Jäger für erkenntnisreiche Diskussionen zu Hydrophobie und Süßkraftrezeptor.

Dr. Marco Müller für die freundschaftlichen Aufmunterungen.

Dr. Sven Hauptmann für die Einweisung in den damaligen Molekulardynamikcode.

Dr. Thorsten Borosch, Dr. Kristine Kast und Dr. Richard Marhöfer für den wissenschaftlichen Austausch.

Allen Mitgliedern des Arbeitskreises für eine Atmosphäre der Freundschaftlichkeit und Kooperation.

Der Deutschen Forschungsgemeinschaft für die Förderung.

Inhaltsverzeichnis

Teile dieser Arbeit sind in den Publikationen veröffentlicht:

D. Zahn, B. Schilling, S. M. Kast, *J. Phys. Chem. B* **106**, 10725 (2002).

S. M. Kast, K. F. Schmidt, B. Schilling, *Chem. Phys. Lett.* **367**, 398 (2003).

B. Schilling, J. Brickmann, S. M. Kast, submitted to *Phys.Chem.Chem.Phys.*.

1 Introduction

Modern computer simulation technique have evolved to a third column in science besides experiment and theory. In various branches of science computer simulation provides a powerful tool to study the properties and development of model systems. Famous applications are for example cosmology, weather forecast or vehicle design. In chemistry such computer experiments are applied to analyze the behavior of molecular many-particle systems at atomic level. These molecular simulations[1,2] allow a direct route from microscopic, molecular quantities to the macroscopic phase properties. All aggregate states of matter can be modeled on the basis of simulation methods and a wide range of chemical aspects has already been studied, for example crystals,[3,4] pure liquids,[5,6] liquid mixtures,[7] solvation thermodynamics,[8] phase equilibrium,[9] ligand binding[10] or protein folding.[11,12] The main limiting factor of simulation is the high demand of computer power. Therefore efficiency is an essential task of the algorithms.

Basic simulation techniques of molecular systems are molecular dynamics (MD) and Monte-Carlo (MC) simulation. In classical molecular dynamic simulation usually a system of atomic particles is considered interacting by empirical potential functions, the force field.[13,14] The total potential energy of the system is defined as a sum of intermolecular and intramolecular contributions. The intermolecular potentials are often modeled by a set of Coulomb terms for electrostatic interactions and Lennard-Jones terms for core repulsion and London dispersion. With MD simulation method the microscopic motion of particles is described based on a numerical solution of Newtonian equations of motion. The simulation generates a time series of system configurations, the trajectory. Macroscopic observables are time averages of the generated system configurations and associated with statistical errors, wich can be reduced by increasing the length of simulation time.

The free energy of solvation is a key property of chemical process in liquid phases.[8] It essentially determines the solubility of a solute molecule in a solvent phase and affects partition, aggregation and reaction of solutes in solvent. Regarding

water as solvent in biomolecular systems[15] the free energy of hydration is a fundamental property to understand biomolecular processes such as ligand binding or the structure and function of proteins. A widely discussed phenomenon in aqueous solution is the hydrophobicity.[16,17] It describes on the one hand the low solubility of unpolar solutes in water and on the other hand the tendency of unpolar molecules or groups to accumulate in aqueous solution. Hydrophobic effects are measured by the free energy of hydration or the free energy of transfer between water phase and a nonpolar phase. Therefore methods to calculate the free energy of solvation can also be applied to analyze the phenomenon of hydrophobicity.

Molecular simulation technique can provide a tool for calculating the free energy difference[18-20] between two thermodynamic states. One possible algorithm to perform a free energy simulation is to incrementally transform an initial state to a final state. The corresponding total free energy change can be obtained with thermodynamic integration (TI) or free energy perturbation (FEP) theory. The FEP method is easier to implement but the TI formula allows for a reliable error estimation. The resulting free energy value, which includes enthalpic and entropic terms, is accompanied by significant statistical fluctuations enforcing very long simulation times. Free energy simulation is generally a computer power consuming task justifying all kinds of efficiency improvements.

The computational effort of a molecular simulation strongly depends on the number N of considered atomic particles and the type of interaction potentials. Regarding pair interaction potentials and taking all interactions into account a number of N^2 potentials have to be calculated at each step of simulation.[21] This task quickly exceeds even the power of actual high performance computers, where systems of $N = 10^3 - 10^6$ are reasonable. Thus for efficient simulation strategies the number of particles is chosen as small as possible and finite size simulation boxes with periodic boundary conditions[1] are used to model bulk phase properties. A further increase of efficiency can be achived by introducing spherical cutoff conditions.[22] The potentials are set to zero beyond a sperical cutoff distance R_C. To guarantee the numerical stability of simulation an additional shifting function acts on

the potential such that the first derivative is also set to zero at cutoff distance. As a result the calculational expense is reduced to a $N \cdot R_C^3$ proportionality, wich is quite efficient at small cutoff distances. A more sophisticated techique to handle long-range electrostatic potentials is the Ewald summation,[23] originally constructed to calculate the lattice sum of ionic crystals and modified for use with simulation. The total Coulomb sum of a periodic grid of simulation boxes is summed up partly in real and reciprocal space. The most efficient smooth particle mesh Ewald (SPME) algorithm[23] reaches a $N \cdot \log(N)$ proportionality. In principle the Ewald sum offers the best description of the real potential, but even the efficient SPME version is still considerably more CPU time consuming than small spherical cutoff schemes.

Molecular simulation results are regarded to be reasonable as long as the force field parameters are accurately fitted. But the efficient simulation stategies induce systematic errors in calculated observables.[22,24,25] Two main sources of error can be recognized: The use of a limited number of molecules in the simulation box generates a finite size effect whereas the approximation of interaction potentials causes a truncation effect. At spherical cutoff conditions within a periodic boundary framework both errors affect the results. Finite size effects can only be avoided by increasing the system size but truncation effects can be reduced with appropriate simulation algorithms or estimated with theoretical approaches. For some observables special truncation error correction techniques[26] have been reported but a general truncation error correction method is not described in the literature.

An alternative approach to calculate thermodynamic properties of fluid systems based on statistical theory of liquids is the integral equation (IE) theory.[27-30] Here a set of nonlinear integral equations have to be solved iteratively. Among the algorithms known to solve the integral equations the one-dimensional reference interaction side model[30] (RISM) with the hypernetted chain closure (HNC) has the reputation to be a robust technique. The numerical solution of RISM directly yields radial pair distribution functions (g functions) of the defined system. These functions describe the probability of finding a pair of atoms at a distance r apart. Other thermodynamic properties can be deduced by analytic expressions. RISM integral equation

theory can use the same atomic force field functions and parameters as molecular simulation. Therefore both methods are two alternative treatments to get thermodynamic properties of a fluid phase based on atomic force field data. The RISM theory has the advantages to be much faster for moderatly sized molecules and generates results free from statistical noise whereas the simulation is slow and associated with statistical error. But the current level of approximation of for instance a water phase within integral equation theory is not adequate for evaluation of absolute thermodynamic quantities. Molecular simulation and integral equation results have been compared to each other[31] and simulation truncation artifacts have been studied on basis of IE theory.[32,33] A new concept is the combination of simulation and integral equation in hybrid approaches[34,35] in order to overcome the deficiencies of the single methods.

The aim of this work was the development of strategies to increase the efficiency of molecular simulation, especially free energy simulation. This can be achieved by simulating with spherical cutoff conditions at small cutoff distances but causes significant potential truncation effects. Two ideas seem to be promising to reduce these truncation errors. The first attempt intends to minimize truncation effects as far as possible by performing simulations with an adequatlely parameterized effective interaction potential. A second approach based on the linkage of simulation and integral equation aims at the development of a general truncation effect correction method which can be employed after simulation. The most successful strategy will finally be applied to the simulation of absolute hydration free energies. Subject of this work is the analysis, minimization and correction of truncation effects at molecular simulations with spherically truncated interaction potentials.

The structure of this work is as follows: In chapter 2 an alternative spherical cutoff function with physical background is described. The original potential, wich was designed for ion melts, is modified and parameterized for use with liquid water. With different truncation schemes several static and dynamic properties of water are calculated and compared to each other. In chapter 3 an integral equation based

correction method for observables of molecular simulations with truncated potentials is derived and verified for energetic and structural properties of liquid water. In chapter 4 the new RISM correction strategy is applied to thermodynamic integration free energy simulations. The absolute hydration free energy of argon in water and the electrostatic contribution to the excess chemical potential of water are calculated with various cutoff conditions. The RISM corrected free energy values are compared to the Ewald reference data.

1.1 References

(1) M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

(2) W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem.* **102**, 1020 (1990).

(3) I. Borzsak, P. T. Cummings, *Chem. Phys. Lett.* **300**, 359 (1999).

(4) N. Grishina, V. Buch, *J. Chem. Phys.* **120**, 5217 (2004).

(5) T. Head-Gordon, G. Hura, *Chem. Rev.* **102**, 2651 (2002).

(6) D. J. Price, C. L. Brooks III, *J. Chem. Phys.* **121**, 10096 (2004).

(7) K. M. Kast, J. Brickmann, S. M. Kast, R. S. Berry, *J. Phys. Chem. A* **107**, 5342 (2003).

(8) M. Orozco, F. J. Luque, *Chem. Rev.* **100**, 4187 (2000).

(9) S. A. Best, K. M. Merz,Jr., C. H. Reynolds*, J. Phys. Chem. B* **103**, 714 (1999).

(10) S. Boresch, F. Tettinger, M. Leitgeb, *J. Phys. Chem.* **107**, 9535 (2003).

(11) M. Karplus, G. A. Petsko, *Nature* **347**, 631 (1990).

(12) C. L. Brooks III, D. A. Case, *Chem. Rev.* **93**, 2487 (1993).

(13) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).

(14) A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, M. Karplus*, J. Am. Chem. Soc.* **117**, 11946 (1995).

(15) S. K. Pal, A. H. Zewail, *Chem. Rev.* **104**, 2099 (2004).

(16) L. R. Pratt, A. Pohorille, *Chem. Rev.* **102**, 2671 (2002).

(17) W. Blokzijl, J. B. F. N. Engberts, *Angew. Chem.* **105**, 1610 (1993).

(18) D. L. Beveridge, F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431 (1989).

(19) W. F. van Gunsteren, P. K. Weiner, Ed., *Computer Simulation of Biomolecular Systems* (Escom, Leiden, 1989).

(20) P. Kollmann, *Chem. Rev.* **93**, 2395 (1993).

(21) C. Sagui, T. Darden, *J. Chem. Phys.* **114**, 6578 (2001).

(22) P. J. Steinbach, B. R. Brooks, *J. Comp. Chem.* **15**, 667 (1994).

(23) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).

(24) Y. Y. Sham, A. Warshel, *J. Chem. Phys.* **109**, 7940 (1998).

(25) M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys.* **119**, 9129 (2003).

(26) H. Resat, J. A. McCammon, *J. Chem. Phys.* **108**, 9617 (1998).

(27) J.-P. Hansen, I. R. McDonald, *Theory of simple liquids*, 2nd edn., Academic Press, London, 1991.

(28) D. Chandler and H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

(29) F. Hirata and P. J. Rossky, *Chem. Phys. Lett.* **83**, 329 (1981).

(30) D. Chandler, H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

(31) Q. Du, D. Beglov, B. Roux, *J. Phys. Chem. B* **104**, 796 (2000).

(32) C. L. Brooks III, *J. Chem. Phys.* **86**, 5156 (1987).

(33) C. L. Brooks, B. Montgomery, and M. Karplus, *J. Chem. Phys.* **83**, 5897 (1985).

(34) K. F. Schmidt and S. M. Kast, *J. Phys. Chem. B* **106**, 6289 (2002).

(35) S. M. Kast, *Phys. Chem. Chem. Phys.* **3**, 5087 (2001).

2 Analysis and minimization of truncation effects

2.1 Introduction

In molecular dynamics (MD) and Monte-Carlo (MC) simulations the accurate evaluation of electrostatic interactions is generally the computationally most demanding task. Several methods have been developed for the treatment of long-range nonbonded interactions. Ewald summation[1] is accepted as the most adequate treatment of Coulomb interaction in periodic systems. Besides reaction field techniques,[2,3] spherical cutoff methods[4] where interactions are taken into account only within a sphere of a certain cutoff radius are computationally less expensive and therefore widely used. It is well known that a discontinuous truncation of the electrostatic potential at the cutoff distance introduces severe truncation artifacts.[5] Truncation errors may be reduced by the use of *switching* or *shifting* schemes. Switching functions that smooth a potential to zero over a certain interval are known to lead to unrealistic fluctuations of the effective molecular pair potentials near the cutoff distance. The best compromise so far seems to be a continuous shifting of the potential at all distances such that value *(shifted potential*) or value and first derivative of the potential both become zero at the cutoff distance (*shifted-force potential*), as implemented for instance in the commonly used CHARMM program.[6] An improved shifted-force potential, particularly with respect to the forces as used in MD simulations has been proposed by Schrimpf *et al*.[7] Both the functional form and the cutoff distance as well as the methodology (atom-based or group-based truncation schemes) influence the quality of the approximation to the real long-range interactions. Several comparative studies have critically examined the various schemes with respect to different observables.[4,8,9]

Recently Wolf *et al*.[10] investigated into the origin of cutoff artifacts. In their clear analysis they first showed that the cutoff artifacts are related to the existence of net charges within the cutoff spheres, caused by the finite number of interaction sites considered. In the second step, it was demonstrated that shifting the potential by a neutralizing term corresponding to the net charge results in faster convergence to the

Madelung limit. Ultimately, they came up with an expression for a damped Coulomb potential satisfying the requirements of charge neutrality and force shifting for use in MD simulations. Physically interpreted, the effective Coulomb interaction in a polarizable medium was shown to decay as the inverse 5th power of the distance[11] so that the existence of net charges leads to considerably longer-ranged behavior explaining a large portion of the artifacts. Moreover, in molecular simulations using conventional shifted-force schemes, particles penetrate the cutoff sphere and instantly change the net charge. This is effectively circumvented by damping the potentials such that interactions are switched on and off more smoothly.

In light of the exceptional success of the damping scheme in their simulations of simple ionic systems even for small cutoff radii,[10] it appeared promising to apply the methodology to molecular systems. Demontis *et al*.[12] quite recently used the original form of the damped Coulomb potential in a simulation study of liquid water and anhydrous and hydrated aluminosilicates, and discussed criteria for the selection of proper combinations of the damping parameters. With their final set a significant speed-up compared to the traditional Ewald scheme could be achieved.

As has been noted in Wolf *et al*.'s original work, the forces and potentials are not consistent. In this work, the inconsistency is removed yielding a significant improvement of dielectric properties when applied to dipolar liquids. For evaluation of this enhanced potential simulations of pure liquid water are performed with the commonly used TIP3P[13] and SPC[14] water models. The performance of the enhanced damped Coulomb scheme is compared with benchmark Ewald simulations and the conventional CHARMM cutoff method.

The chapter is organized as follows: In section 2.2 the basic theory is described and the consistent formulation for potential and force is given along with a discussion of the dielectric properties. In the following a pragmatic strategy is given for parameterization of damping constants and cutoff radii in order to achieve simulation results of near Ewald quality. In section 2.4 static, dynamic, and dielectric properties as obtained from MD simulations with the various Coulomb summation schemes are reported. In the final section the results are summarized.

2.2 Theory

2.2.1 The enhanced damped Coulomb potential

The derivation of the damped Coulomb potential was given by Wolf *et al.*[10] for purely ionic systems. This formalism is summarized and extended to molecular sytems. The total Coulomb energy in units of 4π times the vacuum dielectric constant for an ion i is given by

$$E_i^{\mathrm{tot}}\left(r_{ij}\right)=\sum_{j\neq i}\frac{q_i q_j}{r_{ij}} \tag{2.1}$$

where q_i is the charge of ion i and q_j is the charge of another ion j at distance r_{ij}. In practice the evaluation of the infinite sum has to be approximated by a finite sum. This may be done either by introducing periodicity and performing Ewald summation or by considering only ions j which are within a cutoff sphere of radius R_C. Eq. (2.1) is then approximated by

$$E_i^{\mathrm{tot}}\left(r_{ij}\right)\approx \sum_{\substack{j\neq i\\ r_{ij}<R_C}}\frac{q_i q_j}{r_{ij}}. \tag{2.2}$$

It is well known that this radial summation of truncated potentials does not converge to the true limit. Even for large cutoff radii evaluation of eq. (2.2) leads to dramatic errors. Wolf *et al.*[10] have shown that this error is related to the existence of a net charge within the cutoff sphere. Thus the accuracy may be significantly improved by introducing a correction term E_i^{neutral} that compensates for the lack of charge neutrality in the truncated sum in eq. (2.2). The error in evaluating the total Coulomb energy is then of the order R_C^{-2}:

$$E_i^{\mathrm{tot}}\left(r_{ij}\right)=\sum_{\substack{j\neq i\\ r_{ij}<R_C}}\frac{q_i q_j}{r_{ij}}-E_i^{\mathrm{neutral}}+O\left(R_C^{-2}\right). \tag{2.3}$$

For ionic crystals E_i^{neutral} may be determined in the following way: The ions are grouped as charge-neutral "molecules". For a simple crystal of only two species A^+ and B^- these "molecules" are most simply chosen as dipoles containing one ion of each species at minimum distance. The sum of Coulomb potential terms is now taken for those ions only that belong to "molecules" that are entirely within the cutoff

sphere, yielding a "charge-neutralized" summation of interaction terms. E_i^{neutral} is given by the difference of the sum according to eq. (2.2) and the "charge-neutralized" sum of the Coulomb potential.

Wolf *et al.*[10] derived an elegant method to approximate E_i^{neutral}. Let the maximum ion distance within a charge-neutral group be b. All ions j that belong to "molecules" truncated by the cutoff sphere are within a distance $R_C - b < r_{ij} < R_C + b$ to ion i. If b is small compared to the cutoff radius, the Coulomb interaction of the truncated "molecules" becomes approximately

$$E_i^{\text{neutral}} \approx \sum_{j'} \frac{q_i q_{j'}}{R_C} = \frac{q_i \Delta q_i}{R_C} \tag{2.4}$$

where the prime on the index j indicates summation over ions belonging to truncated "molecules" only and Δq is the net charge within the cutoff sphere. Using this approximation the total Coulomb energy for a configuration **r** is given by

$$E^{\text{tot}}(\mathbf{r}) \approx \frac{1}{2} \sum_{\substack{i,j \neq i \\ r_{ij} < R_C}} \frac{q_i q_j}{r_{ij}} - \frac{1}{2} \sum_i \frac{q_i \Delta q_i}{R_C} = \frac{1}{2} \sum_{i,j \neq i} V^{\text{sh}}(r_{ij}) \tag{2.5}$$

with the shifted pair potentials

$$V^{\text{sh}}(r_{ij}) = \begin{cases} \dfrac{q_i q_j}{r_{ij}} - \dfrac{q_i q_j}{R_C} & \Leftrightarrow \quad r_{ij} < R_C \\ 0 & \Leftrightarrow \quad \text{else}. \end{cases} \tag{2.6}$$

Although the summation in eq. (2.4) is performed over truncated "molecules" only, the corresponding sum over the charge-neutral sites within a sphere exactly cancels to zero. Accordingly, the j sum in eq. (2.5) can be done over all ions.

The shifted pair potential continuously reaches zero at the cutoff radius. In molecular dynamics simulations it is necessary to use potentials whose first derivatives also satisfy this condition. This may be accomplished by using shifted-force potentials. Wolf *et al.*[10] made an attempt to derive a shifted force from the potential eq. (2.6) directly. This was achieved by a questionable operation where differentiation and taking the limit are illegitimately assumed to commute:

$$\frac{\partial}{\partial r_{ij}} V^{sh}\left(r_{ij}\right) = -\frac{q_i q_j}{r_{ij}^2} - \frac{\partial}{\partial r_{ij}} \frac{q_i q_j}{R_C} = -\frac{q_i q_j}{r_{ij}^2} - \frac{\partial}{\partial r_{ij}} \lim_{r_{ij} \to R_C} \frac{q_i q_j}{r_{ij}} \neq -\frac{q_i q_j}{r_{ij}^2} - \lim_{r_{ij} \to R_C} \frac{\partial}{\partial r_{ij}} \frac{q_i q_j}{r_{ij}}. \tag{2.7}$$

From a physical point of view it is reasonable to demand that all magnitudes like potential, force, and higher derivatives become zero at R_C, avoiding a truncation. Mathematically, this would mean either the function to be zero everywhere or a function that is not analytic. Neither option is physically plausible, so using the shifted potential (2.6) in a strict sense implies forces that, if required to be zero at R_C as well, are not consistent with the potential. Alternatively the formally constant R_C-dependent shifting terms might be viewed as the result of an approximate average of contributions from many distances. In this way the shift would be interpreted as an r-dependent variable thus validating the formal differentiation procedure. One can of course write down an expression for an *approximate* force according to the prescription of eq. (2.7) and *enforce* consistency with the potential. Therefore in the present work a shifted-force potential is used which is derived from integration of the shifted force. Proceeding in this way is essential for the dielectric properties of dipolar liquids as shown later. Although the intuitive physical meaning of the charge neutralization is lost the approach appears to be an acceptable compromise if the original potential is not changed too much.

In practical work, it is advantageous to use Wolf *et al.*'s damping approach. One has to make a compromise between the absence of truncation effects and the deformation of the original pair potential. Wolf *et al.*[10] found an elegant way to combine their results of the charge neutralization analysis with the idea of a damping function $d(r_{ij})$ as a factor to the pair potential. Though the derivatives do not exactly reach zero, they may be damped closely to zero. The total energy of ion i reads

$$E_i^{\text{tot}} = \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} d(r_{ij}) + \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} [1 - d(r_{ij})] \tag{2.8}$$

which formally resembles the Ewald summation procedure. Consequently, Wolf *et al.*[10] decided to use the error function complement erfc($\alpha\ r_{ij}$) as damping function, though the choice of $d(r_{ij})$ is somewhat arbitrary. Here, α is a damping

parameter chosen in accordance with the cutoff radius R_C. By adding and subtracting the self term associated with the second contribution

$$E_{\text{self}} = \lim_{r_{ij}\to 0}\left\{\frac{1}{2}\sum_i \frac{q_i^2}{r_{ij}}\operatorname{erf}(\alpha r_{ij})\right\} = \frac{\alpha}{\pi^{1/2}}\sum_i q_i^2 \tag{2.9}$$

one obtains

$$E^{\text{tot}} = \frac{1}{2}\sum_{i,j\neq i}\frac{q_i q_j}{r_{ij}}\operatorname{erfc}(\alpha r_{ij}) - \frac{\alpha}{\pi^{1/2}}\sum_i q_i^2 + \frac{1}{2}\sum_{i,j}\frac{q_i q_j}{r_{ij}}\operatorname{erf}(\alpha r_{ij}) \tag{2.10}$$

where now the last term can be shown to be very small[10] and negligible.

For a truncated potential, applying the charge neutralization procedure [eqs. (2.4) and (2.5)] finally yields

$$E^{\text{tot}} \approx \frac{1}{2}\sum_{\substack{i,j\neq i\\ r_{ij}<R_C}}\left(\frac{q_i q_j}{r_{ij}}\operatorname{erfc}(\alpha r_{ij}) - \frac{q_i q_j}{R_C}\operatorname{erfc}(\alpha R_C)\right) - \left(\frac{1}{2R_C}\operatorname{erfc}(\alpha R_C) + \frac{\alpha}{\pi^{1/2}}\right)\sum_i q_i^2 \,. \tag{2.11}$$

The last term in eq. (2.11) is now a self energy associated with each ion *i*. Using the formally incorrect prescription (2.7) for the derivation of damped Coulomb pair forces then gives[10]

$$F^{\text{dc}}(r_{ij}) = \begin{cases} q_i q_j\left\{\dfrac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}^2} + \dfrac{2\alpha}{\pi^{1/2}}\dfrac{\exp(-\alpha^2 r_{ij}^2)}{r_{ij}} - \right. & \\ \left.\left[\dfrac{\operatorname{erfc}(\alpha R_C)}{R_C^2} + \dfrac{2\alpha}{\pi^{1/2}}\dfrac{\exp(-\alpha^2 R_C^2)}{R_C}\right]\right\} & \Leftrightarrow \quad r_{ij} < R_C \\ 0 & \Leftrightarrow \quad \text{else} \,. \end{cases} \tag{2.12}$$

For molecular liquids consisting of neutral molecules without internal Coulombic interaction, the ions are replaced by the atomic sites and the ionic charges by partial charges. In this case, the self term of eq. (2.11) cancels and it is safe to compute the interaction potential consistently by integrating the pair force, yielding

$$V^{\text{dc}}(r_{ij}) = \begin{cases} q_i q_j\left\{\dfrac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} - \right. & \\ \left.\left[\dfrac{\operatorname{erfc}(\alpha R_C)}{R_C^2} + \dfrac{2\alpha}{\pi^{1/2}}\dfrac{\exp(-\alpha^2 R_C^2)}{R_C}\right](r_{ij} - R_C)\right\} & \Leftrightarrow \quad r_{ij} < R_C \\ 0 & \Leftrightarrow \quad \text{else} \,. \end{cases} \tag{2.13}$$

This relation constitutes the enhanced damped Coulomb potential. In case of charged solutes, a modified expression for the self term must be added to give the best approximation to full Ewald summation. Here only pure liquids are treated and simulated with the damped Coulomb formulae eqs. (2.12) and (2.13).

The limiting factor in using the damped Coulomb scheme is the maximum atom distance within a molecule. As was shown above, eq. (2.5) is valid only if this distance is small compared to the cutoff radius. For small molecules like water this condition is satisfied even better than for ionic systems. Larger molecules instead require larger cutoff radii or alternatively the introduction of charge-neutral subgroups within the entities.

2.2.2 Dielectric properties

The basic theory for treating dielectric properties in systems with toroidal boundary conditions was presented by Neumann and Steinhauser[15,16] almost two decades ago: The relative dielectric permittivity ε_r is given by the fluctuation formula

$$\frac{\varepsilon_r - 1}{\varepsilon_r + 2} = \frac{4\pi}{3} \frac{\langle M^2 \rangle}{3Vk_{\mathrm{B}}T} \left[1 - \frac{3}{4\pi} \frac{\varepsilon_r - 1}{\varepsilon_r + 2} \widetilde{T}(0) \right] \tag{2.14}$$

where $\langle M^2 \rangle$ is the fluctuation of the total dipol moment, k_{B} is the Boltzmann constant, T the absolute temperature, and V the volume of the basic cell. For a spatially isotropic system $\widetilde{T}(0) \equiv \widetilde{T}_{xx}(0) = \widetilde{T}_{yy}(0) = \widetilde{T}_{zz}(0)$ denotes the **k** = 0 limit of the Fourier transform (i.e. the integral) of the dipolar tensor **T** representing the effective interaction $-\boldsymbol{\mu}\cdot\mathbf{T}\cdot\boldsymbol{\mu}$ between molecular dipoles **μ** in Gaussian units. For the general case of a radially symmetric damping function $f(r)$ acting multiplicatively on the original Coulomb r^{-1} potential we have to evaluate the **T** tensor elements

$$T_{ab} = \nabla_a \nabla_b \left(r^{-1} f(r) \right), \quad a,b \in \{x,y,z\}. \tag{2.15}$$

One obtains after some algebra

$$\mathbf{T} = \mathbf{T}^0 \left(f - r\frac{df}{dr} + \frac{1}{3} r^2 \frac{d^2 f}{dr^2} \right) + \frac{1}{3} r^{-1} \frac{d^2 f}{dr^2} \mathbf{I} \tag{2.16}$$

with **I** being the identity matrix, $\mathbf{T}^0$ is the dipolar tensor for a pure Coulomb potential. Considering the integral over the volume the first term in eq. (2.16) vanishes due

to angular symmetry.[15] The only nonzero contribution to $\widetilde{T}(0)$ comes from the second term:

$$\frac{4\pi}{3}Q \equiv \widetilde{T}(0) = \int_V d\mathbf{r}\left(\frac{1}{3}r^{-1}\frac{d^2 f}{dr^2}\right) = \frac{4\pi}{3}\int_0^{R_C} dr\, r\frac{d^2 f}{dr^2}. \tag{2.17}$$

Here it is assumed that the molecular diameter is small compared with the cutoff distance as required by the damped Coulomb model. Q = 1 for Ewald summation represents the ideal infinite box size limit, any smaller value influences the dielectric constant.[15]

Turning now to the special cases of the original and the enhanced damped Coulomb potentials we get from eq. (2.17) and the damping term $f(r)$ = erfc(αr)-erfc(αR_C)

$$Q^{\mathrm{dc,orig}} = \mathrm{erf}(\alpha R_C) - 2\alpha R_C \pi^{-1/2}\exp(-\alpha^2 R_C^2) \tag{2.18}$$

which evaluates to 0.9095 for a typical combination of α = 0.2 Å^{-1} and R_C = 9 Å. This result would also be obtained with a truncated Ewald scheme. In contrast, the enhanced form, eq. (2.13), yields

$$Q^{\mathrm{dc,enhanced}} = \mathrm{erf}(\alpha R_C) + \mathrm{erfc}(\alpha R_C) = 1 \tag{2.19}$$

independent of α and R_C. Removing the inconsistency between force and potential by a simple shifted-force technique obviously improves the dielectric properties significantly compared to the original form.

2.3 Methods

2.3.1 Simulation conditions

Molecular dynamics simulations have been performed using a the DL_POLY software.[17] Both SPC[14] and TIP3P[13] water models were investigated. Force field parameters and geometry of the water models are summarized in Table 2.1. Internal coordinates were constrained with SHAKE algorithm.[18] The simulation systems each contained 1000 water molecules with periodic boundaries applied to rectangular cells.

Table 2.1: Geometric and force field parameters for SPC and TIP3P water models (*r*: distances; *q*: partial charges; ε, σ: Lennard-Jones parameters).

	SPC	TIP3P
r_{OH} / Å	1.000	0.9572
r_{HH} / Å	1.633	1.5139
q_O / e	-0.82	-0.834
q_H / e	0.41	0.417
ε_O / kcal mol^{-1}	0.1554	0.1521
ε_H / kcal mol^{-1}	0	0
σ_O / Å	3.166	3.150
σ_H / Å	0	0

Reference data was obtained from MD simulations with Ewald summation using a real space cutoff of 12 Å and a Ewald coefficient chosen as 0.3 Å^{-1}. For each system the initial equilibration and the determination of static observables were performed under *NpT* conditions[19,20] with the temperature set to 300 K and the pressure to 1 bar. In both models a time step of 2 fs was applied. In addition to the damped Coulomb potential the CHARMM shifted force potential[6]

$$V_{\mathrm{CHARMM}}^{\mathrm{shf}}(r_{ij}) = \begin{cases} \dfrac{q_i q_j}{r_{ij}} \left[1 - \left(\dfrac{r_{ij}}{R_C} \right)^2 \right]^2 & \Leftrightarrow \quad r_{ij} < R_C \\ 0 & \Leftrightarrow \quad \text{else}. \end{cases} \tag{2.20}$$

was used for comparison.

When investigating dynamical properties it is useful to switch to a micro-canonical ensemble as the constant temperature and pressure algorithms have a significant effect on the system dynamics. The transition to *NVE* simulation was accomplished by first determining the average box volume and potential energy in *NpT* simulations. The *NpT* simulation run was stopped at a time when the box volume was equal to its average by a relative accuracy of 10^{-4}. Starting from this

configuration a *NVT* ensemble was simulated. This simulation was stopped as soon as the instantaneous potential energy reached its average value within the same accuracy. This procedure leads to well defined systems for subsequent micro-canonical simulations.

All systems were equilibrated for at least 200 ps. Static properties were computed from 200 ps sampling runs, dynamic properties from 2 ns simulations except for the diffusion coefficients where short 20 ps runs were used (see below).

2.3.2 Parameterization of the damped potential parameters

In contrast to Demontis *et al.*[12] where a general relation for an optimal parameter combination was sought on the basis of fixed box lengths, here a simple two-step strategy is used for finding a suitable combination of the damped potential parameters α and R_C in liquid systems with variable shape:

a) At a given cutoff radius, *NpT* simulations were run for a range of damping constants. The initial cutoff radius was chosen to be the same as the real space cutoff radius of the Ewald simulations. The damping constant was varied from 0.15 to 0.4 $Å^{-1}$. For later cutoff reduction it is favorable to choose the damping constant as large as possible. However, for both TIP3P and SPC water the average potential energy and the density deviates increasingly from the reference values with larger damping constants. To ensure a simulation of near-Ewald quality the damping constant was chosen such that the average potential energy per molecule and the average density did not deviate more than 5 % from the values of the reference simulation. For both SPC and TIP3P water models a reasonable α value was found to be 0.2 $Å^{-1}$. During the next step the damping constant was kept fixed at this value.

b) In a new series of *NpT* simulations the cutoff radii were varied from 12 to 6 Å. Again the resulting average potential energy and density were compared with the Ewald simulations. For a cutoff radius of 9 Å and larger both properties were found to be within the desired agreement of 5 % with the reference data. Thus we chose the combination of $\alpha = 0.2$ $Å^{-1}$ and $R_C = 9$ Å for the simulation of SPC and TIP3P water at near-Ewald quality. The original and the enhanced damped Coulomb

potential with these parameters are visually practically undistinguishable. Stronger deviations occur for much smaller values of α at the given R_C. It is interesting to note that the optimal result corresponds roughly to an estimate of $\alpha = 0.22$ Å^{-1} along the lines of reference 12.

The optimal parameter combination was selected under constant pressure conditions because of the practical significance for the simulation of solutions where a specific constant bulk water density is difficult to attain. For these problems, a pragmatically chosen parameter set determined for the pure solvent is sufficient and would be used for solutions without change.

2.4 Simulation results and discussion

2.4.1 Static properties

To elucidate general trends, both water models, TIP3P and SPC, were simulated under *NpT* conditions using the damped potential with cutoff radius of 12 Å and the damping constant α varying from 0.15 to 0.4 Å^{-1}. Figures 2.1 and 2.2 show average potential energy and average density as a function of α. With increasing damping constant both properties systematically deviate from the Ewald reference data. This is readily explained by the damping of the Coulomb potential. The long-range interaction of the water molecules is dominated by dipol-dipol attraction. Thus the damping leads to an underestimation of the attractive part of the intermolecular interaction. Therefore the particle density decreases with increasing α. The error in the average potential energy per molecule is related to the approximate neglect of the last term in eq. (2.10). It becomes more crucial for larger damping constants and smaller cutoff radii.

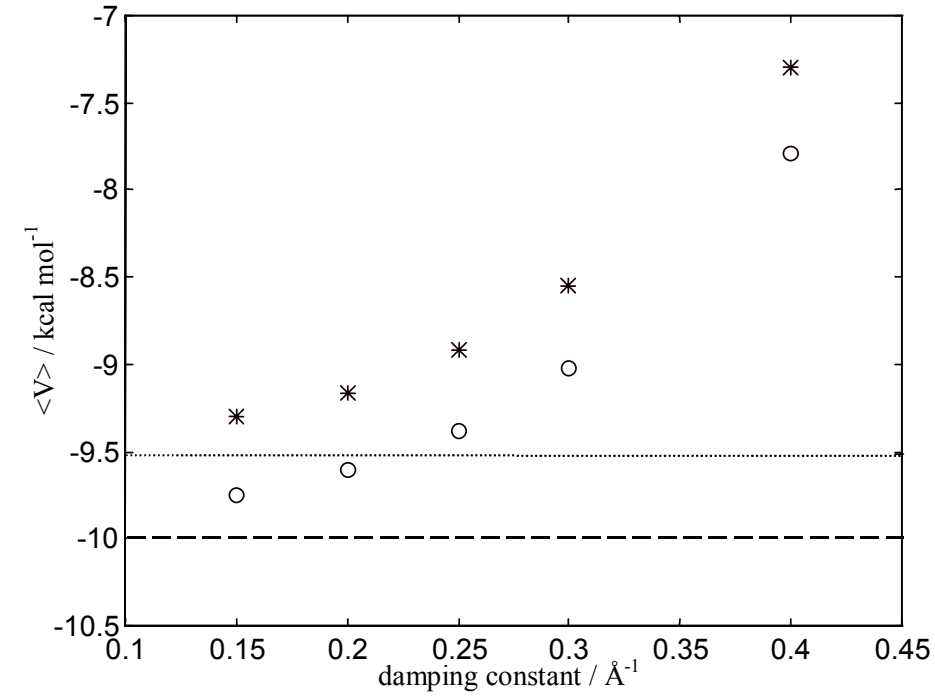

Figure 2.1: Average potential energy per water molecule for TIP3P (*) and SPC (O) models. Ewald summation yields –9.51 kcal/mol (dotted) and –10.00 kcal/mol (dashed) for TIP3P and SPC water, respectively.

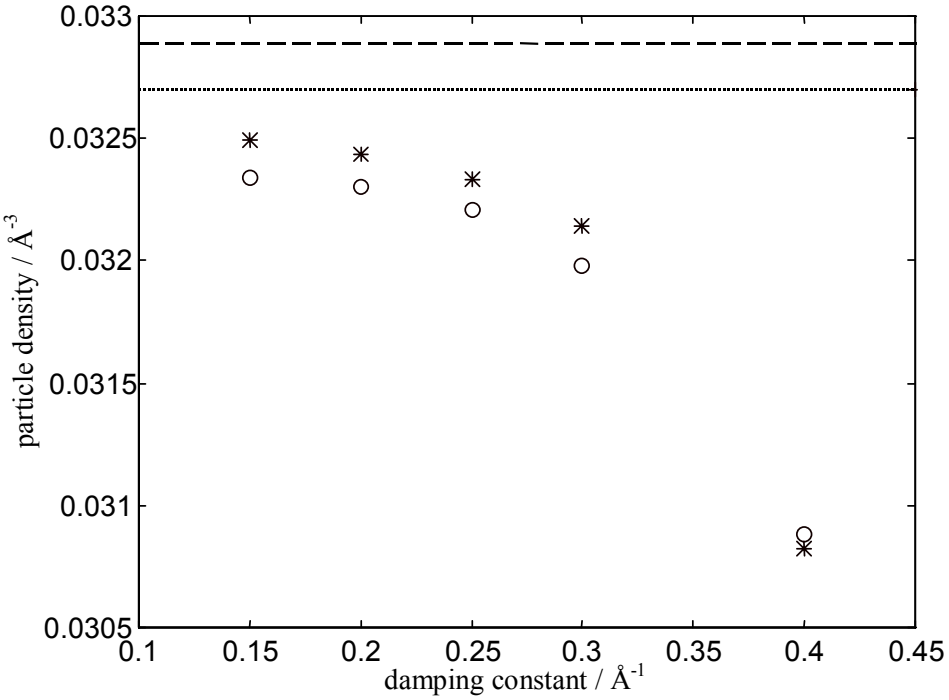

Figure 2.2: Average density of water molecules for TIP3P (*) and SPC (O) models. Ewald summation yields 0.0329 $Å^{-3}$ (dashed) and 0.0327 $Å^{-3}$ (dotted) for TIP3P and SPC water, respectively.

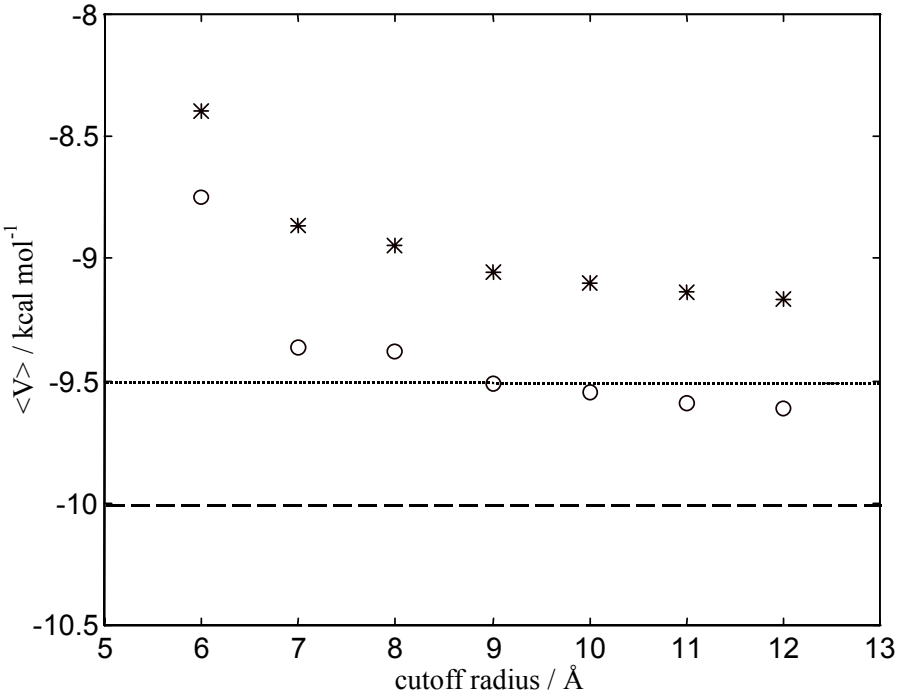

Figure 2.3: Average potential energy per water molecule as a function of the cutoff radii for TIP3P (*) and SPC (O) models. Ewald summation yields -9.51 kcal/mol (dotted) and –10.00 kcal/mol (dashed) for TIP3P and SPC water, respectively.

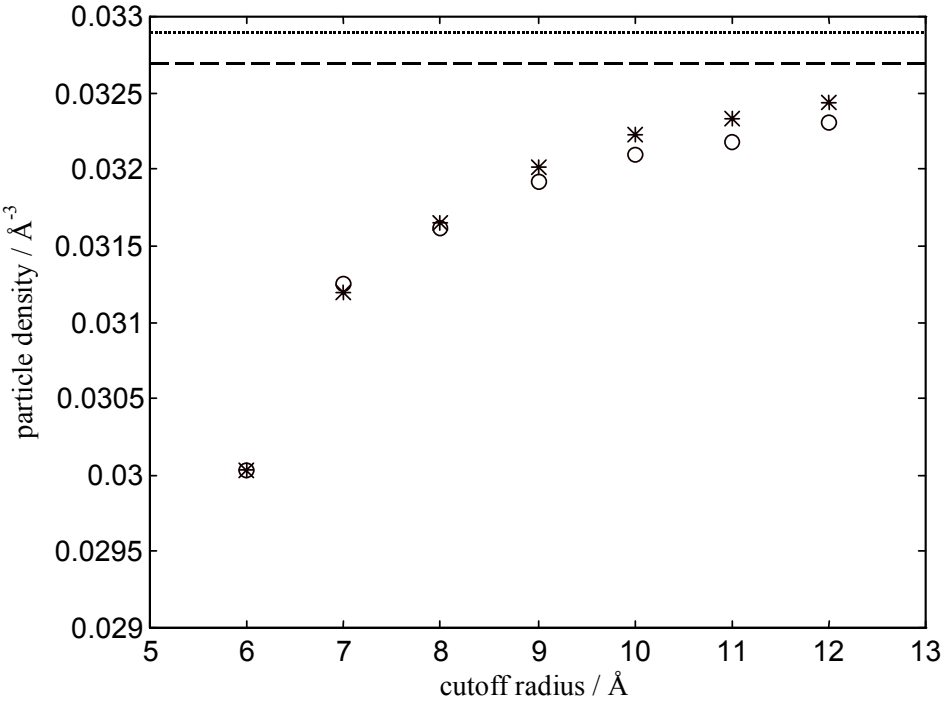

Figure 2.4: Average density of water molecules as a function of the cutoff radii for TIP3P (*) and SPC (O) models. Ewald summation yields 0.0329 Å-3 (dotted) and 0.0327 Å-3 (dashed) for TIP3P and SPC water, respectively.

According to the parameterization scheme described in the previous section, the damping constant was chosen as 0.2 Å^{-1} in order to reach the desired 5 % accuracy for both average density and energy. Then the cutoff radii were varied from 12 to 6 Å. Average potential energy and density were recorded as functions of R_C (Figures 2.3 and 2.4). The deviation of energy and density from the reference values increases for decreasing cutoff radii, as expected. However, the 5 % accuracy threshold is reached at a cutoff distance of only 9 Å. The combination of $\alpha = 0.2$ Å^{-1} and $R_C = 9$ Å yielding near-Ewald quality was used in all subsequent investigations.

Radial pair distribution functions were computed according to

$$g_{ij}(r) = \frac{3}{4\pi}\left\langle \frac{n_j(r-\Delta r/2; r+\Delta r/2)}{\rho_j[(r+\Delta r/2)^3 - (r-\Delta r/2)^3]} \right\rangle_{NpT} \tag{2.21}$$

where r is the distance of atoms i and j, n_j is the number of atoms j within distance r-Δr/2 and r+Δr/2 of atom i with $\Delta r = 0.1$ Å, and ρ_j is the density of the atoms j. Figure 2.5 shows the pair distribution functions of O-O, O-H, and H-H atom pairs of TIP3P water model, while Figure 2.6 illustrates the corresponding data for SPC. The solid curves correspond to the damped potential and the dotted curves were obtained from simulation with Ewald summation. For comparison, we also give the corresponding distribution functions obtained with a CHARMM shifted-force potential (dashed curve) using the same cutoff distance as for the damped Coulomb method. Despite the small cutoff radius of only 9 Å our simulations using the damped potential are able to provide radial distribution functions in very close agreement with those obtained from Ewald simulations. On the other hand the radial distribution functions taken from simulations using the standard CHARMM shifted-force potential show significantly stronger deviations from the reference data. These are particularly striking in the proximity of the cutoff radius. This is related to more pronounced perturbing effects at the cutoff distance as compared to the damped Coulomb case, thereby proving again its value.

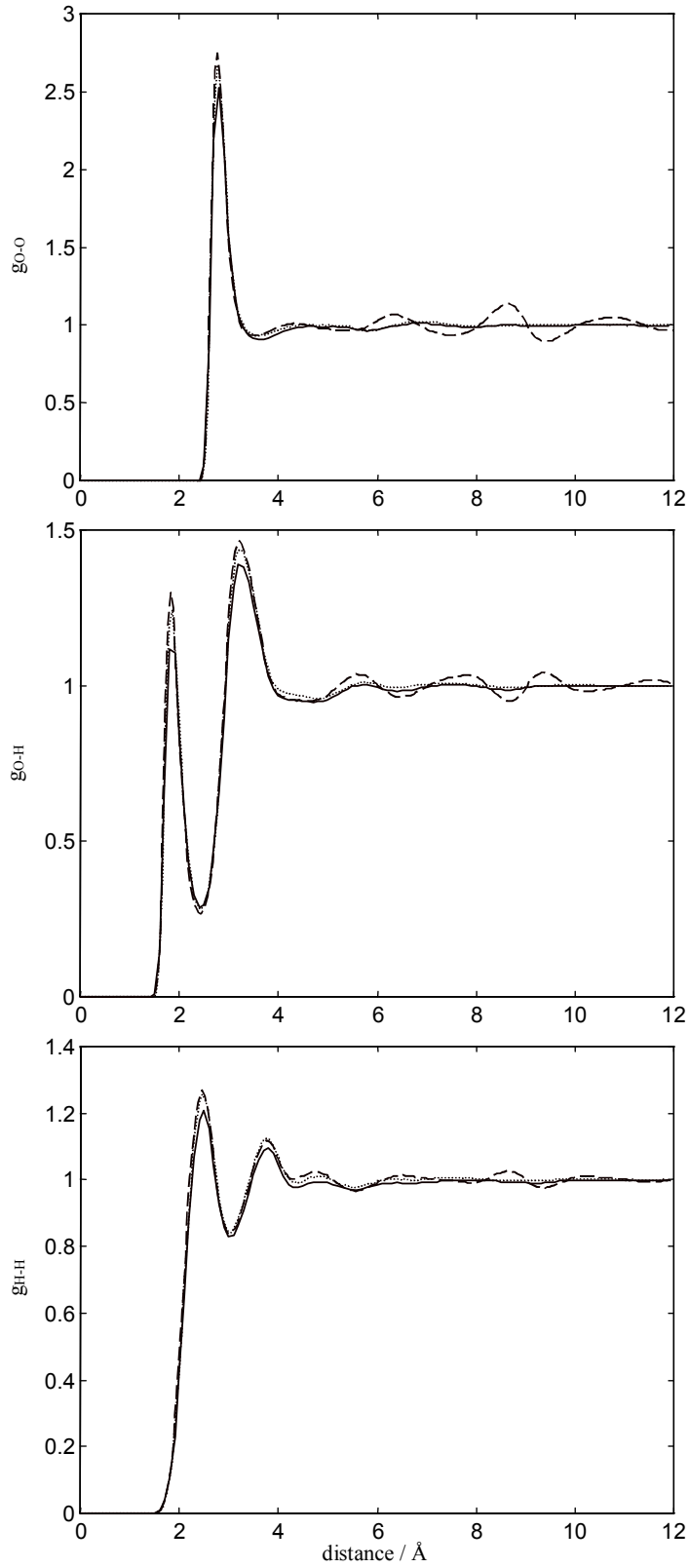

Figure 2.5: Pair correlation function for O-O (top), O-H (middle), H-H (bottom) distances for TIP3P water using damped potential (solid), CHARMM shifted-force (dashed), and Ewald summation (dotted).

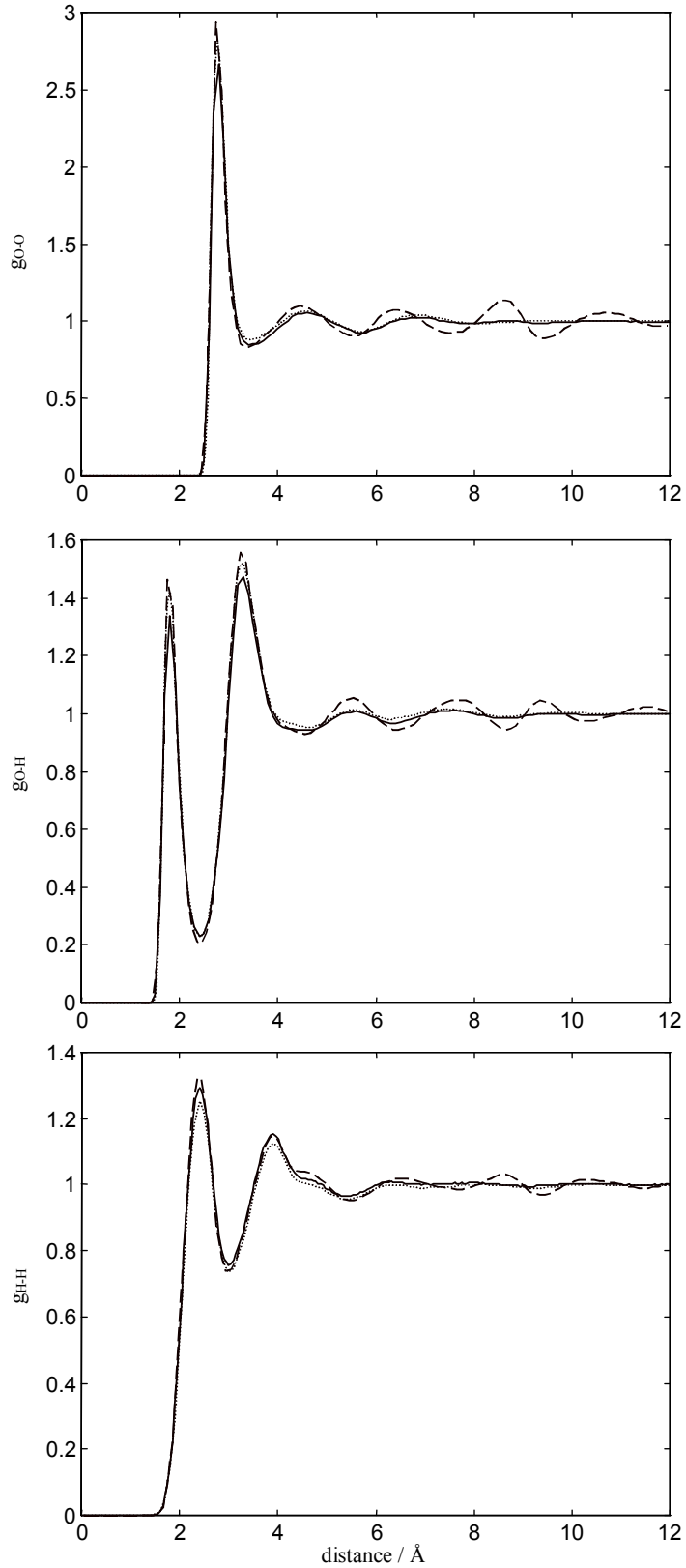

Figure 2.6: Pair correlation function for O-O (top), O-H (middle), H-H (bottom) distances for SPC water using damped potential (solid), CHARMM shifted-force (dashed), and Ewald summation (dotted).

All static properties for the final damped Coulomb parameter set and the reference simulations are summarized in Table 2.2. For the average potential energy additionally the values obtained from constant-volume simulations at the average density of the *NpT* Ewald runs are given. As expected (see also ref. 12), these correspond even better to the Ewald results. Note particularly for the CHARMM potential that the average density is very close to the experimental value while the average potential energy strongly deviates from the Ewald simulation data. This is related to the fact that the parameterization of the water models was done mainly with respect to experimental data by simulation with shifted-force potentials. One can therefore, on the other hand, not expect too close correspondence between Ewald simulation results and experiment.

Table 2.2: Average potential energy $\langle V \rangle$ (numbers in parentheses: *NVT* results at Ewald *NpT* density), water density ρ, dielectric permittivity ε_r, Debye relaxation time τ_D, diffusion constant D, and relative RMS fluctuation of the total energy from *NVE* runs, for TIP3P and SPC water models with damped potential (Da), CHARMM shift (Ch) and Ewald sum (Ew).

	$\langle V \rangle$ / kcal mol^{-1}	ρ / Å^{-3}	ε_r	τ_D / ps	D / 10^{-5}cm^{2}s^{-1}	RMS / 10^{-5}
$TIP3P_{Da}$	-9.05 (-9.14)	0.0320	91	6.4	6.8	9.10
$TIP3P_{Ew}$	-9.51	0.0329	97[23],95[24]	7.3[23],6.1[24]	6.3	2.51
$TIP3P_{Ch}$	-10.60	0.0331	/	/	3.9	3.23
SPC_{Da}	-9.51 (-9.60)	0.0319	64	5.9	4.6	5.51
SPC_{Ew}	-10.00	0.0327	65[23]	7.6[23]	5.8	3.02
SPC_{Ch}	-11.13	0.0333	/	/	2.4	4.39
exp.	/	0.0334[22]	78.54[22]	9.3[21]	2.35[22]	/

2.4.2 Dynamic properties

For evaluating the dynamic properties first the velocity autocorrelation function was calculated

$$c_v(t) = \frac{\langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle}{\langle \mathbf{v}_i(0) \cdot \mathbf{v}_i(0) \rangle} \tag{2.22}$$

where $\mathbf{v}_i$ is the velocity vector of the center of mass of water molecule *i* and the averages are taken over all molecules. As shown in Figure 2.7 excellent agreement with the reference Ewald simulations for both water models is again confirmed, whereas the CHARMM shifted-force results deviate drastically.

The diffusion constants may be calculated by integration of the velocity auto-correlation function according to

$$D = \frac{k_B T}{m} \int_0^\infty c_v(\tau)\, d\tau \tag{2.23}$$

where *m* is the mass of a water molecule. Since the velocity autocorrelation function decays rapidly to zero, integration was only performed up to an upper limit of 1 ps. Table 2.2 shows the diffusion coefficients as obtained from simulation and from experiment. Replacing explicit Ewald summation by damped potentials yields small errors in the computation of *D*. The large discrepancy as compared to the experimental value is entirely due to model deficiencies. Here the correspondence between Ewald and damped Coulomb data is of importance. Again, huge deviations by using the CHARMM potential are observed. Table 2.2 additionally contains the relative RMS fluctuations of the total energy from the *NVE* simulation runs proving stable trajectories.

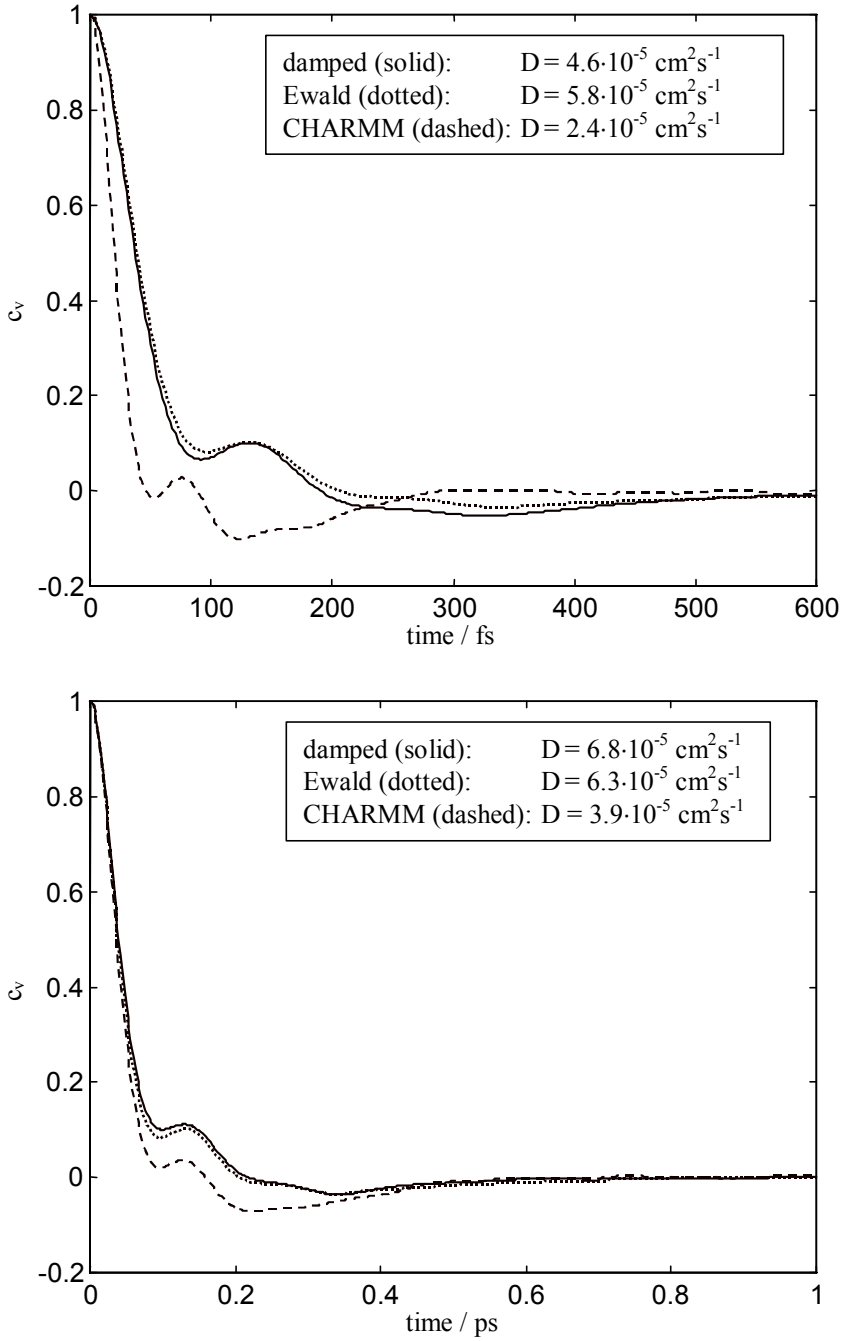

Figure 2.7: Velocity autocorrelation function for SPC (top), TIP3P (bottom) water. The solid curve corresponds to damped Coulomb potential, the dotted curve was obtained from Ewald simulation, the dashed line represents the CHARMM shifted-force result.

2.4.3 Dielectric Properties

Because the Q factor equals to 1 for all combinations of α and R_C as discussed in section 2.3.2, the usual relations can be applied for characterizing the dielectric behavior: The time correlation function of the total dipole moment **M**

$$c_M(t) = \frac{\langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle}{\langle \mathbf{M}(0) \cdot \mathbf{M}(0) \rangle} \tag{2.24}$$

for the various models is depicted in Figure 2.8. The correlation may be approximated by simple exponential decay according to Debye dielectrics; the corresponding Debye relaxation times τ_D are listed in Table 2.2.

The relative dielectric permittivity ε_r was computed from the fluctuation of the total dipole moment via the Kirkwood G factor

$$G = \frac{\left\langle \left(\mathbf{M} - \langle \mathbf{M} \rangle \right)^2 \right\rangle}{N\mu^2} \tag{2.25}$$

as

$$\varepsilon_r = 1 + 3\,G\,\frac{4\pi\rho\mu^2}{9k_BT} \tag{2.26}$$

where N is the total number of water molecules and μ is the dipole moment of a single water molecule. The results are summarized in Table 2.2 and compared to experiments[21,22] and to simulations of Steinhauser and coworkers[23] and of Chandra and Ichiye[24] using the Ewald summation method. Both works used the experimental water density while here the system density was kept at the value corresponding to 1 bar from earlier simulations. The deviation is however small. More significantly, while Chandra and Ichiye applied *NVE* conditions as in this work, Steinhauser and coworkers used a thermostat while sampling dynamical quantities. This affects the system dynamics dramatically. Chandra and Ichiye[24] therefore provide a much better source for comparison, though they give data for TIP3P model only. Since typical errors for the dielectric permittivity amount to roughly 5 and for the Debye relaxation times ca. 0.7 ps,[23] the results correspond to Ewald data within error bars, at least in the case of TIP3P water. However, as can be seen from the diffusion coefficients, using the damped Coulomb scheme leads to slightly slower molecular motion. There

is no apparent reason for this behavior, though one might argue that the use of the damping scheme implies some loss of long-range dynamical correlation. On the other hand, it cannot ruled out the possibility that the Ewald results are contaminated by artifacts due to the artificial periodicity imposed on a liquid system.

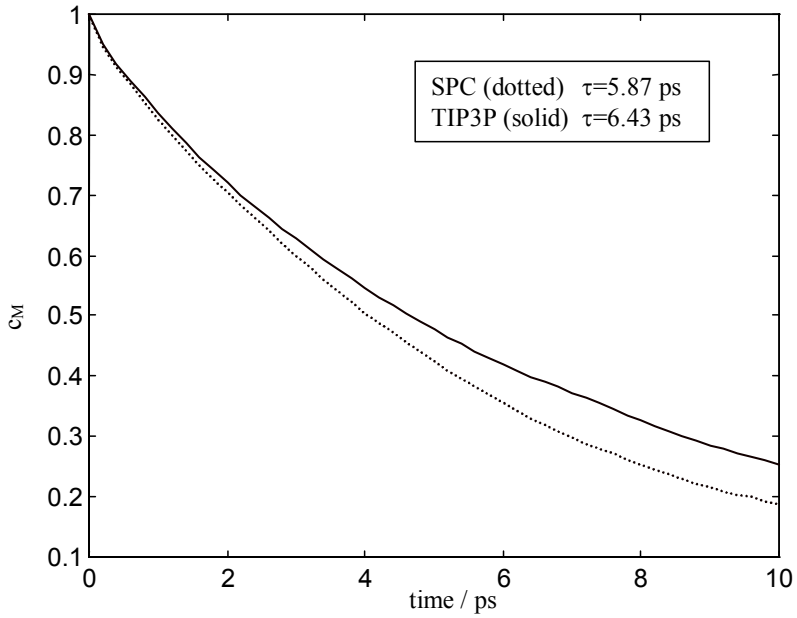

Figure 2.8: Dipole moment time correlation function for damped TIP3P (solid) and SPC (dotted) water models.

2.5 Conclusion

The damped Coulomb potential was applied to molecular dynamics simulations of TIP3P and SPC water models in order to demonstrate its general applicability to molecular liquids. To this end, the original theory given by Wolf *et al.*[10] was adapted and enhanced to yield potential and force consistency with the side effect of significant improvement of dielectric properties. With referral to Ewald summation data, the parameterization of the damping constant and cutoff radius was accomplished by a strategy which is generally transferable to other systems as well. The set of final parameters was chosen with respect to minimizing cutoff effects in a number of observables, both static and dynamic ones. For TIP3P and SPC water a combination of a damping constant of 0.2 $Å^{-1}$ and a cutoff distance of 9 Å was found to be appropriate. As a general result, it is found that close correspondence with

Ewald data is achieved at less computational cost than the Ewald scheme: With the barely optimized code the damped Coulomb computations compared to the traditional Ewald implementation are roughly twice as fast, compared to particle-mesh Ewald[25] the speed-up is around 25 %, in line with expectations.

Static properties like radial distribution functions were produced in excellent agreement to the reference simulations. The same applies to the average potential energy per molecule and average particle density that were the primary properties used for parameterization of damping constant and cutoff radius. As an example for the improvement over existing shifted-force schemes, additionally the CHARMM function was applied. Particularly in the cutoff distance proximity, well-known artifacts are effectively eliminated by using the damped Coulomb formalism. Furthermore, fluctuations, time correlation functions and related data were computed. The dielectric permittivity and the Debye relaxation time of damped Coulomb simulations correspond to Ewald summation data within error bars, while the diffusion coefficients are slightly smaller in the case of the damped Coulomb potential. The deviations are more significant than for static properties, yet very small. Again, applying the CHARMM potential introduces huge deviations.

2.6 References

(1) P.P Ewald, *Ann. Phys. (Leipzig)* **64**, 253 (1921).

(2) G. Hummer, D. K. Soumpasis, *Phys. Rev. E* **49**, 591 (1994).

(3) D. van der Spoel, P. J. van Maaren, H. J. C. Berendsen, *Chem. Phys.* **108**, 10220 (1998).

(4) P. J. Steinbach, B. R. Brooks, *J. Comp. Chem.* **15**, 667 (1994).

(5) C. L. Brooks, *J. Chem. Phys.* **86**, 5156 (1987).

(6) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.* **4**, 187 (1983).

(7) G. Schrimpf, M. Schlenkrich, J. Brickmann, P. A. Bopp, *J. Phys. Chem.* **96**, 7404 (1992).

(8) H. Dufner, S. M. Kast, J. Brickmann, M. Schlenkrich, *J. Comp. Chem.* **18**, 660 (1997).

(9) H. Resat,J. A. McCammon, *J. Chem. Phys.* **108**, 9617 (1998).

(10) D. Wolf, P. Keblinski,S. R. Phillpot, J. Eggebrecht, *J. Chem. Phys.* **110**, 8254 (1999).

(11) D. Wolf, *Phys. Rev. Lett.* **68**, 3315 (1992).

(12) P. Demontis, S. Spanu, G. B. Suffritti, *J. Chem. Phys.* **114**, 7980 (2001).

(13) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).

(14) H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht 1981, p 331.

(15) M. Neumann, O. Steinhauser, *Chem. Phys. Lett.* **95**, 417 (1983).

(16) M. Neumann, *Molec. Phys.* **50**, 841 (1983).

(17) T. R. Forester, W. Smith, *DL_POLY molecular dynamics code*, version 2.14, CCP5 of the EPSRC (1995).

(18) J. P. Rykaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).

(19) H. J. C. Berendsen,J. P. M. Postma, W. F. van Gunsteren, A. DiNola,J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).

(20) S. M. Kast, J. Brickmann, *J. Chem. Phys.* **104**, 3732 (1996).

(21) J. B. Hasted, In *Water: A Comprehensive Treatise*, Franks, F. Ed.; Plenium: New York, 1972, Vol. 1, Chap. 7.

(22) *CRC Handbook of Chemistry and Physics*, Lide, D. R., Ed.; Chemical Rubber: Boca Raton, 1994.

(23) P. Höchtl, S. Boresch, W. Bitomsky, O. Steinhauser, *J. Chem. Phys.* **109**, 4927 (1998).

(24) A. Chandra, T. Ichiye, *J. Chem. Phys.* **111**, 2701 (1999).

(25) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).

3 Correction of truncation effects

3.1 Introduction

Despite recent progress in the fast computational treatment of long-range interactions in molecular simulations, like the particle-mesh Ewald simulation technique[1], the speed increase achieved by using truncated potentials with small cutoff distances is undisputed. If only the real space and pairwise spherical potentials are considered, cutting a truncation distance by half nominally leads to a decrease of the computational burden for calculating nonbonded interactions by a factor of eight. The distorting effects of applying simple potential truncation, switching or shifting functions as compared to the presumably "exact" Ewald summation[2] for Coulomb potentials are known for quite some time.[3-5] Besides molecular simulations, truncation artifacts have also been studied on the basis of an approximate integral equation theory.[6]

If substantial artificial shifts of computed observables are expected from simulation with truncated potentials, a fast corrective tool is needed in order to allow for an extrapolation towards results for the full potential. Statistical-mechanical integral equation (IE) theory,[7] though approximate in nature, provides the necessary tools: Verlet[8] was the first who used IE theory for the purpose of extrapolating the radial distribution functions of simple liquids from molecular dynamics (MD) simulations beyond the range determined by the simulation cell. Similarly, Trokhymchuk *et al.*[9] estimated the long range structure of liquid water and later applied the extrapolation technique to the calculation of small wavenumber structure factors.[10] Besides mere extrapolation, Kambayashi and Chihara[11] corrected potential truncation artefacts in the case of simple monoatomic liquids by extracting the so-called "bridge function", that serves as the link between approximate and exact IE theory, from a truncated simulation. Assuming independence of the truncation, the simulated bridge function was then used for solving the Ornstein-Zernike integral equation with the full potential.

Quite recently, it was shown based on perturbation-theoretic arguments[12] that an IE-based cutoff artifact correction is also possible for molecular liquids. In this work both correction strategies are taken into account, the perturbation approach as well as the Kambayashi/Chihara technique. Two descriptors measuring the performance are developed: The internal energy as a global property and a similarity measure characterising local differences between corrected and Ewald-limit radial distribution functions. Corrections are applied to a range of cutoff distances used in simulations of liquid water, considering two shifted-force potentials acting on the long range Coulomb interaction: The well-known CHARMM function[4] and the damped Coulomb potential[13] in its recently enhanced form.[14] In contrast to most other shifted-force functions, the latter is derived from strictly physical arguments: The effect of spherical truncation is rationalised in the sense of a net charge generated within the cutoff sphere that is compensated by a smoothing function. This term contains adjustable parameters, the parameterisation of which will also be addressed on the basis of integral equation theory in this chapter.

3.2 Theory and methods

3.2.1 Cutoff correction

The one-dimensional reference interaction site model (RISM) integral equation[7,15] for fluids composed of molecules with sites α and γ as used in this chapter is

$$\mathbf{h} = \boldsymbol{\omega} * \mathbf{c} * \boldsymbol{\omega} + \boldsymbol{\omega} * \mathbf{c} * (\rho \mathbf{h}) \tag{3.1}$$

where $\mathbf{h} = (h_{\alpha\gamma}(r))$ is the matrix of the total correlation functions (the radial distribution function is $g = h+1$), $\mathbf{c} = (c_{\alpha\gamma}(r))$ are the direct correlation functions, $\boldsymbol{\omega} = (\omega_{\alpha\gamma}(r))$ are the intramolecular correlation functions (for rigid molecules these are normalized Dirac delta functions constraining the site distance), ρ is the site density, r is the distance and the star denotes convolution. This equation must be supplied with a closure relation,

$$h_{\alpha\gamma}(r) + 1 = \exp(-\beta u_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r) + B_{\alpha\gamma}(r)) , \tag{3.2}$$

with the pair potential $u_{\alpha\gamma}$ and $\beta = 1/kT$ where k is the Boltzmann constant and T is the absolute temperature; $B_{\alpha\gamma}$ are the bridge functions. In the hypernetted chain

(HNC) approximation B identically vanishes. The bridge functions are extracted from MD simulation data by constraining the RISM g functions to the values of the simulation, g^{MD}, according to Roux and co-workers:[16]

$$B_{\alpha\gamma}(r) = f(r)[\ln(g_{\alpha\gamma}^{\mathrm{MD}}(r)) + \beta u_{\alpha\gamma}(r) - h_{\alpha\gamma}(r) + c_{\alpha\gamma}(r)] \tag{3.3}$$

where $f(r)$ is a cubic polynomial acting as switching function that varies smoothly between 1 to 0 over a range defined by the simulation cell. Beyond this range the correlation functions are extrapolated within the HNC approximation.

The first method to correct for cutoff artifacts is adopted from the Kambayashi / Chihara direct bridge function approach:[11] By assuming that the bridge function is largely independent of the cutoff distance r_c,

$$B_{\alpha\gamma}(r; r_c \to \infty) \approx B_{\alpha\gamma}(r; r_c), \tag{3.4}$$

B can be taken from a truncated potential simulation and directly applied in solving the coupled eqs. (3.1) and (3.2) with the full potential. As can be seen from (3.3), in regions where g^{MD} is zero the explicit value of B is numerically ill-defined, i.e. particularly for $r \to 0$. This does not play a role for the truncation correction since the potential is strongly repulsive in this region, absorbing the bridge function.

The second approach originates from a first order functional perturbation expansion of the free energy in terms of Mayer functions of the bridge term. In references 12 and 17 the derivation was given for the chemical potential. For the Helmholtz free energy, A, it can be analogously obtained

$$A^{\mathrm{ex,MD}} \approx A^{\mathrm{ex,HNC}}[\mathbf{h}^{\mathrm{HNC}}, \mathbf{c}^{\mathrm{HNC}}] + \frac{\beta^{-1}\rho N}{2} \sum_{\alpha\gamma} \int d\mathbf{r}\, g_{\alpha\gamma}^{\mathrm{HNC}}(r)[\exp(B_{\alpha\gamma}(r)) - 1]. \tag{3.5}$$

Since the dominant truncation effects on g are located around the cutoff distance[4,6] where the bridge term is small, the second term approximately cancels upon subtracting expressions for full and truncated potential, yielding

$$A^{\mathrm{ex,MD}}(r_c \to \infty) - A^{\mathrm{ex,MD}}(r_c) \approx A^{\mathrm{ex,HNC}}(r_c \to \infty) - A^{\mathrm{ex,HNC}}(r_c). \tag{3.6}$$

This relation can now be used in the reversible work theorem, applied for the true (superscript MD) or integral equation distribution functions (superscript HNC)

$$\ln \frac{g^{\mathrm{MD/HNC}}(r;r_c \to \infty)}{g^{\mathrm{MD/HNC}}(r;r_c)} = \\ -\beta\left(\Delta A^{\mathrm{ex,MD/HNC}}(r;r_c \to \infty) - \Delta A^{\mathrm{ex,MD/HNC}}(r;r_c) \right. \\ \left. + u(r;r_c \to \infty) - u(r;r_c)\right) \tag{3.7}$$

with

$$\Delta A^{\mathrm{ex,MD/HNC}}(r;r_c) = A^{\mathrm{ex,MD/HNC}}(r;r_c) - A^{\mathrm{ex,MD/HNC}}(r \to \infty;r_c) \tag{3.8}$$

where $A(r)$ means the free energy of a system in which a certain atomic distance r is constrained. Together with (3.6) finally the compact expression

$$\frac{g^{\mathrm{MD}}(r;r_c \to \infty)}{g^{\mathrm{MD}}(r;r_c)} \approx \frac{g^{\mathrm{HNC}}(r;r_c \to \infty)}{g^{\mathrm{HNC}}(r;r_c)} \tag{3.9}$$

is obtained which offers a convenient route to a cutoff correction even in the absence of explicit bridge function data.

3.2.2 Models and numerical methods

The rigid 3-site TIP3P water model[18] was studied using either the full potential (implying Ewald summation[1] in the case of MD simulations) or shifted-force potentials with varying cutoff distances. The Coulomb interaction due to partial site charges q_i in these cases was modified according to the CHARMM function,[4]

$$u_{ij}^{\mathrm{CH}}(r) = \frac{q_i q_j}{r}[1-(r/r_c)^2]^2, \tag{3.10}$$

and to the enhanced damped Coulomb formalism,[14]

$$u_{ij}^{\mathrm{DC}}(r) = q_i q_j \left\{ \frac{\mathrm{erfc}(\alpha r)}{r} - \left[\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 r_c^2)}{r_c} \right](r - r_c) \right\}, \tag{3.11}$$

where the damping constant α needs to be optimised for a given r_c. The Lennard-Jones interaction was truncated with the corresponding CHARMM expression[4] at the cutoff specified for Coulomb interactions, up to a maximum of 12 Å that was also used as real space cutoff distance in the Ewald simulation. Within the RISM theory small Lennard-Jones parameters were attributed to the hydrogen sites in order to avoid numerical singularities.

MD simulations were performed in the canonical ensemble[19] at a temperature of 300 K and a density of 0.03341 $Å^{-3}$ with a rectangular simulation cell comprising 1000 water molecules under periodic boundary conditions. After equilibration for at least 200 ps, 2000 configurations were sampled from a 200 ps trajectory using a time step of 2 fs. From these, reference radial distribution functions were obtained with a histogram spacing of 0.05 Å. Assuming statistical independence, the absolute error of the distribution functions reaches a maximum of around 0.004 at the onset of the first peak and is on average 0.001-0.002 throughout the rest.

The RISM solutions were obtained on a logarithmic grid of 512 points ranging from $5.98 \cdot 10^{-3}$ Å to a maximum distance of 164.02 Å at the same density and temperature as in the MD simulations. As discussed in reference 12 neither the dielectrically consistent RISM equations[20] nor the simple charge-scaling procedure[21] were used to adjust the dielectric constant at full potential. The nonlinear system of equations was solved using a variant of the "modified direct inversion of iterative subspace" (MDIIS) method[22] for coarse optimisation of the correlation functions and the traditional Picard iteration scheme for ultimate convergence to a threshold of $\max(\Delta\mathbf{c}) < 10^{-7}$ between two successive iteration steps.

In the case of eq. (3.3) for extracting the bridge functions, the function $f(r)$ switching between the simulation results and the HNC solution was applied over a range between 10.975 and 11.975 Å. The sampling points from the reference radial distribution functions were smoothed and transformed to those of the integral equations by means of cubic spline interpolation with a smoothing factor given by the number of histogram bins in g^{MD}, considering the respective statistical uncertainties.[23] It turned out that stable numerical results appropriate for further computations can be obtained by first iterating up to $\max(\Delta(\mathbf{h}\text{-}\mathbf{c})) < 10^{-3}$ within the MDIIS scheme, after which the bridge function was explicitly obtained from eq. (3.3) by setting g^{MD} values less than 10^{-5} to exactly 10^{-5} and potentials larger than 10^5 kJ mol^{-1} to 10^5 kJ mol^{-1}. The resulting bridge functions were then used in iterations towards the target convergence threshold. An analogous procedure was applied for the perturbative correction where now $g^{MD}(r_c)g^{HNC}(r_c \rightarrow \infty)/g^{HNC}(r_c)$

plays the role of the reference distribution function. In this case, the ratio of HNC solutions was set to one for distances where any g^{HNC} becomes less than 0.1.

3.3 Results and discussion

The performance of the various correction schemes was measured by a global property, the excess internal energy

$$U^{\mathrm{ex}} = \frac{\rho}{2} \sum_{\alpha\gamma} \int d\mathbf{r}\, g_{\alpha\gamma}(r) u_{\alpha\gamma}(r) \tag{3.12}$$

and by a quantity characterising local differences, the relative similarity

$$S = 1 - \frac{\sum_{ij=\mathrm{HH,OH,OO}} \int_0^{r_{\max}} d\mathbf{r} \left| g_{ij}(r;r_c) - g_{ij}^{\mathrm{ref}}(r;r_c \to \infty) \right|}{\sum_{ij=\mathrm{HH,OH,OO}} \int_0^{r_{\max}} d\mathbf{r}\, g_{ij}^{\mathrm{ref}}(r;r_c \to \infty)}. \tag{3.13}$$

While eq. (3.13) is used here (with $r_{\max}$ = 11.975 Å) for comparing reference Ewald (superscript "ref"="MD") and corrected data (no superscript), it also represents a useful means for the parameterisation of the damped Coulomb potential. In chapter 2 ist was shown that an energy-based parameterisation leads to static, dynamic and dielectric results in good agreement with results from Ewald simulations. Both, the cutoff distance and the damping parameter, were simultaneously optimised with respect to minimal deviation from Ewald simulation energies, while all other properties were examined keeping the parameters fixed. Now, the opportunity is used to optimise α for a given r_c with respect to structural properties only since energies will be computed with the original, full potential by integrating over corrected distribution functions. The tedious simulation work can be avoided by maximising S on the basis of HNC integral equation solutions alone: With $g^{\mathrm{ref}} = g^{\mathrm{HNC}}(r_c \to \infty)$ and $g = g^{\mathrm{HNC,DC}}(\alpha, r_c)$, and integration up to $r^{\max}$ = 20 Å, the optimal combinations to be used subsequently are (α, r_c) = (0.37 Å^{-1}, 8 Å), (0.41 Å^{-1}, 7 Å), and (0.44 Å^{-1}, 6 Å).

The performance of the two correction techniques is qualitatively shown in Figs. 3.1 and 3.2: The H-H, O-H and O-O radial distribution functions are depicted with particular emphasis on the cutoff region where the most pronounced artifacts are expected. For the CHARMM function (as shown in ref. 12 with Ewald simulation results under constant pressure conditions at a slightly different average density) the

improvement is quite evident while for the damped Coulomb form the original *g* functions have, as expected, already such a good quality that neither correction leads to a substantial change. It cannot be expected however to compensate for the remaining small differences due to the inherently approximate nature of eqs. (3.4) and (3.9): As is clear from (3.3), even if the correlation functions were identical for the full and damped Coulomb approach, the bridge functions will differ since the full and the damped Coulomb potential are significantly dissimilar. It can be furthermore not expected to correct for arbitrarily small cutoff distances due to a breakdown of the underlying assumptions if the range of the bridge functions interferes with regions of notable cutoff artifacts.

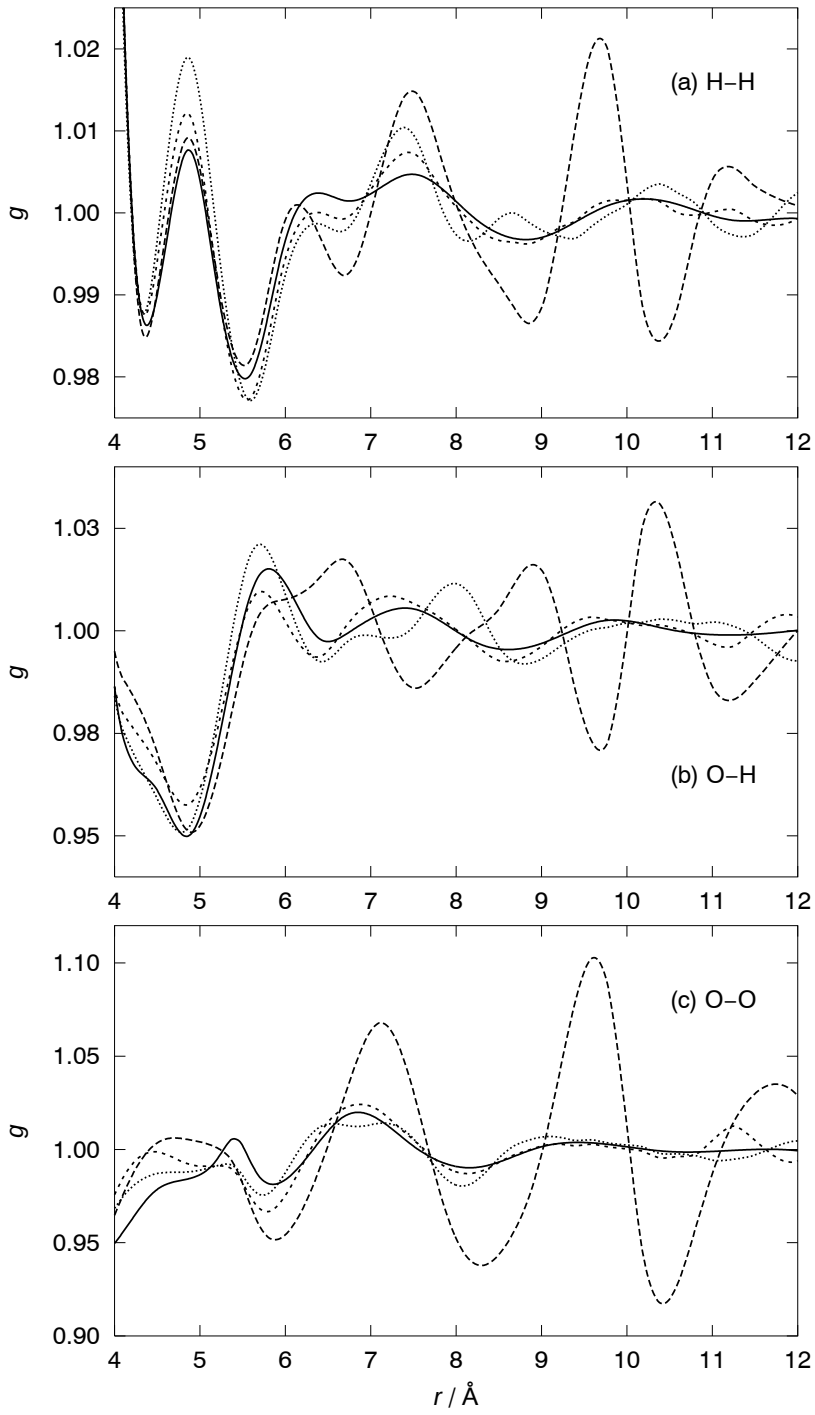

Figure 3.1: Radial pair distribution functions of TIP3P water: (a) hydrogen-hydrogen, (b) oxygen-hydrogen, (c) oxygen-oxygen; Ewald simulation result (solid), 10 Å CHARMM truncation simulation (long-dashed), corrected with explicit bridge function (short-dashed), corrected with perturbative method (dotted).

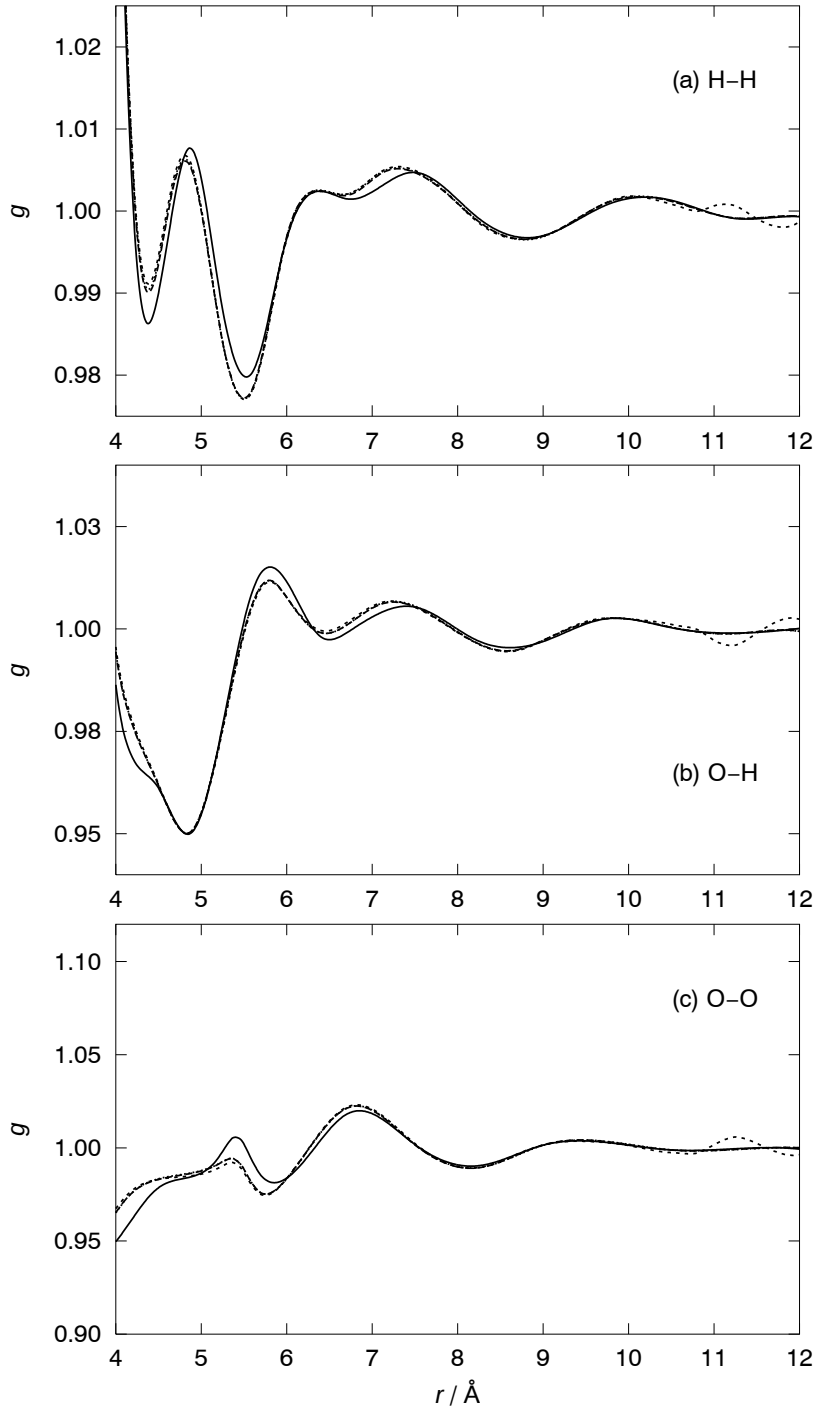

Figure 3.2: Radial pair distribution functions of TIP3P water: (a) hydrogen-hydrogen, (b) oxygen-hydrogen, (c) oxygen-oxygen; Ewald simulation result (solid), 8 Å damped Coulomb truncation simulation (long-dashed), corrected with explicit bridge function (short-dashed), corrected with perturbative method (dotted).

Quantitative results for various truncation conditions are summarized in the Table 3.1 and in Fig. 3.3. In order to check whether the long-range HNC extrapolation influences the results for the internal energy, the respective potential used in the simulation was integrated over the corresponding, extrapolated radial distribution functions to give $U_B^{\mathrm{ex,orig}}$, and compared with the expectation value directly obtained from the simulation, $U_{\mathrm{MD}}^{\mathrm{ex}}$: For all cutoff distances studied the results agree within 0.5 kJ mol^{-1}. As expected, the damped Coulomb energies deviate much more drastically from the Ewald result than the CHARMM data. A rather unexpected outcome is obtained if the full Coulomb potential is integrated over the *uncorrected g* functions ($U_B^{\mathrm{ex},0}$): Despite the large discrepancies between Ewald-limit and shifted-force distribution functions in the CHARMM case, the energy is very close to the full Coulomb energy, while the damped Coulomb results are even slightly worse.

The corrected data in Table 3.1 and in Figure 3.3, supplemented by the relative similarity S_B^{orig} between original and Ewald *g* functions, indicates the performance of the presented approaches: The CHARMM potential can be safely used with either method down to a cutoff distance of 10 Å, while the similarity, corrected or not, of the damped Coulomb potential is unsurpassed. Beyond 9-10 Å, the correction of the CHARMM function becomes unstable, even more pronounced so for the perturbative technique. With smaller cutoff distances only the damped Coulomb function is usable. In this case also the explicit bridge function approach appears to be slightly advantageous for the energy, although the perturbative correction works better for the relative similarity. Still smaller cutoff distances cannot be managed with the present formalism due to the approximations invoked and numerical difficulties in finding convergent solutions to the integral equations.

Table 3.1: Energetic und structural results for various correction methods and cutoff distances (CH: CHARMM, DC: enhanced damped Coulomb, EW: Ewald):

r_c / Å	∞(EW)	12(CH)	10(CH)	9(CH)
$U_{\mathrm{MD}}^{\mathrm{ex}}$ / kJ mol^{-1}	-40.98	-42.24	-43.32	-44.30
$U_{B}^{\mathrm{ex,orig}}$ / kJ mol^{-1}	-40.68	-42.85	-43.85	-44.76
$U_{B}^{\mathrm{ex,0}}$ / kJ mol^{-1}		-40.70	-40.65	-40.67
$U_{B}^{\mathrm{ex,corr}}$ / kJ mol^{-1}		-40.37	-40.62	-40.80
$U_{\mathrm{perturb}}^{\mathrm{ex,corr}}$ / kJ mol^{-1}		-41.00	-40.90	-43.98
S_{B}^{orig}		0.9892	0.9803	0.9756
S_{B}^{corr}		0.9968	0.9966	0.9965
$S_{\mathrm{perturb}}^{\mathrm{corr}}$		0.9961	0.9961	0.9943

r_c / Å	8(CH)	8(DC)	7(DC)	6(DC)
$U_{\mathrm{MD}}^{\mathrm{ex}}$ / kJ mol^{-1}	-45.66	-32.50	-29.66	-26.80
$U_{B}^{\mathrm{ex,orig}}$ / kJ mol^{-1}	-44.98	-32.51	-30.15	-27.39
$U_{B}^{\mathrm{ex,0}}$ / kJ mol^{-1}	-39.50	-38.95	-39.07	-38.81
$U_{B}^{\mathrm{ex,corr}}$ / kJ mol^{-1}	-38.95	-39.72	-40.25	-40.41
$U_{\mathrm{perturb}}^{\mathrm{ex,corr}}$ / kJ mol^{-1}	-36.72	-40.78	-39.64	-38.80
S_{B}^{orig}	0.9715	0.9985	0.9984	0.9981
S_{B}^{corr}	0.9958	0.9978	0.9972	0.9971
$S_{\mathrm{perturb}}^{\mathrm{corr}}$	0.9941	0.9985	0.9984	0.9983

Excess internal energy from simulation $U_{\mathrm{MD}}^{\mathrm{ex}}$ and from extracted bridge function $U_{B}^{\mathrm{ex,orig}}$, original distribution functions with full Coulomb potential $U_{B}^{\mathrm{ex,0}}$, corrected with explicit bridge function $U_{B}^{\mathrm{ex,corr}}$, corrected with perturbative method $U_{\mathrm{perturb}}^{\mathrm{ex,corr}}$; relative similarity to Ewald distribution functions for shifted-force simulation S_{B}^{orig}, corrected with explicit bridge function S_{B}^{corr}, corrected with perturbative method $S_{\mathrm{perturb}}^{\mathrm{corr}}$.

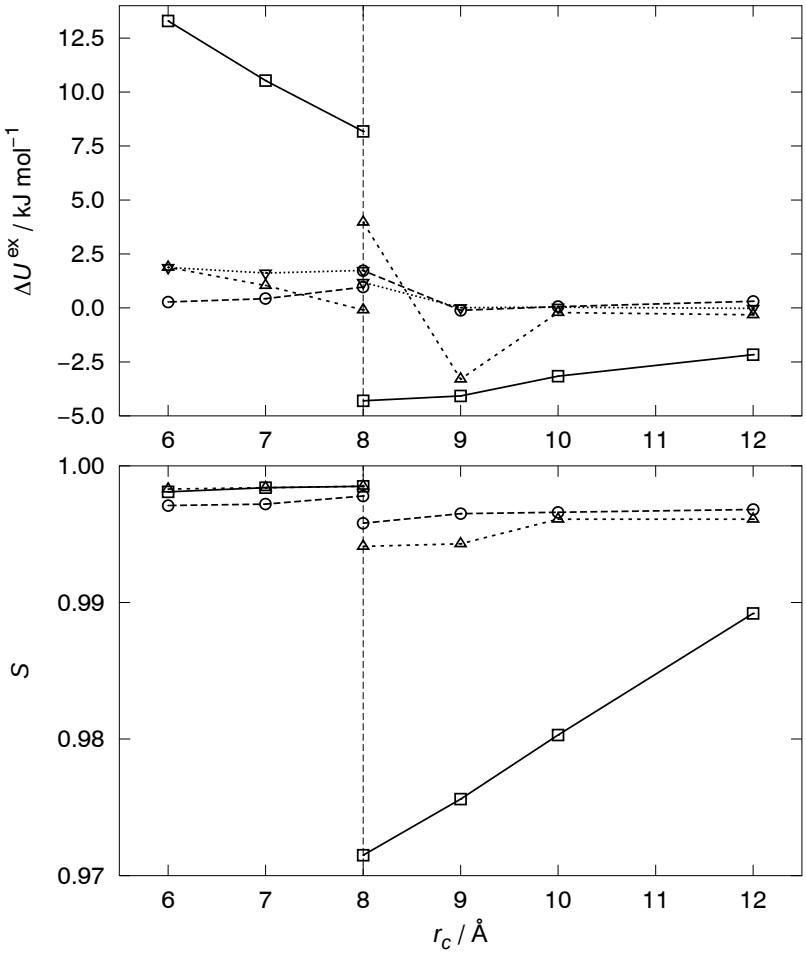

Figure 3.3: Energetic und structural results for various correction methods and cutoff distances, up to 8 Å: damped Coulomb potential, 8 Å and above: CHARMM potential. Top: Difference to Ewald excess internal energy $U_B^{\mathrm{ex}}(\mathrm{EW})$ for shifted-force simulation energy $U_B^{\mathrm{ex,orig}}(\mathrm{CH/DC})$ (solid), corrected with explicit bridge function $U_B^{\mathrm{ex,corr}}$ (long-dashed), corrected with perturbative method $U_{\mathrm{perturb}}^{\mathrm{ex,corr}}$ (short-dashed), original distribution functions with full Coulomb potential $U_B^{\mathrm{ex,0}}$ (dotted). Bottom: Relative similarity to Ewald distribution functions for shifted-force simulation S_B^{orig} (solid), corrected with explicit bridge function S_B^{corr} (long-dashed), corrected with perturbative method $S_{\mathrm{perturb}}^{\mathrm{corr}}$ (short-dashed).

3.4 Conclusion

It has been demonstrated that integral equation theory is a useful tool for correcting molecular simulation results generated with truncated potentials, and for the parameterisation of a specific truncation scheme, the enhanced damped Coulomb potential. The observables examined in this work were first-order expectation values, namely distribution functions and internal energy. As it turns out, the correction based on the bridge functions explicitly extracted from structural information is overall slightly more robust compared to the perturbation scheme, the latter being quite useful nonetheless. In the damped Coulomb case, good results are obtained down to a cutoff distance of only 6 Å.

In contrast to pure phases, solute-solvent distribution functions, e.g. for a single molecule in solution are difficult to obtain with sufficient accuracy, prohibiting a straightforward application of the explicit bridge function technique. In these cases the perturbative scheme will be much more useful at least for qualitative purposes. For quantitative corrections in the presence of statistical noise an alternative strategy could be employed: Within the perturbation-theoretic formalism the corrected first-order expectation value of a spatially dependent quantity $X(r)$ can be written as $\langle X(r)\, g^{\mathrm{HNC}}(r;r_c\to\infty) / g^{\mathrm{HNC}}(r;r_c)\rangle$, the average taken over the truncated potential ensemble. This allows for a reweighting correction during the sampling run itself.

3.5 References

(1) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **103** , 8577 (1995).

(2) P. P. Ewald, *Ann. Phys. (Leipzig)* **64**, 253 (1921).

(3) P. E. Smith, B. M. Pettitt, *J. Chem. Phys.* **95**, 8439 (1991).

(4) P. J. Steinbach, B. R. Brooks, *J. Comp. Chem.* **15**, 667 (1994).

(5) H. Dufner, S. M. Kast, M. Schlenkrich, J. Brickmann, *J. Comp. Chem.* **18**, 660 (1997).

(6) C. L. Brooks III, *J. Chem. Phys.* **86**, 5156 (1987).

(7) J.-P. Hansen, I. R. McDonald, *Theory of simple liquids*, 2nd edn., Academic Press, London, 1991.

(8) L. Verlet, *Phys. Rev.* **163**, 201 (1968).

(9) A. D. Trokhymchuk, M. F. Holovko, E. Spohr, K. Heinzinger, *Mol. Phys.* **77**, 903 (1992).

(10) A. D. Trokhymchuk, M. F. Holovko, K. Heinzinger, *Z. Naturforsch*., **50a,** 18 (1995).

(11) S. Kambayashi, J. Chihara, *Phys. Rev. E* **50**, 1317 (1994).

(12) S. M. Kast, *Phys. Chem. Chem. Phys.* **3**, 5087 (2001).

(13) D. Wolf, P. Keblinski, S. R. Phillpot, J. Eggebrecht, J., *J. Chem. Phys.* **110**, 8254 (1999).

(14) D. Zahn, B. Schilling, S. Kast, *J. Phys. Chem. B* **106**, 10725 (2002).

(15) D. Chandler, H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

(16) Q. Du, D. Beglov, B. Roux, *J. Phys. Chem. B* **104**, 796 (2000).

(17) A. Kovalenko, F. Hirata, *J. Chem. Phys.* **113**, 2793 (2000).

(18) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).

(19) S. M. Kast, J. Brickmann, *J. Chem. Phys.* **104**, 3732 (1996).

(20) J. Perkyns, B. M. Pettitt, *J. Chem. Phys.* **97**, 7656 (1992).

(21) P. J. Rossky, B. M. Pettitt, G. Stell, *Mol. Phys.*, **50**, 1263 (1983).

(22) A. Kovalenko, S. Ten-no, F. Hirata, *J. Comp. Chem.* **20**, 928 (1999).

(23) P. Diercks, *Curve and surface fitting with splines*, Oxford University Press, Oxford, 1993.

4 Application to free energy simulation

4.1 Abstract

The calculation of absolute hydration free energies using thermodynamic integration or free energy perturbation theory within molecular simulations is known to be contaminated by considerable artifacts when potential truncation is employed to speed up the computation. In this work corrective contributions beyond the dielectric continuum approximation for the hydration free energy (i.e. the excess chemical potential upon solvation) with respect to the full potential are analyzed for the case that both the solute-solvent as well as the solvent-solvent potentials are truncated and modified by a shifted-force term. The model systems argon in TIP3P water and pure TIP3P water are used as examples for apolar and polar solutes, revealing significant shifting and truncation contributions even for the short-ranged dispersive interactions and the magnitude of solute-solvent and solvent-solvent components. A variety of techniques is applied to predict this effect solely from the truncated simulation, based upon perturbation-theoretic and other simulation methodology on one hand and the extended reference interaction site model integral equation theory on the other. The latter is shown to be capable of quantitatively predicting truncation artifacts at negligible computational overhead, thereby offering a cost-effective and powerful tool to be used in conjunction with simple truncated potential simulations. Furthermore, free energy finite size effects in shifted-force simulations associated with small simulation cells are discussed.

4.2 Introduction

Free energy calculations have become an important and most valuable tool to study the thermodynamics of solvation at the atomic level. The main fields of interest are free energies of solvation,[1-4] phase equilibria[5,6] or the thermodynamics of ligand binding.[7] In combination with molecular dynamics (MD) or Monte-Carlo (MC) simulations a number of methods are established to obtain free energy differences.[8-10] Widely used are thermodynamic integration (TI) and free energy perturbation (FEP).

The TI method is considered to be more robust and can handle even extensive system changes,[2] whereas FEP calculations are sometimes favored due to the simple implementation but have in some cases been shown to produce systematic errors.[10] The umbrella sampling technique[11] provides the construction of free energy profiles along predefined reaction paths. A variety of new strategies have been proposed recently.[12-18] Intense effort is currently invested by several groups for increasing the efficiency of phase space sampling and thus the accuracy of free energy values.

The handling of long-range interactions in molecular simulations is a crucial point concerning the accuracy of simulation results. Efficient simulation algorithms generally use finite size simulation boxes under periodic boundary conditions.[19] Long-range potentials are approximated by spherical cutoffs,[20] shifting functions,[20] the reaction field method,[21] or Ewald summation.[22,23] All methods have been shown to introduce to various degrees some kind of artifact in calculated observables.[24-31] Two main sources of error are recognized: The use of a limited number of molecules in the simulation cell produces finite size effects whereas the approximation of interaction potentials results in truncation artifacts. At spherical cutoff conditions within a periodic boundary framework both effects influence the results. In principle the Ewald technique offers the best description of the real potential though even the Ewald sum exhibits finite size effects due to periodicity at small simulation cells.[24,27,28]

The accurate calculation of the free energy of solvation constitutes a major task in the field of molecular simulations. For quite some time computations of solvation free energies were restricted to relative free energy differences.[32,33] But with the advent of sophisticated scaling methods[34-39] the routine evaluation of absolute solvation free energies has been accessible. The hydration free energy of apolar,[40-43] polar,[44-46] and ionic[47-52] solutes have been studied. All systems turned out to be considerably sensitive to truncation and finite size effects.[43,46,51] In the case of ionic and polar solutes correction methods based on continuum electrostatics[25,27,28,31] have been applied successfully to account for and correct the electrostatic free energy component. Ewald simulation leads to accurate results for the electrostatic (charging)

free energy of hydration of atomic ions with very small simulation cells as long as a suitable self-term is included in the calculation.[49,51] The question of finite size effects and influences of solute-solvent as well as solvent-solvent truncation on electrostatic free energies of solvated ions has quite recently been addressed by Hünenberger and co-workers[53] using dielectric continuum theory.

Less attention has been paid so far to apolar contributions of the free energy.[42,43] Sometimes simple empirical or experimentally deduced terms are added to account for the solvation cavity. In highly charged systems like solvated ions these apolar contributions are rather negligible. This also holds for mutations that preserve the solvation cavity. But whenever the simulation of absolute hydration free energies of large polar or apolar solutes is intended the contributions of the cavity require a closer look. The problem of how to correct for artifacts coming from application of *shifted-force* modifications to the truncated potential and their effect on the free energy is equally unclear; Brunsteiner and Boresch[46] provide the most recent and accurate material on electrostatic chemical potentials of water subject to different shifting and truncation schemes. It is clear that an additional error component is introduced that cannot be compensated for by theories covering the effect of mere truncation. In practice, simulations typically involve shifted-force modifications in order to avoid rapid heating due to sharp truncation. Inadequate treatment of simple truncation in dynamical simulations is responsible for inconsistencies between MC and MD results based on the same Hamiltonian, as has been pointed out by Trokhymchuk and Alejandre.[54]

An alternative description of solvation thermodynamics with a statistical yet atomistically resolved solvent model is provided by statistical-mechanical integral equation theory,[55-68] dominated by the reference interation site model (RISM) theory in its various approximation forms. From the solution to a set of nonlinear equations the RISM theory directly yields radial distributions functions of the system under consideration. Other observables like the free energy of solvation, i.e. the excess chemical potential at infinite dilution can be easily deduced under certain circumstances.[64] Compared to the time-consuming simulation methods the numerical

solution to RISM theories can be performed very quickly for moderately sized molecules, the results are free from statistical noise and apply directly to the thermodynamic limit. For absolute quantities however the current level of approximation of for instance liquid water phases within integral equation theory results in incorrect solvent structure and absolute solvation free energy errors of sometimes large magnitude.

By recognizing the fact that RISM theory in its simplest form, i.e. within the so-called hypernetted chain (HNC) approximation (extended RISM, RISM-HNC), when combined with molecular simulations yields almost quantitatively correct results whenever error cancellation to some degree can be expected, such hybrid framework has been applied to the computation of relative selectivities for ion/crown ether complexes,[65] the correction of radial distribution (*g*) functions from truncated simulations[66,67] and the direct determination of free energies from simulating a single state.[64,67] The general approach has been outlined and summarized recently.[68] In particular, the successful correction of *g* functions (besides the direct approach used in reference 67 that is related to earlier work by Trokhymchuk *et al.*[69] and by Kambayashi and Chihara[70]) was conceptually based on the assumption that the *difference* of chemical potentials of relatively similar systems is the same when computed either from (exact) simulations or from (approximate) RISM-HNC theory. This assumption can be deduced from perturbation-theoretic arguments based on the short-ranged character of the correction terms characterizing the difference between approximate and (hypothetically) exact integral equation solutions.[66,67] It appeared therefore in order to directly apply this methodology to the correction of free energy data from truncated potential simulation. By means of this methodology it should be possible to correct for free energy artifacts including and beyond those that are caught by electrostatic continuum theories. Another advantage of the integral equation approach is related to the fact that density variation effects due to application of a constant-pressure algorithm (that leads to different average densities for the truncated and the Ewald case) can also be considered. Such a strategy has

initially been suggested by Wood[48] quite some time ago; but no detailed formalism with that goal in mind has been described yet.

In this work the focus is set on model systems where *both* the solute-solvent as well as the solvent-solvent potentials are truncated and/or modified by a shifted-force technique. In this case maximal computational efficiency can be expected, particularly when the solvent model is simplified by drastic truncation as is possible by utilizing the enhanced damped Coulomb potential.[66,71] Various strategies are investigated, simulation-based as well as integral equation approaches based on the (truncated) reference trajectory only, to correct for these artifacts in a quantitative manner in order to reach the infinite-potential limit. The total correction can be split up into separate solute-solvent and solvent-solvent as well as long-range contribution, yielding some insight into the origin and magnitude of the correction. Argon in TIP3P water has been chosen as model system representing the purely apolar case and pure TIP3P water as an example for a polar system where the electrostatic free energy is severely contaminated by truncation artifacts. The models are evaluated by TI/MD simulations taking the corrections into account. Ultimately, integral equation theory even in the simple RISM-HNC approximation is suggested as a general and non-empirical strategy for correcting free energy simulation artifacts coming from quite a broad range of sources. Even the remaining approximation besides the finite simulation cell typically applied in Ewald simulations can be quantitatively elaborated: the truncation of short-ranged dispersive interactions. Furthermore, indications for the presence and magnitude of finite size effects in shifted-force potential simulations are discussed.

4.3 Methods

4.3.1 Free energy simulations

Thermodynamic integration and thermodynamic perturbation[8,9] are based upon the definition of a reversible path composed of intermediate states, linking the initial and final states of the system. In most applications the path is constructed by introducing a coupling parameter λ into the system Hamiltonian H where $H(\lambda = 0)$

represents the initial (*i*) state and $H(\lambda = 1)$ the final (*f*) state. Since the free energy is a state function, it is independent of the (reversible) path chosen. The corresponding change of solvation free energy, the excess chemical potential $\Delta\mu$, can be evaluated either by the free energy perturbation formula,[8,9]

$$\Delta\mu_{i\to f} = -\beta^{-1}\sum_{\lambda_j} \ln\left\langle \exp\{-\beta[H(\lambda_{j+1}) - H(\lambda_j)]\}\right\rangle_{\lambda_j}, \tag{4.1}$$

by summing the contributions of a sufficient amount of intermediate steps *j*, where $\beta = 1/kT$ is an inverse temperature with the Boltzmann constant *k*, absolute temperature *T*, and $\langle ... \rangle_\lambda$ denoting an ensemble average performed with the Hamiltonian $H(\lambda)$. The accuracy of the FEP method relies upon the capability of each reference ensemble at λ_j to sample also the next state λ_{j+1} adequately.[10]

On the other hand, by thermodynamic integration,[8,9]

$$\Delta\mu_{i\to f} = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial\lambda} \right\rangle_\lambda d\lambda, \tag{4.2}$$

the averages of $\partial H/\partial\lambda$ are sampled at intermediate states λ and integrated numerically. The accuracy of the numerical integration depends on the smoothness of the function in coupling parameter space. Assuming independent statistical error for each ensemble average the total error can be estimated reliably.[19]

In order to calculate the absolute solvation free energy of a solute in a solvent, the initial state is typically represented by the pure solvent with a non-interacting solute whereas the final state is the solute immersed in the solvent with full interaction potentials. Modeling the solute molecule as rigid body allows for a reduction of the solute-solvent interaction to a sum of Coulomb and Lennard-Jones terms within typical empirical force fields. The coupling parameter λ then defines a path along which the solute-solvent interaction potentials increases from zero to the full potential *V*.

A specific λ path is chosen only by consideration of computational efficiency in order to ensure numerically stable simulations.[39] To scale single charges or whole Coulomb terms V_C the simple linear scaling theme[34,36] is well established:

$$V_C(\lambda) = \lambda\, V_C = \lambda(4\pi\varepsilon_0)^{-1} q_i q_j / r \tag{4.3}$$

where $q_{i,j}$ denote partial site charges, r is the distance, and ε_0 is vacuum dielectric permittivity. If apolar Lennard-Jones interaction potentials V_{LJ} are involved, a nonlinear scaling scheme like[35,36]

$$V_{\mathrm{LJ}}(\lambda) = \lambda^n V_{\mathrm{LJ}} = 4\lambda^n\varepsilon[(\sigma/r)^{12} - (\sigma/r)^6], \tag{4.4}$$

with $n > 3$ is necessary to achieve adequate free energy results. Here, ε and σ are the Lennard-Jones parameters. In previous test simulations we have found the best performance with $n = 6$, i.e. λ^6 scaling. Similar power-law scaling schemes have been used in the earliest free energy simulation of liquid water.[72,73] Scaling of the total intermolecular interaction potentials comprising both Coulomb and Lennard-Jones terms can also be handled with the soft-core scaling method by Beutler *et al.*[37,38] Here we use it only for the Lennard-Jones part in the form

$$V_{\mathrm{LJ}}(\lambda) = 4\lambda^n\varepsilon\{(\alpha(1-\lambda)^2 + (r/\sigma)^6)^{-2} - (\alpha(1-\lambda)^2 + (r/\sigma)^6)^{-1}\} \tag{4.5}$$

where α and n are constants that define the softness of the core potential.

To analyze the influence of potential truncation (tr) and shifting (sh) on free energy values we use various cutoff distances r_c with typical CHARMM shifting functions for Coulomb and Lennard-Jones terms:[20]

$$V_{\mathrm{C,sh}} = V_{\mathrm{C}}(1-(r/r_c)^2)^2 , \tag{4.6}$$

$$V_{\mathrm{LJ,sh}} = V_{LJ} + Cr^6 + D \tag{4.7a}$$

with

$$C = 8\varepsilon\sigma^{12}/r_c^{18} - 4\varepsilon\sigma^6/r_c^{12}, \; D = -12\varepsilon\sigma^{12}/r_c^{12} + 8\varepsilon\sigma^6/r_c^6 . \tag{4.7b}$$

The λ derivatives of the CHARMM shifted Coulomb potential (4.6) with linear scaling (4.3) and shifted Lennard-Jones potential (4.7) with λ^6 (4.4) or soft-core scaling (4.5) are easily derived.

Instead of spherical truncation the Ewald summation technique[22,23] can be applied to describe the long-ranged Coulomb interaction. Ewald-quality free energy simulations can be regarded as full potential reference states. The total Ewald sum E_{tot} consists of a reciprocal space term E_k, the real space term E_r, the surface-dipole term E_{M} and the self-interaction term E_{S}:[22]

$$E_{\mathrm{tot}} = E_k + E_r + E_{\mathrm{M}} + E_{\mathrm{S}} . \tag{4.8}$$

When scaling the solute charges with the linear method (4.3), all terms of the Ewald sum become functions of λ. The real space and self-interaction terms have the form of simple pair potential sums. Their coupling parameter derivatives can be directly calculated. To cope with reciprocal and dipole terms a simple but effective trick is used: Both terms together are calculated at the actual λ value, $E_{k+\mathrm{M}}(\lambda)$, and simultaneously with the same configuration for slightly scaled solute charges, $E_{k+\mathrm{M}}(\lambda+\Delta\lambda)$. The partial derivative with respect to λ can be written as limit of the difference quotient

$$\left(\frac{\partial E_{k+\mathrm{M}}}{\partial\lambda}\right)_{\lambda} = \lim_{\Delta\lambda\to 0}\frac{E_{k+\mathrm{M}}(\lambda+\Delta\lambda)-E_{k+\mathrm{M}}(\lambda)}{\Delta\lambda}. \tag{4.9}$$

Thus, if Δλ is chosen small enough the coupling parameter derivative of total reciprocal and dipole Ewald terms can be approached by calculating the difference of two slightly varying sums. The eq. (4.9) was implemented within the conventional Ewald sum[23] and within the smooth particle mesh (SPME) version.[22] Since one can use the same program loops of the Ewald algorithm for both sums in parallel, the method is quite efficient.

An interesting quantity guiding further analysis of free energy simulation techniques is given by the spatial solvation free energy density. In the case of pairwise additive potentials the integrand can be rearranged in the TI expression as

$$\Delta\mu = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial\lambda}\right\rangle d\lambda = \sum_{\alpha,\gamma}\int_0^1 4\pi\rho_\gamma \int_0^\infty \frac{\partial V_{\alpha\gamma}(\lambda,r)}{\partial\lambda} g_{\alpha\gamma}(r;\lambda) r^2 d\lambda dr = 4\pi\int_0^\infty dr\, r^2 \rho_\mu(r) \tag{4.10}$$

such that the radially averaged free energy density is given as

$$\rho_\mu(r) = \sum_{\alpha,\gamma}\rho_\gamma \int_0^1 d\lambda \frac{\partial V_{\alpha\gamma}(\lambda,r)}{\partial\lambda} g_{\alpha\gamma}(r;\lambda). \tag{4.11}$$

where, here, α and γ denote solute and solvent sites, respectively, and ρ_γ is the solvent site density. This form is particularly easily accessible from RISM-HNC theory. Although implying high computational cost, the free energy density can also be derived from simulations by recording radial distribution functions for each coupling parameter value and integration.

4.3.2 Simulation-based truncation correction

Because of the rigid body approximation valid for the model systems, the configurational energy of the system consists additively only of intermolecular Coulomb- and Lennard-Jones interactions. Since the solute-solvent (uv) as well as the solvent-solvent (vv) interactions are truncated or shifted, separate uv and vv contributions to the total correction $\Delta\Delta\mu$ can be expected. For the various short and long range terms the application of different correction methods is appropriate.

a) Long range correction: If it is assumed that the solvent already reaches the bulk phase density at the cutoff distance, and the solute-solvent radial distribution function approaches unity, then the long range correction (LRC) of the truncation error of solute-solvent Lennard-Jones potentials can be estimated by integration of the *g* functions starting at the cutoff to infinity[42,43] as

$$\Delta\Delta\mu_{\mathrm{MD,LRC,uv}} = 4\pi \sum_{\alpha\gamma} \rho_{\gamma} \int_{r_c}^{\infty} dr\, r^2\, V_{\mathrm{LJ}}(r). \tag{4.12}$$

For electrostatic contributions beyond the cutoff distance, r_c, in the case of polar solutes and solvents used with truncated (not shifted!) potentials the general reaction field (GRF) continuum electrostatics correction by Resat and McCammon[25] is applicable and yields good results. One should however keep in mind that the contribution of the shifted potential region in the case of shifted-force potentials is not negligible and has not been assessed in earlier work.

b) Free energy perturbation correction: The free energy perturbation formula applied as a single step can also be used as a truncation correction method. If difference between a shifted and truncated potential is definded as $\Delta V = V_{\mathrm{tr}} - V_{\mathrm{sh}}$ then the FEP method yields the corresponding free energy difference as

$$\Delta\Delta\mu_{\mathrm{MD,FEP,uv/vv}} = -\beta^{-1} \ln\left\langle \exp(-\beta\Delta V) \right\rangle_{\mathrm{sh}}, \tag{4.13}$$

i.e., by an average of the shifted potential ensemble. The correction, in principle applicable to the solute-solvent and the solvent-solvent case, leads to the truncated system to be complemented by a long range term at least for the solute-solvent terms. In principle the FEP correction could also be used for the long range part, but in practice potential values for distances larger than the cutoff are not sampled. Because

of known numerical restrictions of the FEP method only small steps ΔV lead to converged results.[10]

c) Umbrella sampling correction: The truncation effect of using a shifting function with respect to the truncated case can also be estimated by a method derived from umbrella sampling techniques.[11] If the difference of shifted and truncated pair potential is defined as biasing potential $\Delta V = V_{\mathrm{sh}} - V_{\mathrm{tr}}$ in the sense of an umbrella method (notice the sign reversal as compared to the last paragraph) then the classical umbrella formula provides the calculation of the unbiased (unshifted) observables. Together with the TI formula (4.3) the corrected free energy follows from coupling parameter integration of

$$\left(\frac{\partial \mu}{\partial \lambda}\right)_{\mathrm{corrected}} = \frac{\left\langle \frac{\partial V}{\partial \lambda} \exp\left(+\beta \Delta V\right) \right\rangle_{\mathrm{sh}}}{\left\langle \exp\left(+\beta \Delta V\right) \right\rangle_{\mathrm{sh}}}. \tag{4.14}$$

Similarly to the FEP method the calculation of unknown ensemble averages with a known reference ensemble is limited to small ΔV values. In principle the formula is not restricted to a correction in the sense of shifted to truncated potential, the full potential is also possible. In practice however the solvent shell size entering the expression does not exceed the cutoff distance.

The total potential difference ΔV can be split into contributions of solute-solvent interaction ΔV_{uv}, solvent-solvent interaction ΔV_{vv} within a solvation shell around the solute of a given radius R and the solvent-solvent potential outside the shell

$$\Delta V = \Delta V_{\mathrm{uv}} + \Delta V_{\mathrm{vv}}(r < R) + \Delta V_{\mathrm{vv}}(r > R). \tag{4.15}$$

The solute-solvent corrected chemical potential then has the form

$$\Delta \mu_{\mathrm{MD,UMB,uv}} = \int_0^1 \frac{\left\langle \frac{\partial V}{\partial \lambda} \exp\left(+\beta \Delta V_{\mathrm{uv}}\right) \right\rangle_{\mathrm{sh}}}{\left\langle \exp\left(+\beta \Delta V_{\mathrm{uv}}\right) \right\rangle_{\mathrm{sh}}} d\lambda \tag{4.16}$$

and within the solvent shell one has

$$\Delta \mu_{\mathrm{MD,UMB,vv}} = \int_0^1 \frac{\left\langle \frac{\partial V}{\partial \lambda} \exp\left(+\beta \Delta V_{\mathrm{vv}}(r < R)\right) \right\rangle_{\mathrm{sh}}}{\left\langle \exp\left(+\beta \Delta V_{\mathrm{vv}}(r < R)\right) \right\rangle_{\mathrm{sh}}} d\lambda. \tag{4.17}$$

In order to use the UMB method, ΔV values have to be calculated and stored during the simulation protocol. In a second pass the simulated average is rescaled using eqs. (4.16) or (4.17). The umbrella correction is finally given by the difference of corrected and shifted solvation free energy:

$$\Delta\Delta\mu_{\mathrm{MD,UMB}} = \Delta\mu_{\mathrm{MD,UMB}} - \Delta\mu_{\mathrm{MD,sh}} \,. \tag{4.18}$$

4.3.3 Integral equation theory

The one-dimensional RISM integral equation for fluids composed of molecules with sites α and γ as used in this work is given by[55]

$$\mathbf{h}^{\mathrm{vv}} = \boldsymbol{\omega}^{\mathrm{vv}} * \mathbf{c}^{\mathrm{vv}} * \boldsymbol{\omega}^{\mathrm{vv}} + \boldsymbol{\omega}^{\mathrm{vv}} * \mathbf{c}^{\mathrm{vv}} * (\rho\mathbf{h}^{\mathrm{vv}}) \tag{4.19}$$

where $\mathbf{h} = (h_{\alpha\gamma}(r))$ is the matrix of the total correlation functions (the radial distribution function is $g = h+1$), $\mathbf{c} = (c_{\alpha\gamma}(r))$ are the direct correlation functions, $\boldsymbol{\omega} = (\omega_{\alpha\gamma}(r))$ are the intramolecular correlation functions (for rigid molecules these are normalized Dirac delta functions constraining the site distance), and the star denotes convolution. For a solute system in infinite dilution the simpler relation

$$\rho\mathbf{h}^{\mathrm{uv}} = \boldsymbol{\omega}^{\mathrm{uu}} * \mathbf{c}^{\mathrm{uv}} * \boldsymbol{\chi}^{\mathrm{vv}} \tag{4.20}$$

holds where

$$\boldsymbol{\chi}^{\mathrm{vv}} = \rho\boldsymbol{\omega}^{\mathrm{vv}} + \rho^2\mathbf{h}^{\mathrm{vv}} \,, \tag{4.21}$$

the solvent susceptibility (density correlation function), can be computed beforehand and independently of any solute. These equations must be supplied with a so-called "closure" relation,

$$h_{\alpha\gamma}(r) + 1 = \exp(-\beta V_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r) + B_{\alpha\gamma}(r)) \tag{4.22}$$

where $B_{\alpha\gamma}$ are (at least formally in the site-averaged molecular theory) the bridge functions. In the HNC approximation (extended RISM) B identically vanishes.

The particular advantage of using a set of equations like (4.20)-(4.22) for computing radial distribution functions is due to the fact that the excess chemical potential can be computed in a path-independent closed form for the HNC closure from[57]

$$\Delta\mu_{\mathrm{RISM}} = 4\pi\rho\beta^{-1}\int_0^\infty dr\, r^2\left(\frac{1}{2}h_{\alpha\gamma} - c_{\alpha\gamma} - \frac{1}{2}h_{\alpha\gamma}c_{\alpha\gamma}\right). \tag{4.23}$$

A radially averaged free energy density corresponding to eq. (4.11) is directly given by the integrand in eq. (4.23) divided by $4\pi r^2$. For other closures with nonzero bridge functions, path independence can be achieved while not necessarily a closed form expression is obtained.[64] The RISM-HNC equations have been used earlier for interpreting[58,59] and correcting[66,67] truncation errors in molecular simulations on the level of g functions and internal energy, whereas a direct free energy correction has not been attempted until now.

4.3.4 Correction with RISM-HNC theory

A first order functional perturbation expanson of the chemical potential in terms of Mayer functions of the bridge term, as discussed in the corresponding references[61,66,67], leads to

$$\Delta\mu_{\mathrm{MD}} \approx \Delta\mu_{\mathrm{RISM}}[\mathbf{h}_{\mathrm{RISM}}, \mathbf{c}_{\mathrm{RISM}}] + 4\pi\rho\beta^{-1}\sum_{\alpha\gamma}\int_0^\infty dr\, r^2\, g_{\mathrm{RISM},\alpha\gamma}(r)[\exp(B_{\alpha\gamma}(r)) - 1]. \tag{4.24}$$

Since the dominant truncation effects on g are located around the cutoff distance where the bridge term is small, the second term approximately cancels upon subtracting expressions for full and truncated potential, yielding

$$\Delta\Delta\mu_{\mathrm{MD}} = \Delta\mu_{\mathrm{MD}}(r_c \to \infty) - \Delta\mu_{\mathrm{MD}}(r_c) \approx \Delta\Delta\mu_{\mathrm{RISM}} = \Delta\mu_{\mathrm{RISM}}(r_c \to \infty) - \Delta\mu_{\mathrm{RISM}}(r_c). \tag{4.25}$$

A similar relation holds for correcting from one cutoff distance to another. Given the successful structural and energetic corrections in chapter 3, now the accuracy of eq. (4.25) can be explicitly studied in comparison with the simulation-based correction methods and full-potential reference simulations.

From this basic scheme various expressions covering solute-solvent, solvent-solvent, and long range contributions to the total correction in correspondence to the simulation corrections defined in the last subsection can be derived: The total correction is formally given by

$$\Delta\Delta\mu_{\mathrm{RISM,tot}} = \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{full}}, \mathrm{vv}_{\mathrm{full}}, \rho_{\mathrm{EW}}) - \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{sh}}, \mathrm{vv}_{\mathrm{sh}}, \rho_{\mathrm{sh}}) \approx \Delta\Delta\mu_{\mathrm{MD,tot}} \tag{4.26}$$

where $[uv,vv]_{full,sh}$ denote RISM-HNC solutions for the [solute-solvent, solvent-solvent] potential with the full or shifted potentials, respectively, and for the *average* solvent density taken from pure-solvent isothermal-isobaric simulations under Ewald (EW) and shifted-force (sh) conditions, or for the prescribed density used in isothermal runs. Here a subdivision of the total correction into separate terms is expected, according to

$$\Delta\Delta\mu_{\mathrm{RISM,tot}} = \Delta\mu_{\mathrm{RISM,uv}} + \Delta\mu_{\mathrm{RISM,vv}} + \Delta\mu_{\mathrm{RISM,LRC}} \,. \tag{4.27}$$

The individual terms are

$$\Delta\Delta\mu_{\mathrm{RISM,uv}} = \Delta\mu_{\mathrm{RISM}}(0 \to r_c, \mathrm{uv}_{\mathrm{full}}, \mathrm{vv}_{\mathrm{sh}}, \rho_{\mathrm{sh}}) - \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{sh}}, \mathrm{vv}_{\mathrm{sh}}, \rho_{\mathrm{sh}}) \,, \tag{4.28}$$

where the additional argument, $0 \to r_c$, means that the upper integration limit in eq. (4.23) is replaced by the cutoff distance,

$$\Delta\Delta\mu_{\mathrm{RISM,vv}} = \Delta\mu_{\mathrm{RISM}}(0 \to r_c, \mathrm{uv}_{\mathrm{full}}, \mathrm{vv}_{\mathrm{full}}, \rho_{\mathrm{EW}}) - \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{sh}}, \mathrm{vv}_{\mathrm{sh}}, \rho_{\mathrm{sh}}) - \Delta\Delta\mu_{\mathrm{RISM,uv}} \tag{4.29}$$

and finally

$$\Delta\Delta\mu_{\mathrm{RISM,LRC,uv}} = \Delta\mu_{\mathrm{RISM}}(r_c \to \infty, \mathrm{uv}_{\mathrm{full}}, \mathrm{vv}_{\mathrm{full}}, \rho_{\mathrm{EW}}) \,. \tag{4.30}$$

In the last expression, eq. (4.24) is integrated from the cutoff distance up to infinity; it is assumed that the long range RISM structure corresponds to the simulation distribution. The total correction, eq. (4.26) does not necessarily correspond to the sum of individual terms, eq. (4.27) together with (4.28)-(4.30). For the model systems described below however the difference never exceeds the second decimal figure on the kcal/mol scale, further supporting the subdivision suggested above.

A straightforward extension of the above scheme for the correction of purely electrostatic free energy components that is also explored below can be formulated as

$$\begin{aligned}\Delta\Delta\mu_{\mathrm{RISM,elec}} = &(\Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{full}}, \mathrm{vv}_{\mathrm{full}}, \rho_{\mathrm{EW}}) - \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{LJ,full}}, \mathrm{vv}_{\mathrm{full}}, \rho_{\mathrm{EW}})) - \\ &(\Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{sh}}, \mathrm{vv}_{\mathrm{sh}}, \rho_{\mathrm{sh}}) - \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}_{\mathrm{sh,LJ}}, \mathrm{vv}_{\mathrm{sh}}, \rho_{\mathrm{sh}}))\end{aligned} \tag{4.31}$$

where the subscript LJ,full denotes a computation with the full Lennard-Jones potential only. Note that the path independence of the HNC chemical potential functional has been explicitly exploited in order to construct the electrostatic component relation although any component is certainly path-dependent since the apolar cavitiy has to exist *before* charging. Despite the validity of this general

relation, the electrostatic free energy data from simulations has been corrected only in the canonical ensemble, such that the two densities in eq. (4.31) are equal.

The magnitude of real-space Lennard-Jones truncation typically used in Ewald simulation schemes can also be quantified now: In variation of eq. (4.26) it is

$$\Delta\Delta\mu_{\mathrm{RISM,vv}}^{\mathrm{EW-LRC}} = \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}, \mathrm{vv}_{\mathrm{full}}, \rho_{\mathrm{EW}}) - \Delta\mu_{\mathrm{RISM}}(\mathrm{uv}, \mathrm{vv}_{\mathrm{tr}}, \rho_{\mathrm{EW}}) \tag{4.32}$$

such that the true full potential limit is obtained without the need of Ewald-analogue summation of Lennard-Jones potentials.

4.4 Computational details

Two model cases were considered for the analysis: argon solvated in water for studying apolar system,[40-43] and pure water as an example for polar systems dominated by electrostatic interactions.[44-46] The argon / water case is particularly interesting since the effect of potential shifting and truncation of both, the solute-solvent and the solvent-solvent interaction on the free energy of hydration has not been conclusively treated and assessed in the past. The general approach can be summarized as follows: After simulations with varying cutoff conditions that yield the basic value for all corrective strategies, the simulation-based and RISM-based truncation correction methods were applied to individual (uv, vv, LRC) components and summed up to give a total correction, to be compared with reference data from Ewald simulations.

4.4.1 Free energy simulations

The simulations were performed with the DL_POLY[74] code, extended by the free energy computation functionality. Additionally to the positions of all atomic centers, information about potential differences ΔV and free energy contributions were stored in the trajectory file. Throughout this chapter, water is described by the rigid point charge TIP3P model, extended by small Lennard-Jones terms originally included to allow for absolute free energy simulations that are also necessary for RISM computations.[75,76] The argon Lennard-Jones parameters were taken from

Guenot and Kollman.[77] For the Lennard-Jones interactions the Lorentz-Berthelot combination rules were applied.[19] All force field parameters are listed in Table 4.1.

Table 4.1: Force field parameters of extended TIP3P water[75,76] and argon.[77]

atom type	q / e	ε / kcal mol^{-1}	σ / Å
O_{water}	-0.834	0.152	3.15
H_{water}	+0.417	0.046	0.4
Ar	0.0	0.253	3.414

For the argon / water case, two simulation cells were constructed, one with 1 argon and 1000 water molecules, a second larger system was composed of 1 argon and 2000 water molecules. MD simulations were performed under isothermal-isobaric (*NpT*) conditions at a temperature of 300 K and a pressure of 1 bar in rectangular simulation cells with periodic boundary conditions. The second system under investigation, pure water, was examined with respect to the electrostatic chemical potential. Again, two cubic boxes composed of 1000 and 2000 molecules were constructed, simulation conditions were chosen as isothermal (*NVT*) at 300 K at the experimental density of 1.0 g cm^{-3} corresponding to a number density of 0.033456 Å^{-3}. An ensemble-optimized variant of the constant temperature algorithm by Kast *et al.*[78,79] was used for *NVT* simulations, combined with the Berendsen barostat[80] in the *NpT* cases. The coupling parameter for the constant temperature algorithm was 0.001, the barostat relaxation time was 2 ps. The rigid molecules were constrained by using the SHAKE algorithm.[81] All simulations were performed with a time step of 2 fs using the Verlet leapfrog algorithm.[19]

To perform the thermodynamic integration the argon-water interaction potential was scaled with the soft-core method (4.5) ($\alpha = 0.5$ and $n = 2$) and with the λ^6 scheme (4.4). 21 independent simulations at λ values of 0.0, 0.05, 0.1,…, 1.0 were produced with 50 ps of equilibration and 250 ps of production simulations for each coupling parameter value, totaling around 5 ns simulation time per free energy computation. The trajectory data was stored every 50 time steps (100 fs). The free

energy of hydration of argon was calculated by numerical integration of eq. (4.2) using the trapezoidal rule. Statistical error was separately determined for each λ value by means of the blocking technique[19] and accumulated to give the total free energy error. With the CHARMM shifted-force functions, eqs. (4.6) and (4.7), the hydration free energy of argon was computed for the cutoff distances 8, 10 and 12 Å with the small argon-water box and for 16 Å with the large box. Additionally we performed free energy simulations with Ewald summation of the water Coulomb interactions based on the small simulation box using a real-space cutoff of 12 Å (note that in the DL_POLY implementation Lennard-Jones terms used with the Ewald sum are simply truncated at the cutoff, not shifted), a broadening parameter of 0.3, and a reciprocal space dimension of 10 at conducting boundary conditions. To examine finite size effects the simulations with spherical cutoff at 12 Å and using the Ewald sum were also repeated with 2000 water molecules and a reciprocal dimension of 12. For SPME simulations grid sizes 8^3 and 16^3 were used for small and large box, respectively.

In the pure water case, the electrostatic free energy was determined from TI simulations where the charges of one water molecule were scaled to zero with the linear method of eq. (4.3) while maintaining the value of the Lennard-Jones terms. The charge scaling is generally associated with smoother functions in λ space. Therefore a more coarsely spaced set of points λ = 0.0, 0.1, 0.2, … , 1.0 was sufficient. All λ points were independently equilibrated for again 50 ps and sampled for 250 ps. The electrostatic hydration free energy of water was computed with CHARMM shifted-force cutoffs at 8, 10 and 12 Å, and with the Ewald method using the small box. Again, to check finite size effects, the 8 Å cutoff and Ewald free energy simulations were repeated with 2000 water molecules.

Corrective strategies as described above to be compared with reference data were applied in the following form: In the apolar case the solute-solvent correction turned out to be computed accurately enough by the free energy perturbation technique only, eq. (4.13), whereas the solvent-solvent component was advantageously derived from the umbrella technique (4.17). These terms were

complemented by the long range correction, eq. (4.12). These terms were then recomputed using the RISM strategy and compared. For the second model system, the pure water case, only the RISM method is applied to calculate the total correction and compare the results with reference data. The Coulombic long range truncation error was also estimated by using the GRF technique.[25]

4.4.2 RISM-HNC computations

The solution to the set of integral equations (4.20)-(4.23) within the HNC approximation was done with precisely the same potential functions as used in the simulation, i.e. even including the effect of Lennard-Jones truncation in the real-space sum within Ewald simulations. The RISM solutions were obtained on a logarithmic grid of 512 points ranging from $5.98 \cdot 10^{-3}$ Å to a maximum distance of 164.02 Å at the same (average) density and temperature as in the MD simulations. For the cutoff distances 8, 10, 12, 16 Å, and for Ewald conditions the corresponding average *NpT* water densities used in eqs. (4.20) and (4.21) were 0.03343, 0.03340, 0.03316, 0.03312, and 0.03291 $Å^{-3}$, respectively. The temperature was set to 300 K as in all the simulations.

As discussed in Kast *et al.*[67] and in analogy with the *g* function correction in chapter 3 neither the dielectrically consistent RISM equations[82] nor the simple charge-scaling procedure[83] were used to adjust the dielectric constant at full potential for the pure-solvent calculations. The nonlinear system of equations was solved using a variant of the "modified direct inversion of iterative subspace" (MDIIS) method[84] for coarse optimization of the correlation functions and the traditional Picard iteration scheme for ultimate convergence up to a threshold of $\max(\Delta\mathbf{c}) < 10^{-6}$ for the pure solvent and 10^{-5} for the solute case between two successive iteration steps. Final integrations of the chemical potential, eq. (4.24), and its related correction forms were performed analytically from a cubic spline interpolant.

4.5 Results and discussion

4.5.1 Argon in water: simulation-based correction results

First the proper use of coupling parameter scaling was checked for apolar case. The soft-core scaling and λ^6 methods yielded practically the same free energy results but the total statistical error is somewhat smaller in the λ^6 scaling case. To explain this result, in Figure 4.1 the corresponding coupling parameter derivative averages are plotted against the λ path. Particularly around the steep maximum of the soft-core scaling curve quite long simulation times and a dense grid of λ points is needed to ensure accurate numerical integration. The λ^6 path is obviously smoother and more efficient for scaling a single Lennard-Jones center (not necessarily larger molecules). The major advantage of the soft-core method is realized when dealing with molecules or Lennard-Jones and Coulomb potentials simultaneously which is not possible with the λ^6 scheme. Here however the latter method is prefered to compute the basic values $\Delta\mu_{MD}$.

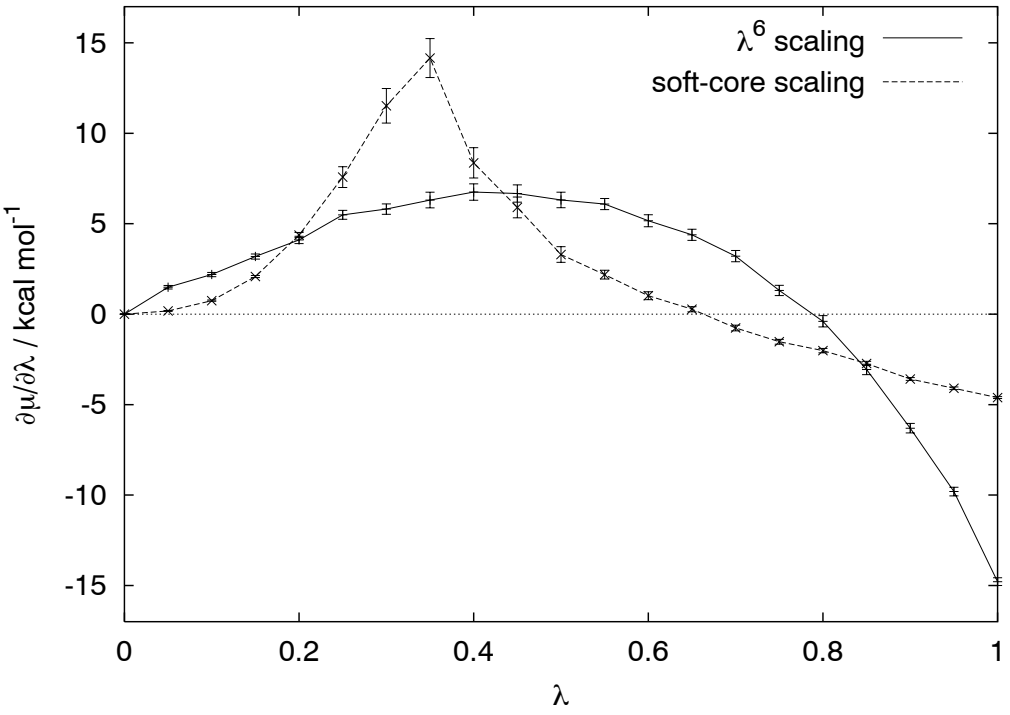

Figure 4.1: Comparison of the λ^6 (solid) and the soft-core scaling scheme (dashed) for the argon-water interaction potential with respect to the free energy derivative

Turning now to the various correction terms the shifted to truncated solute-solvent potential correction, $\Delta\Delta\mu_{MD,FEP,uv}$, could successfully be computed with a single step FEP evaluation according to eq. (4.13). The formal statistical errors of FEP type calculations are still debated in the literature,[10] so that here a similar order of magnitude for the error of the corrected values was assumed as was obtained for the original numbers. The long range correction $\Delta\Delta\mu_{MD,LRC,uv}$, was derived from eq. (4.12) after ensuring by inspection that the radial distribution functions are sufficiently close to unity near the cutoff distance. The solvent-solvent contribution was more difficult to compute. The umbrella technique, eq. (4.17), was successfully applied to yield the corrected value $\Delta\Delta\mu_{MD,UMB,vv}$ from a careful consideration of the size of the solvent shell to be taken into account: To find the suitable solvation shell radius the correction term was calculated for successively increased shell radius. As illustrated in Figure. 4.2, a smooth curve was obtained up to roughly the distance of the first distinct solvent shell at 4.8 Å beyond which the statistical error gets to large. It is clear that one should not have too much confidence in these values due to the fuzzy definition of the reference radius, however, the numbers obtained agree quite well with the RISM results.

All terms, the total accumulated correction, $\Delta\Delta\mu_{MD,tot}$, and the final, corrected value, $\Delta\mu_{MD,corr}$, are summarized in Table 4.2 for the cutoff distances 8, 10, 12, and 16 Å under consideration, together with the Ewald summation data. For the latter, the long range correction for 12 Å also applies and is calculated explicitly since the Lennard-Jones terms are affected by truncation. The results are further depicted in Figure 4.3 suggesting a successive accumulation in arbitrary ordering to the total sum. The simulated free energies show significant truncation effects but approach the Ewald limit with increasing cutoff distance. For example, the free energy at 8 Å cutoff shows some 50 % deviation as compared to the Ewald reference, which is quite an unexpected effect that has, in contrast to ionic systems,[49-52] been largely neglected for apolar systems in the literature. The magnitude of the shift correction is even more surprising if the small deviation between shifted and original potential is considered as illustrated in Figure 4.4.

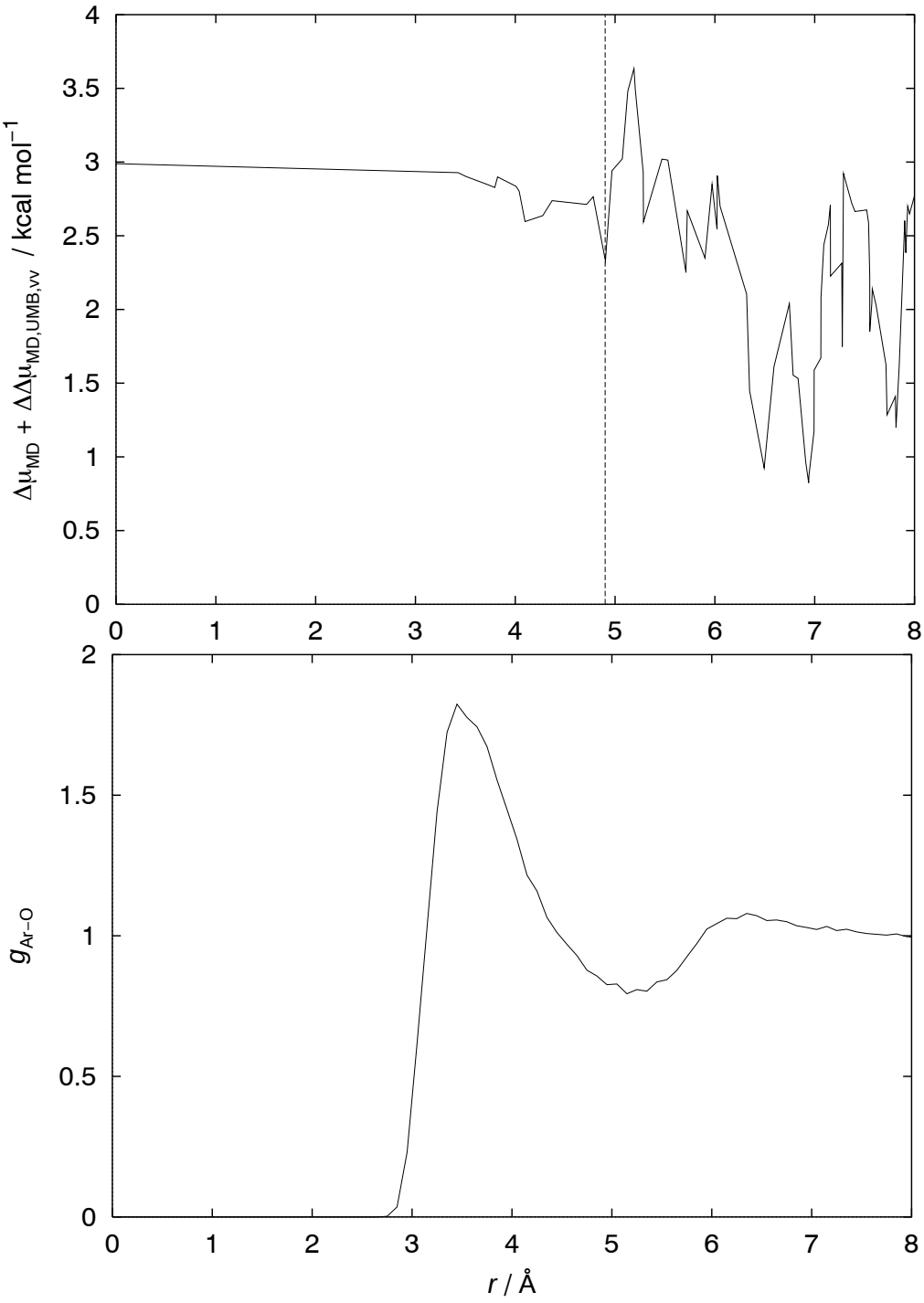

Figure 4.2: Solvent-solvent umbrella-corrected hydration free hydration of argon as function of the solvent shell radius (top) and argon-oxygen radial distribution function (bottom) for CHARMM shift-force potential at 8 Å cutoff distance.

Table 4.2: Free energy of argon in extended TIP3P water for different cutoff conditions and Ewald (EW) simulation results; solute-solvent, solvent-solvent, and long range free energy corrections (in kcal mol^{-1}) derived from simulation-based correction.

r_c / Å	8	10	12	16	EW
$\Delta\mu_{MD}$	2.99 ± 0.06	2.54 ± 0.05	2.24 ± 0.03	2.10 ± 0.03	2.03 ± 0.02
$\Delta\Delta\mu_{MD,FEP,uv}$	-0.43	-0.22	-0.13	-0.05	/
$\Delta\Delta\mu_{MD,LRC,uv}$	-0.26	-0.13	-0.076	-0.033	-0.076
$\Delta\Delta\mu_{MD,UMB,vv}$	-0.54	-0.43	-0.24	-0.06	/
$\Delta\Delta\mu_{MD,tot}$	-1.23	-0.78	-0.446	-0.143	-0.076
$\Delta\mu_{MD,corr}$	1.76	1.76	1.79	1.96	1.95

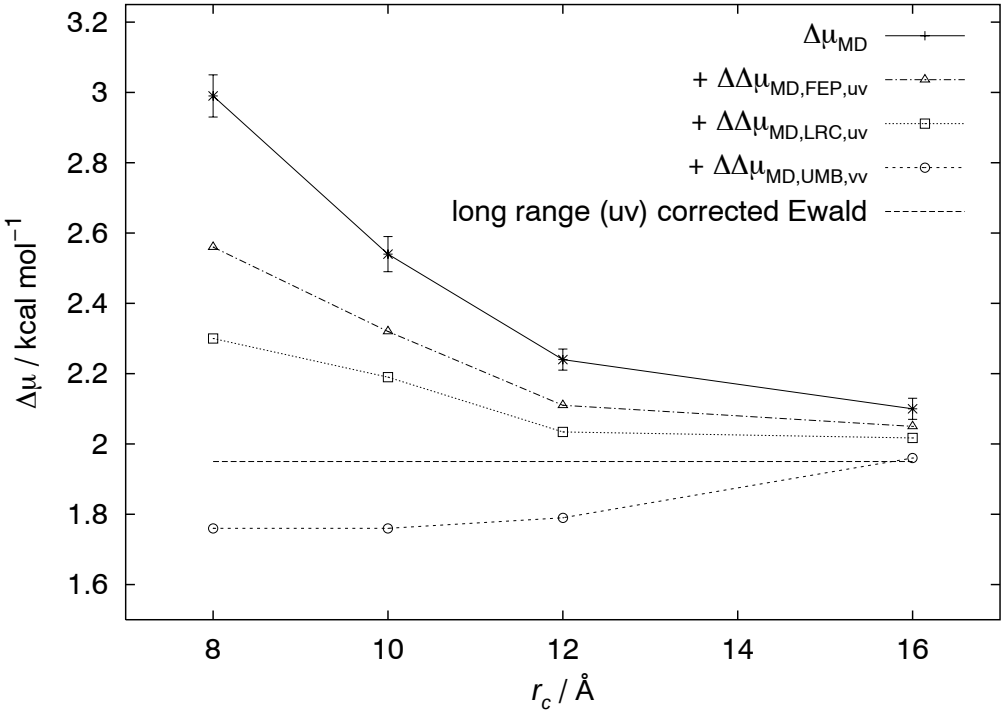

Figure 4.3: Successive accumulation of the total free energy from simulation-based correction for the argon-water case taking into account various cutoff distances and the Ewald reference value according to Table 4.2.

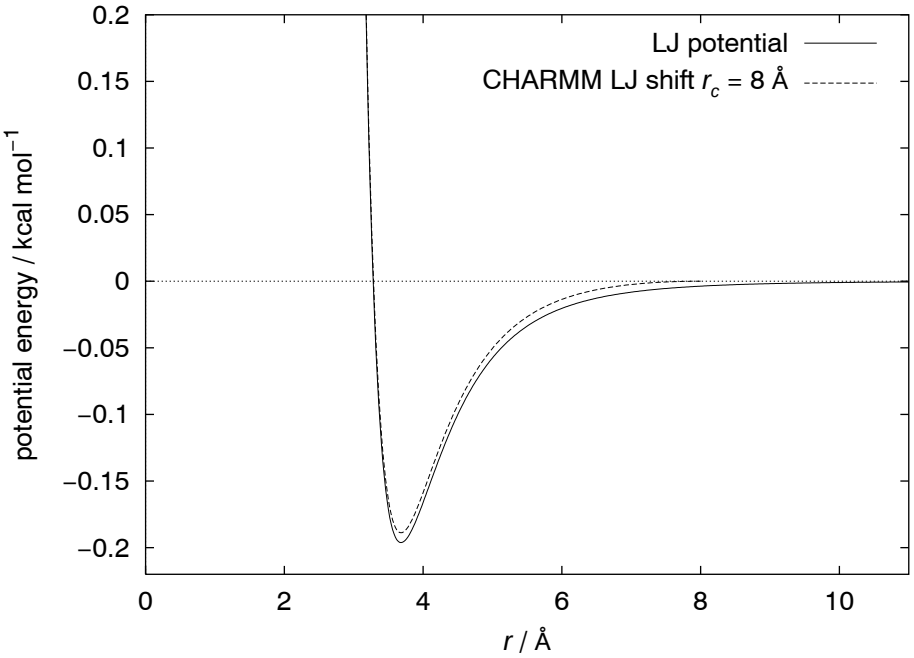

Figure 4.4: Lennard-Jones argon-oxygen potential (solid) and CHARMM shifted-force version with a cutoff distance of 8 Å (dashed).

Comparing the results of all the correction methods described at varying cutoff conditions in Table 4.2, the consequence of spherically truncating the solute-solvent potential, measured as $\Delta\Delta\mu_{MD,LRC,uv}$, is obviously the smallest truncation effect, whereas the influence of using a solute-solvent shifting function, as expressed by $\Delta\Delta\mu_{MD,FEP,uv}$, and the error induced by solvent-solvent shifting, $\Delta\Delta\mu_{MD,UMB,vv}$, amount to similar contributions. As expected all error contributions grow with smaller cutoff and diminish to zero with increasing cutoff range. In parallel to increasing cutoffs the uncorrected simulation results quite slowly also approach the Ewald limit. Even the corrected results agree only for the 16 Å simulation within error bars with the Ewald reference, for smaller cutoffs a consistent residual error of around 0.2 kcal mol^{-1} remains. This, while still acceptable, is most likely related to the inaccuracy of the solvent-solvent term and poses a limitation on the simulation-based correction techniques.

Another source of error is posed by the possible existence of finite size effects that should be less likely for the 16 Å simulation where the larger box has been used. Therefore the free energy simulations with CHARMM shifted-force cutoff of 12 Å

and the Ewald sum were performed again with 2000 water molecules. The results for the Ewald system were $\Delta\mu_{MD}(2000) = 2.037 \pm 0.02$ kcal mol^{-1} and $\Delta\mu_{MD}(1000) = 2.037 \pm 0.02$ kcal mol^{-1}, rendering a finite site effect unlikely, while for the CHARMM system $\Delta\mu_{MD}(2000) = 2.29 \pm 0.03$ kcal mol^{-1} and $\Delta\mu_{MD}(1000) = 2.24 \pm 0.03$ kcal mol^{-1}. A slight finite size effect around 0.05 kcal/mol for the CHARMM cutoff indicates a further contribution to the total error in Table 4.2.

A substantial dependence of absolute free energy values on the truncation scheme and the long range contributions is relevant for any kind of parametrization work where experimental free energy data is taken into account. The (hypothetical) full potential limit should be taken as a reference state in order to avoid the cutoff distance and the shifting scheme acting as another empirical parameter. Such an idea, although not considering free energy data, has already been adopted in very recent work on the reparametrization of the TIP3P water model[85] that has originally been designed for 9 Å truncation simulations. Particularly with the RISM correction described below it is now possible to correct to the true full potential even beyond present simulation limitations.

4.5.2 Argon in water: RISM-based correction results

Analogously to the simulation-corrections now the RISM results within the HNC approximation according to eqs. (4.27)-(4.30) are regarded. The results are summarized in Table 4.3, the syntax corresponds to that of the preceding subsection except for replacing the subscript "MD" by "RISM". The total correction can be independently calculated directly from eq. (4.26) and gives the same results. The final extrapolated value for the free energy of hydration, $\Delta\mu_{RISM,corr}$, is computed by adding the total correction to the uncorrected TI/MD results, $\Delta\Delta\mu_{MD}$. Again, the results are plotted in Figure 4.5 together with the simulation-based corrections from the last subsection for comparison. In both, the table and the figure, the long range solvent-solvent contribution from eq. (4.32) was added for correcting the Lennard-Jones truncation applied to the Ewald simulation. No simulation-based strategy known up till now could have provided this value.

Table 4.3: Free energy of argon in extended TIP3P water for different cutoff conditions and Ewald (EW) simulation results; solute-solvent, solvent-solvent, and long range free energy corrections (in kcal mol^{-1}) derived from RISM-based correction.

r_c / Å	8	10	12	16	EW
$\Delta\mu_{MD}$	2.99 ± 0.06	2.54 ± 0.05	2.24 ± 0.03	2.10 ± 0.03	2.03 ± 0.02
$\Delta\Delta\mu_{RISM,uv}$	-0.45	-0.24	-0.14	-0.06	/
$\Delta\Delta\mu_{RISM,LRC,uv}$	-0.27	-0.14	-0.082	-0.034	-0.076
$\Delta\Delta\mu_{RISM,vv}$	-0.44	-0.36	-0.16	-0.14	-0.076[a]
$\Delta\Delta\mu_{RISM,tot}$	-1.16	-0.74	-0.38	-0.234	-0.152
$\Delta\mu_{RISM,corr}$	1.83	1.80	1.87	1.87	1.88

[a]computed with eq. (4.32)

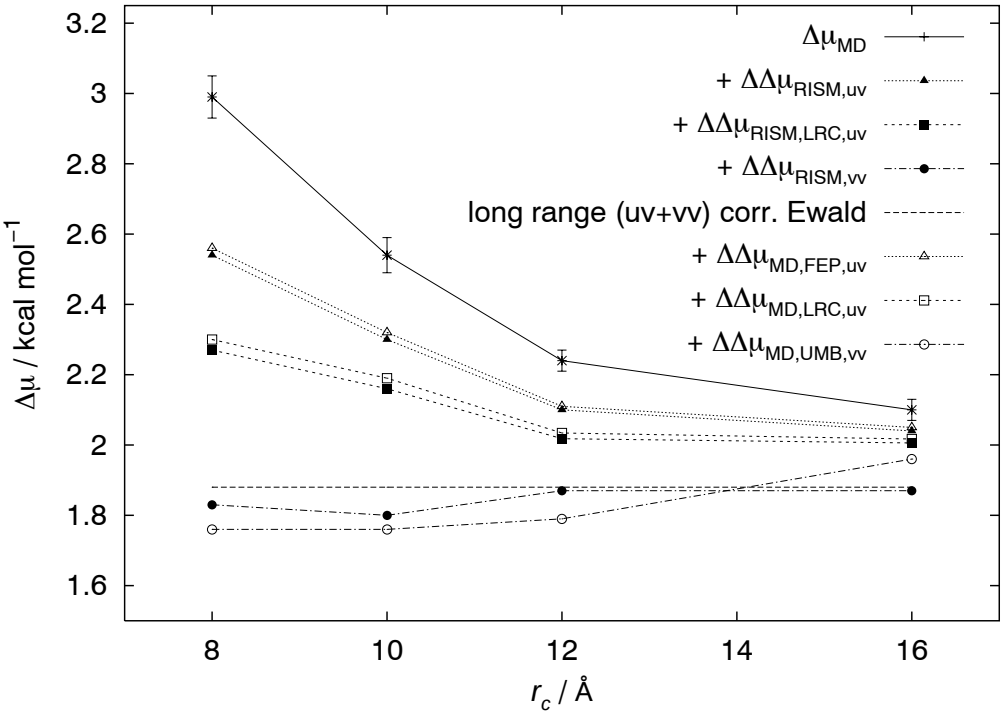

Figure 4.5: Successive accumulation of the total free energy from RISM-based correction (and for comparison purposes from simulation-based correction) for the argon-water case taking into account various cutoff distances and the Ewald reference value according to Table 4.3. Note that the simulation-based corrections given in the legend are to be added to the simulation result, $\Delta\mu_{MD}$.

Note the excellent extrapolation agreement for the various cutoffs to the reference value. Although quite close to the simulation-based results, the predictive correction capability of the RISM strategy is higher and the method is much more robust. The remaining statistically insignificant differences can be interpreted on the basis of the finite size effect discussed in the last subsection. The total RISM-based correction of long range solute-solvent and solvent-solvent truncation yields a considerable amount of 0.152 kcal mol^{-1}, i.e. around 8 % of the simulated value. In summary, the RISM theory is able to quantitatively correct for shifted-force and truncation artifacts to the best possible extent, and even yields a good estimate for the full potential limit for the hydration free energy of argon in water: 1.88 kcal mol^{-1} including Ewald truncation corrections.

Further insight into truncation effects is gained from the radial free energy density to be computed from eq. (4.11). Accordingly, for a number of distances the thermodynamic integration has to be carried out which is a very expensive task in the context of a molecular simulation, while the RISM-HNC result follows in a straightforward way via the integrand of eq. (4.23) directly from the correlation functions. The statistical error of the simulation data is considerably large, particularly near the repulsive core region where no result can be obtained without application of special techniques like for instance umbrella sampling. The integral equation on the other hand has no problem with the range of small distances. For illustrative purposes such free energy densities for the 12 Å shifted-force and for the Ewald system (full Coulomb but truncated Lennard-Jones potential) are shown in Figure 4.6. As expected, huge discrepancies are found between absolute RISM and simulation results that can be traced back to the crude HNC approximation, yet the core region simulation results are certainly wrong while the RISM density shows reasonable finite numbers in that region.

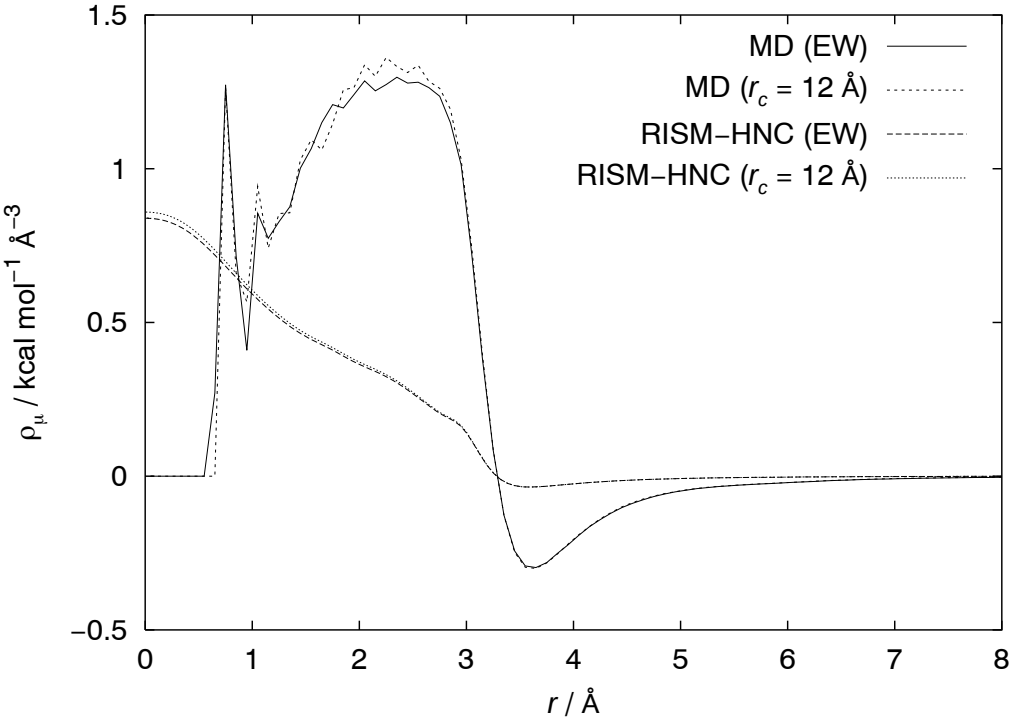

Figure 4.6: Solvation free energy density for argon in water as a function of distance. Solid/short-dashed: simulation results obtained from Ewald and 12 Å shifted-force conditions, short-dashed/dotted: RISM-HNC results under the same set of conditions.

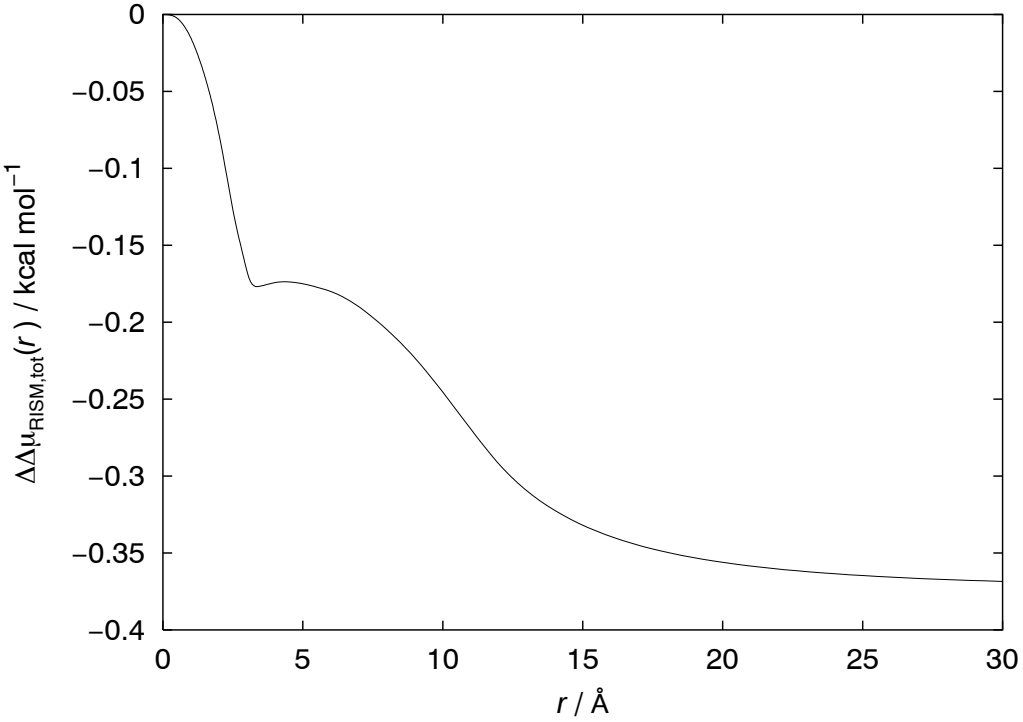

Figure 4.7: Integral function of the RISM-HNC free energy density difference between Ewald and 12 Å shifted-force conditions for the argon-water system.

The simulated densities can not be safely subtracted in order to get a function that, upon integration, yields the total correction discussed above. The difference of integrated RISM results however gives a reliable prediction so that a meaningful and noiseless curve can be expected for the difference of the RISM free energy densities between full (Ewald) potential and shifted-force conditions. The integral function of the difference (derived from the density difference by integration between zero and a variable *r*) is shown in Fig. 4.7. The asymptotic value for large distance corresponds to the total correction of -0.38 kcal mol^{-1}. We find a sharp slope change near the core region after which the curve smoothly runs out and appears to approach the limiting value after more than 20 Å. Based upon the magnitude it is quite plausible to correlate the two distinct steps to the solute-solvent correction in the core region and to the solvent-solvent correction outside, though no quantitative theory is presently available that supports this view. Such an analysis could prove useful in the future for a unanimous definition of solute cavities.

Even more interesting however is the fact that the total correction is obtained from integration up to at least 20 Å. Since the absolute free energy density in the truncated case is zero beyond the cutoff, the Ewald summation is responsible for the long-range character. This means that a simulation box size not smaller than twice this radius is required for converged results at shifted-force simulation without any correction. This result is the first quantitative indication for the presence of finite size effects that requires more elaboration in future work. It is clear that an integral equation theory that does not mimic a periodic environment will not be able to reproduce a structural finite size effect that modifies the distribution functions, but the free energy density difference is a useful instrument for estimating minimum box sizes for sufficiently converged simulations.

4.5.3 Pure water: RISM- and simulation-based correction results

The final test conducted in this work concerning the appropriateness of RISM-based truncation correction consists of applying it to polar systems, in this case for predicting the Ewald limit of the electrostatic free energy component of pure

extended TIP3P water. Pure water is a very common test case for free energy techniques that has for instance been treated with various cutoff schemes as well as with Ewald summation by Brunsteiner and Boresch.[46] As in the latter work the electrostatic free energy difference was determined from canonical MD simulations at prescribed density and temperature; the solvent susceptibility necessary for RISM-based correction has been computed with exactly the simulation conditions, i.e. including Lennard-Jones truncation at the real-space cutoff for Ewald conditions. Additionally, solute-solvent general reaction field corrections[25] were computed by averaging the reaction field over the simulated trajectory data at $\lambda = 1$. To provide the solvent-solvent correction term the GRF was also averaged over the whole simulation cell for $\lambda = 0$ and $\lambda = 1$ with invisible solute. The reaction field dielectric constant was set to the true TIP3P value of 97.[86]

First the performance and reliability of the finite difference Ewald free energy algorithm were checked as proposed in eq. (4.9). The basic functionality of the algorithm was tested by linear scaling of the charges with the CHARMM shifted force potential where the analytical coupling parameter derivative is easily accessible. With the same parameters our method was able to reproduce the analytical derivative for $\Delta\lambda < 10^{-5}$. Then the method was applied to Ewald simulation of extended TIP3P water, again with linear scaling of charges of one water molecule. The free energy derivative was averaged based on the same trajectory for $\lambda = 1$ with decreasing $\Delta\lambda$ values. As expected the free energy derivative approaches an asymptotic limiting value with decreasing $\Delta\lambda$. There is a broad region of stable and converged free energy derivative results for $\Delta\lambda$ ranging between 10^{-5} and 10^{-9} despite the fact that we have to evaluate the small difference of rather big numbers. To ensure sufficient accuracy we finally chose $\Delta\lambda = 10^{-8}$.

Simulation results under Ewald and CHARMM shifted force conditions with cutoff distances 8, 10 and 12 Å are summarized along with RISM corrections, $\Delta\Delta\mu_{RISM}$, and GRF corrections for both solute-solvent and solvent-solvent truncation,[25] $\Delta\Delta\mu_{MD,GRF,tot}$, with respect to Ewald conditions in Table 4.4. The total corrected values computed either by RISM or by GRF theories are denoted by

$\Delta\mu_{RISM,corr}$ and $\Delta\mu_{MD,GRF,corr}$, respectively. Also given are reference data, $\Delta\mu_{ref}$, documented by Brunsteiner and Boresch.[46] Also according to their work, the dipolar finite size correction is negligible. The obvious slight discrepancies between the uncorrected simulation results and the latter work can be traced back to the smaller simulation cell, differing density, and different force-switching method applied to the Lennard-Jones interactions, and to the fact that they used the conventional TIP3P water model.

Table 4.4: Electrostatic free energy of hydration of extended TIP3P water (in kcal mol^{-1}): own results, reference data,[46] RISM and GRF corrections.

r_c / Å	8	10	12	EW
$\Delta\mu_{MD}$	-9.82 ± 0.05	-9.20 ± 0.04	-8.83 ± 0.04	-8.42 ± 0.04
$\Delta\mu_{ref}$	-9.66 ± 0.07	-9.12 ± 0.06	-8.71 ± 0.06	-8.26 ± 0.06
$\Delta\Delta\mu_{MD,GRF,tot}$	2.396	1.179	0.303	/
$\Delta\mu_{MD,GRF,corr}$	-7.42	-8.02	-8.53	/
$\Delta\Delta\mu_{RISM}$	1.56	0.94	0.61	/
$\Delta\mu_{RISM,corr}$	-8.26	-8.26	-8.22	/

A comparison of the own values with those by Brunsteiner and Boresch confirms the right order of magnitude and the trend of the truncation effects. The relative electrostatic truncation effect at 8 Å cutoff amounts to 16 % compared to the Ewald limit. But as the electrostatic part of the free energy in most cases is the dominating contribution, the total error has significant impact. In such a case the statistical ensemble methods, FEP and UMB, totally failed due to the extensive system change upon shifting or truncating Coulomb potentials. The GRF correction terms indicate that the total truncation effect is clearly dominated by the solvent-solvent error. Thus if the hydration free energy of another polar molecules is to be computed the quality of the water description is crucial.

It is quite evident that, while GRF corrects into the right direction, the final results are rather inconsistent among the various cutoffs with respect to the Ewald

reference, worse so with smaller cutoff distance. This is not unexpected since only truncation, not shifting effects are handled by the reaction field. Additionally, the GRF terms are contaminated by rather large errors of the order 1 kcal mol^{-1} when taken from the same trajectory as the original free energy data. The RISM method on the other hand covers both, truncation and shifting, and yields perfectly consistent corrected noise-free values.

There remains however a small systematic deviation between the Ewald reference and the RISM-corrected data of on average 0.17 kcal mol^{-1} that does not seem to disappear with increasing cutoff. A systematic RISM-HNC approximation failure can safely be ruled out from the cutoff independence of the observed trend and the existence of a structural, distribution function-based finite size effect associated with using only 1000 water molecules appears to be the most likely hypothesis. Further indication for this point of view is revealed by subtracting uncorrected values and RISM-corrected values among cutoff simulations: Because of $\Delta\mu_{MD}(10\ Å) - \Delta\mu_{MD}(8\ Å) = 0.62$ kcal mol^{-1} and $\left|\Delta\Delta\mu_{RISM}(10\ Å) - \Delta\Delta\mu_{RISM}(8\ Å)\right| =$ 0.62 kcal mol^{-1}, and, similarly, $\Delta\mu_{MD}(12\ Å) - \Delta\mu_{MD}(10\ Å) = 0.38$ kcal mol^{-1} and $\left|\Delta\Delta\mu_{RISM}(12\ Å) - \Delta\Delta\mu_{RISM}(10\ Å)\right| = 0.33$ kcal mol^{-1}, the finite size effect is caused by using the CHARMM shifted-force potential alone since a RISM-correction from one cutoff distance to another within the same shifted-force protocol gives excellent results. For ultimate clarification, the 8 Å cutoff and the Ewald simulation were repeated with the larger water box. It turned out that the Ewald method shows no measurable difference, as expected from earlier work,[28] whereas the CHARMM value is 0.11 kcal mol^{-1} smaller than the result given in Table 4.4 such that the RISM corrected value would be -8.37 kcal mol^{-1}, i.e. within the error bar of the simulation result.

In summary, for polar systems the RISM-HNC theory also provides a quantitative correction strategy as long as structural finite size effects do not play a substantial role. Further work is required to understand the origin of finite size free energy artifacts as a function of the shifting scheme.

4.6 Conclusion

It has been demonstrated that absolute hydration free energies of apolar and polar solutes, accessible by thermodynamic integration molecular dynamics simulation, are contaminated by a strong sensitivity to truncation and finite size effects, as expected. A number of correction schemes were worked out and applied, among them the RISM integral equation theory within the simple HNC approximation. RISM theory was identified as a powerful tool to analyze and correct truncation effects of free energy simulations in a quantitative manner, in accord with simulation-based protocols (umbrella sampling, perturbation theory, and general reaction field correction) that are much harder to apply and still suffer from inherent statistical uncertainties. Since solute-solvent and solvent-solvent potential truncation were chosen, it was possible to estimate the effects separately. It turned out that both components are similar in magnitude and therefore equally important. Simulation-based correction and RISM theory agree also for these separate terms.

For the two simple model systems, argon in TIP3P water and pure TIP3P water, excellent correction with respect to Ewald reference simulations was achieved, with the RISM-based technique having an advantage in both, the computational ease of application and the numbers obtained. The magnitude of the correction even for the short-ranged dispersive term in the apolar case turned out to be surprisingly large and to be of the same order of magnitude as the total hydration free energy itself for small cutoffs around 8 Å. In the argon case it was furthermore possible to identify solute-solvent and solvent-solvent components as distinct structural features of the integral function of the radial free energy density based upon which a sufficiently large simulation cell dimension for avoiding finite size effects or, more precisely, finite range effects could be estimated. Corresponding with other works,[42,43] reliable free energy computations are only possible by either Ewald summation with at least a long-range Lennard-Jones correction, or by correction schemes like those proposed here. In contrast to common methods that assume a uniform distribution function integral equation theory is capable of correcting long range contributions even when the distribution functions are still structured but not directly accessible due to the

limited size of the simulation cell. The simulations together with RISM corrections were accurate enough to identify the most likely source of the very small remaining differences as structural finite size effects . These have apparently a larger effect on the truncated simulation and will need more consideration in future work. A viable route will use results from explicitly periodic integral equation theories along the lines of for instance the work by Denton and Egelstaff[87] based upon the pioneering study by Pratt and Haan,[88] and the 3D-RISM equation.[89,90] A further limitation of the RISM theory derives from the fact that we need nonzero Lennard-Jones term on each interaction site in order to avoid Coulomb singularities leading to unstable RISM solutions.

Having only simulation data and methods at disposal is not sufficient for successful correction regardless of the system under consideration, at least not at a reasonable cost comparable to the free energy computation effort itself and not without adjustment of sensible parameters like the solvent shell size used for the solvent-solvent shift correction. Since integral equation theory provides an accurate, simple and non-empirical, cost-effective way to compute necessary corrections, such a technique will in the future be applied also to atomic and molecular solutions and to massively truncated solvent-solvent potentials like the enhanced damped Coulomb model[71] that has far less drastic effects on radial distribution functions than the simple CHARMM shifted-force term. It is expected to be able to use cutoff distances down to only 6-7 Å, as was possible with the *g* function corrections in chapter 3. This corresponds to roughly an order of magnitude increase of the computational performance as compared to a real-space cutoff of 12 Å within a Ewald simulation. Preliminary simulation results indicate that the finite size effects observed with the CHARMM function do not occur in this case, comparably with the Ewald technique. Due to the generality and simplicity of the RISM-HNC correction approach the integral equation theory turns our to be a very versatile tool for correcting free energy artifacts coming from both Coulombic or dispersive/repulsive potential truncation and shifting. In this way, the spirit of earlier purely electrostatic truncation corrections (GRF or continuum theory) is extended to more general cases.

4.7 References

(1) D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Chem.* **18**, 431 (1989).

(2) V. Helms and R. C. Wade, *J. Comput. Chem.* **18**, 449 (1997).

(3) R. H. Henchman and J. W. Essex, *J. Comput. Chem.* **20**, 499 (1999).

(4) M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, *J. Chem. Phys.* **119**, 5740 (2003).

(5) S. A. Best, K. M. Merz Jr., and C. H. Reynolds, *J. Phys. Chem. B* **103**, 714 (1999).

(6) M. J. Vlot, J. Huinink, and J. P. van der Eerden, *J. Chem. Phys.* **110**, 55 (1999).

(7) S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, *J. Phys. Chem. B* **107**, 9535 (2003).

(8) P. Kollmann, *Chem. Rev.* **93**, 2395 (1993).

(9) R. J. Radmer and P. A. Kollmann, *J. Comput. Chem.* **18**, 902 (1997).

(10) N. Lu and D. A. Kofke, *J. Chem. Phys.* **115**, 6866 (2001).

(11) G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).

(12) J. W. Pitera and W. F. van Gunsteren, *J. Phys. Chem. B* **105**, 11264 (2001).

(13) D. A. Hendrix and C. Jarzynski, *J. Chem. Phys.* **114**, 5974 (2001).

(14) S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, *J. Chem. Phys.* **119**, 3559 (2003).

(15) G. Grochola, *J. Chem. Phys.* **120**, 2122 (2004).

(16) D. Rodriguez-Gomez, E. Darve, and A. Pohorille, *J. Chem. Phys.* **120**, 3563 (2004).

(17) C. Peter, C. Oostenbrink, A. van Dorp, and W. F. van Gunsteren, *J. Chem. Phys.* **120**, 2652 (2004).

(18) K. M. Åberg, A. P. Lyubartsev, S. P. Jacobsson, and A. Laaksonen, *J. Chem. Phys.* **120**, 3770 (2004).

(19) M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

(20) P. J. Steinbach and B. R. Brooks, *J. Comp. Chem.* **15**, 667 (1994).

(21) P. H. Hünenberger and W. F. van Gunsteren, *J. Chem. Phys.* **108**, 6117 (1998).

(22) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).

(23) M. Deserno and C. Holm, *J. Chem. Phys.* **109**, 7678 (1998).

(24) F. Figueirido, G. S. Del Buono, and R. M. Levy, *J. Chem. Phys.* **103**, 6133 (1995).

(25) H. Resat and J. A. McCammon, *J. Chem. Phys.* **108**, 9617 (1998).

(26) Y. Y. Sham and A. Warshel, *J. Chem. Phys.* **109**, 7940 (1998).

(27) P. H. Hünenberger and J. A. McCammon, *J. Chem. Phys.* **110**, 1856 (1999).

(28) S. Boresch and O. Steinhauser, *J. Chem. Phys.* **115**, 10793 (2001).

(29) Z. Wang and C. Holm, *J. Chem. Phys.* **115**, 6351 (2001).

(30) B. W. Arbuckle and P. Clancy, *J. Chem. Phys.* **116**, 5090 (2002).

(31) M. Bergdorf, C. Peter, and P. H. Hünenberger, *J. Chem. Phys.* **119**, 9129 (2003).

(32) B. G. Rao and U. C. Singh, *J. Am. Chem. Soc.* **112**, 3803 (1990).

(33) P. M. King, *Mol. Phys.* **94**, 717 (1998).

(34) W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.* **83**, 3050 (1985).

(35) A. J. Cross, *Chem. Phys. Lett.* **128**, 198 (1986).

(36) M. Mezei, *J. Chem. Phys.* **86**, 7084 (1987).

(37) T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, *Chem. Phys. Lett.* **222**, 529 (1994).

(38) T. Z. M. Denti, T. C. Beutler, W. F. van Gunsteren, and F. Diederich, *J. Phys. Chem.* **100**, 4256 (1996).

(39) A. Dejaegere and M. Karplus, *J. Phys. Chem.* **100**, 11148 (1996).

(40) T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, *J. Chem. Phys.* **85**, 6720 (1986).

(41) T. C. Beutler, D. R. Béguelin, and W. F. van Gunsteren, *J. Chem. Phys.* **102**, 3787 (1995).

(42) J. T. Slusher, *J. Phys. Chem. B* **103**, 6075 (1999).

(43) J. T. Wescott, L. R. Fischer, and S. Hanna, *J. Chem. Phys.* **116**, 2361 (2002).

(44) C. Chipot, C. Millot, B. Maigret, and P. A. Kollman, *J. Chem. Phys.* **101**, 7953 (1994).

(45) G. Hummer, L. R. Pratt, and A. E. Garcia, *J. Phys. Chem.* **99**, 14188 (1995).

(46) M. Brunsteiner and S. Boresch, *J. Chem. Phys.* **112**, 6953 (2000).

(47) T. P. Straatsma and H. J. C. Berendsen, *J. Chem. Phys.* **89**, 5876 (1988).

(48) R. H. Wood, *J. Chem. Phys.* **103**, 6177 (1995).

(49) G. Hummer, L. R. Pratt, and A. E. Garcia, *J. Phys. Chem.* **100**, 1206 (1996).

(50) S. G. Kalko, G. Sesé, and J. A. Padró, *J. Chem. Phys.* **104**, 9578 (1996).

(51) G. Hummer, L. R. Pratt, and A. E. Garcia, *J. Chem. Phys.* **107**, 9275 (1997).

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

(53) M. Bergdorf, C. Peter, and P. H. Hünenberger, *J. Chem. Phys.* **119**, 9129 (2003).

(54) A. Trokhymchuk and J. Alejandre, *J. Chem. Phys.* **111**, 8510 (1999).

(55) D. Chandler and H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

(56) F. Hirata and P. J. Rossky, *Chem. Phys. Lett.* **83**, 329 (1981).

(57) S. J. Singer and D. Chandler, *Mol. Phys.* **55**, 621 (1985).

(58) C. L. Brooks, B. Montgomery, and M. Karplus, *J. Chem. Phys.* **83**, 5897 (1985).

(59) C. L. Brooks III, *J. Chem. Phys.* **86**, 5156 (1987).

(60) L. Shao, H.-A. Yu, and J. Gao, *J. Phys. Chem. A* **102**, 10366 (1998).

(61) A. Kovalenko and F. Hirata, *J. Chem. Phys.* **113**, 2793 (2000).

(62) Q. Cui and V. H. Smith Jr., *J. Chem. Phys.* **113**, 10240 (2000).

(61) S. Ten-no, *J. Chem. Phys.* **115**, 3724 (2001).

(62) S. M. Kast, *Phys. Rev. E* **67**, 041203 (2003).

(63) S. Ten-no, *J. Chem. Phys.* **115**, 3724 (2001).

(64) S. M. Kast, *Phys. Rev. E* **67**, 041203 (2003).

(65) K. F. Schmidt and S. M. Kast, *J. Phys. Chem. B* **106**, 6289 (2002).

(66) S. M. Kast, K. F. Schmidt, and B. Schilling, *Chem. Phys. Lett.* **367**, 398 (2003).

(67) S. M. Kast, *Phys. Chem. Chem. Phys.* **3**, 5087 (2001).

(68) S. M. Kast, *ChemPhysChem.* **5**, 449 (2004).

(69) A. D. Trokhymchuk, M. F. Holovko, E. Spohr, and K. Heinzinger, *Molec. Phys.* **77**, 903 (1992).

(70) S. Kambayashi and J. Chihara, *Phys. Rev. E* **50**, 1317 (1994).

(71) D. Zahn, B. Schilling, and S. M. Kast, *J. Phys. Chem. B* **106**, 10725 (2002).

(72) J. Hermans, A. Pathiaseril, and A. Anderson, *J. Am. Chem. Soc.* **110**, 5982 (1988).

(73) J. Quintana and A. D. J. Haymet, *Chem. Phys. Lett.* **189**, 273 (1992).

(74) T. R. Forester and W. Smith, *DL_POLY molecular dynamics code*, version 2.14, CCP5 of the EPSRC (1995).

(75) W. E. Reiher III, PhD thesis, Harvard University (1985).

(76) W. L. Jorgensen, J. Chandrasekhar, J. R. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).

(77) J. Guenot and P. A. Kollman, *J. Comp. Chem.* **14**, 295 (1993).

(78) S. M. Kast, K. Nicklas, H. J. Bär, and J. Brickmann, *J. Chem. Phys.* **100**, 566 (1994).

(79) S. M. Kast and J. Brickmann, *J. Chem. Phys.* **104**, 3732, 3732 (1996).

(80) H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).

(81) J. P. Rykaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comp. Phys.* **23**, 327 (1977).

(82) J. Perkyns and B. M. Pettitt, *J. Chem. Phys.* **97**, 7656 (1992).

(83) P. J. Rossky, B. M. Pettitt, and G. Stell, *Mol. Phys.* **50**, 1263 (1983).

(84) A. Kovalenko, S. Ten-no, and F. Hirata, *J. Comput. Chem.* **20**, 928 (1999).

(85) D. J. Price and C. L. Brooks III, *J. Chem. Phys.* **121**, 10096 (2004).

(86) P. Höchtl, S. Boresch, W. Bitomsky and O. Steinhauser, *J. Chem. Phys.* **109**.

(87) A. R. Denton and P. A. Egelstaff, *Z. Phys. B* **103**, 343 (1997).

(88) L. R. Pratt and S. W. Haan, *J. Chem. Phys.* **74**, 1864 (1981).

(89) D. Beglov and B. Roux, *J. Phys. Chem. B* **101**, 7821 (1997).

(90) A. Kovalenko and F. Hirata., *Chem. Phys. Lett.* **290**, 237 (1998).

5 Conclusion

The key idea of this work was to improve the efficiency of molecular dynamics simulation of liquids by applying small spherical cutoff distances and to minimize or correct the corresponding truncation errors. The main tasks were the analysis, minimization and correction of simulation based potential truncation effects. The most successful strategy was finally employed in free energy simulations to calculate reliable hydration free energy values.

To deal with this challenge a powerful molecular dynamics code as well as a RISM integral equation program were necessary. The parallel MD code DLPOLY[1] was extended by the free energy functionality and implemented on high performance super computer clusters.[2,3] The applied RISM code was developed by F. Schmidt und S. M. Kast[4] and all RISM calculations were accomplished by S. M. Kast. Due to practical relevance the simulations were performed with liquid water phase. To restrict the simulation expense the simple three site water models TIP3P[5] and SPC[6] were used. The CHARMM shifted force[7] function was chosen as typical spherical cutoff method and the Ewald summation[8] as full potential reference.

Truncation effects were analyzed at TIP3P and SPC water models with the CHARMM function at various cutoff distances. This was realized by calculating the difference of observables simulated with spherical cutoff conditions and Ewald summation. Static properties like density, excess potential energy and radial distribution functions, dynamical properties like diffusion constant and dielectric properties like dielectric permitivity have been averaged. In agreement with other authors all observables turned out to be sensitive to significant truncation effects. [9-11]

A strategy to minimize truncation errors was performed with an enhencement of the Wolf damped Coulomb potential,[12] a special effective long-range interaction potential, actually construncted for use with ion melts. The original form was modified in oder to remove the inconsistency of force and potential and to apply it to simulations of dipolar liquids. After fitting the damping constant to 0.2 $Å^{-1}$ at 9 Å cutoff radius the properties of pure TIP3P and SPC water phase were evaluated. The

results of potential energy, density, diffusion coefficient and especially the dielectric permitivity were very close to the corresponding Ewald reference values. Because of the smooth decay of the damping function also radial distrucution functions were in excellent agreement with the Ewald reference. Therefore the enhanced damped Coulomb potential was proven to be a useful spherical cutoff alternative to the conventional CHARMM function with either lower truncation effects at the same cutoff distance or allowing smaller cutoff distances at the same error level. The method further provided a considerable speed up compared to Ewald summation.

The correction of truncation effects after simulation was achieved with RISM integral equation theory. Comparing the radial distribution functions of TIP3P water independently generated by MD simulation and RISM theory it was recognized that both methods exhibit exactly the same truncation effects though the absolute observables were obviously different. This knowledge established the idea to combine the techniques in a hybrid approach and to correct simulation truncation errors with RISM integral equation theory. Two ways were explored to link simulation and integral equation. The first approach adopted simulated g functions to adjust special additional functions describing the short-range interaction at IE theory beyond the RISM approximation, the bridge functions. This method requires relatively smooth g functions which implies long MD simulation times. The second perturbation-theoretic approach rescaled simulated observables with a correcting factor determined by RISM theory. It is more flexible and can be easily adapted to other observables. Both strategies were analysed for extended TIP3P[13] water with CHARMM shift and enhanced damped Coulomb potential calculating the radial distribution functions and the excess potential energy. The RISM corrected observables of spherically truncated simulations reached the Ewald limit within statistical error bars. In combination with the CHARMM shift the RISM correction was successfully applied down to 10 Å cutoff radius, whereas with the enhanced damped Coulomb shift the correction was even robust down to 6 Å cutoff radius. As a result the RISM theory was recognized as reliable prediction and correction tool of simulation truncation effects. Moreover referring to the fast fitting of the damping

constants the RISM theory turned out to be a practical force field parameterization tool. Limitations of the RISM correction are the facts that it needs non-zero Lennard-Jones terms at all atomic centers and that only thermodynamic expectation values can be treated but no dynamical properties.

The new RISM based correction strategy was extended to free energy simulation. With thermodynamic integration molecular dynamics (TI/MD) simulation the absolute hydration free energy of argon in water and the electrostatic contribution to the excess chemical potential of water were investigated, again with the extended TIP3P model.[13] Using the CHARMM shifting function for Lennard-Jones and Coulomb interaction the free energy results with various cutoff distances contained severe truncation artifacts as compared to the Ewald reference values. With ensemble based correction methods and continuum electrostatics[14] it was possible to evaluate the total truncation error and to identify several truncation error contributions. Here the RISM correction was capable to reliably reproduce the total truncation errors at both cases, Lennard-Jones and Coulomb interaction, and all separate truncation error contributions. Correcting to full Lennard-Jones and Coulomb potential the real full potential value of hydration free energy of argon in water was calculated, wich is hardly possible with simulation techniques only. Hence the RISM theory has been shown to be a reliable and universal correction method also for free energy simulations. The small remaining differences were traced back to strunctural finite size effects which have to be further analyzed in the future.

The combination of molecular simulation and integral equation theory has been demonstrated to be a valuable strategy to overcome the limits of the individual methods. The stand-alone RISM theory on one hand is known to yield inaccurate results concerning the liquid water phase. This deficit is ascribed to an insufficient modeling of short-range interaction. The long-range capability of RISM however is excellent. Molecular dynamics simulation on the other hand can in principle produce correct results, but to reach an acceptable expense of computer power the long-range interactions are truncated introducing corresponding errors whereas the short-range part remains proper. Thus the calculation of observables using simulation with

truncated potential and subsequent RISM correction to the full potential combines the advantages of both methods. Using small cutoff distances the simulation can be massively accelerated whereas the correction still reaches full potential results. Limits of the approach are defined by limits of the individual methods. The RISM calculation claims a huge demand of computer memory which actually allows for up to 50 different atomic types in the system. Whereas the simulation technique requires an immense investment of computer power. On modern parallel super computers up to 1 million atomic centers within the simulation cell can be simulated for some nanoseconds of time. Both methods will benefit from the development of computer power in the future.

Due to the limitations of simulation methods the most actual force field parameters[15,16] have been fitted using truncation schemes. The analyzed potential truncation effects in this work imply that observables evaluated with a truncation scheme are generally different from the according Ewald values. Conversely an Ewald simulation with force field parameters which were originally adjusted to the truncation scheme produce wrong results. Thus for reliable Ewald simulations the force field parameters should first be fitted to Ewald conditions. This reparametrization has been recently performed for some water models.[11,17]

The same consideration especially holds for free energy simulations. Some authors so far have calculated absolute free energy values with truncated potentials without any correction,[18,19] argumenting that the employed force field was paramerized with the same potential truncation. The most authors however applied a correction[14,20,21], usually limited to electrostatic contributions. Both strategies appear to be questionable regarding the severe truncation effects within free energy simulations. The absolute hydration free energies turned out to be quite sensitive to all kind of intermolecular potential truncation suggesting the full potential as only reasonable reference point for the parameterization as well as the calculation of free energy values. This full potential free energy can either be approximated by Ewald simulation, with long-range integration of the Lennard-Jones potentials or now completely be obtained by truncated simulation combined with a subsequent RISM

correction. Thus the proposal of a full potential force field for use in simulation and integral equation studies is a fundamental result of this work.

An additional result was the first calculation of a radial free energy density based on molecular dynamics simulation data. This quantity, which is easily provided by RISM theory, but inaccurate for water phase, was successfully averaged over very long simulation trajectories of Argon in Water. The radial free energy function yields important information about the range of solvent phase around a solute influencing the free energy of solvation. Thus the minimum size of simulation boxes can be determined that minimizes finite size effects. An extension of the radial free energy function to a corresponding three dimensional free energy density would allow the calculation of a quantitative local hydrophobicity index[22] based on molecular simulation data.

Future investigations of the enhanced damped Coulomb potential can be dedicated to the simulation of other pure and mixed liquids as well as solute molecules in solution in order to verify the general applicability of the damped Coulomb method. Further research can be invested to implement the RISM correction stategy to free energy simulation of atomic and molecular ions in water. Here the results can directly be compared with existing continuum electrostatic based correction methods. A massive acceleration of free energy simulation can possibly be achieved by simulations with small spherical cutoff radius using the enhanced damped Coulomb shifting function and subsequent RISM correction to the full potential. The concept of the full potential force field has to be investigated for some selected systems. The main advantage should be the possibility to calculate reliable absolute free energy values that can be compared with experimantal data. Here the RISM integral equation theory can evolve to a practical tool for force field development.

5.1 References

(1) T. R. Forester and W. Smith, *DL_POLY molecular dynamics code*, version 2.14, CCP5 of the EPSRC (1995).

(2) Regatta, IBM Power 4 Cluster, HRZ TU-Darmstadt.

(3) CSC Opteron Cluster, Uni Frankfurt.

(4) K. F. Schmidt, S. M. Kast, *J. Phys. Chem. B* **106**,6289 (2002).

(5) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein*, J. Chem. Phys.* **79** (1983) 926.

(6) H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht 1981, p 331.

(7) A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, M. Karplus*, J. Am. Chem. Soc.* **117**, 11946 (1995).

(8) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).

(9) P. J. Steinbach, B. R. Brooks, *J. Comp. Chem.* **15**, 667 (1994).

(10) P. H. Hünenberger, W. F. van Gunsteren, *J. Chem. Phys.* **108**, 6117 (1998).

(11) D. J. Price and C. L. Brooks III, *J. Chem. Phys.* **121**, 10096 (2004).

(12) D. Wolf, P. Keblinski,S. R. Phillpot, J. Eggebrecht,. *J. Chem. Phys.* **110**, 8254 (1999).

(13) W. E. Reiher III, PhD thesis, Harvard University (1985).

(14) H. Resat, J. A. McCammon, *J. Chem. Phys.* **108**, 9617 (1998).

(15) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).

(16) A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, M. Karplus*, J. Am. Chem. Soc.* **117**, 11946 (1995).

(17) S. W. Rick, *J. Chem. Phys.* **120**, 6085 (2004).

(18) V. Helms, R. C. Wade, *J. Comp. Chem.* **18**, 449 (1997).

(19) M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *J. Chem. Phys.* **119**, 5740 (2003).

(20) C. Chipot, C. Millot, B. Maigret, P. A. Kollman, *J. Chem. Phys.* **101**, 7953 (1994).

(21) M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys*. **119**, 9129 (2003).

(22) R. Jäger, F. Schmidt, B. Schilling, J. Brickmann, *J. Comput.-Aided Mol. Des.*, 14, **631** (2000).

6 Zusammenfassung

Die moderne Computersimulationstechnik hat sich neben Experiment und Theorie zu einer dritten Säule der Wissenschaft entwickelt. In vielen Bereichen werden Computersimulationen als ein machtvolles Werkzeug eingesetzt um die Eigenschaften und die Entwicklung von Modellsystemen zu untersuchen. Bekannte Anwendungsfelder sind beispielsweise Weltraumforschung, Wettervorhersage oder Fahrzeugdesign. In der Chemie dienen Computer basierte Experimente unter anderem bei der Beschreibung molekularer Vielteilchensysteme auf atomarem Level. Diese molekularen Simulationen[1,2] ermöglichen einen direkten Zugang von den mikroskopischen Moleküleigenschaften zu den makroskopischen Phaseneigenschaften. Alle Aggregatzustände der Materie lassen sich auf der Basis von Simulationsmethoden modellieren. Ein entsprechend weites Feld chemischer Fragestellungen wurde bereits untersucht, zum Beispiel kristalline Festkörper,[3,4] reine Flüssigkeiten,[5,6] flüssige Mischungen,[7] Solvatationsthermodynamik,[8] Phasengleichgewichte,[9] Ligandenbindung[10] und Proteinfaltung.[11,12] Der entscheidene Begrenzungsfaktor für den Einsatz molekularer Simulationen ist der emorme Bedarf an Rechenleistung. Essenziell ist daher eine hohe Effizienz der Simulationsalgorithmen.

Grundlegende Simulationstechniken molekularer Systeme sind die Molekulardynamik (MD) und die Monte-Carlo (MC) Simulation. In der klassischen molekulardynamische Simulation wird in der Regel ein System atomarer Teilchen betrachtet, die über empirische Potenziale miteiander wechselwirken, das so genannte Kraftfeld.[13,14] Die potenzielle Energie wird als Summe von intermolekularen und intramolekularen Potenzialbeiträgen definiert. Die intermolekularen Potenziale werden häufig mit einem Satz von Coulomb Termen für die elektrostatische Wechselwirkung und Lennard-Jones Termen für Kernabstoßung und die London Dispersion modelliert. Bei der MD-Simulation wird die mikroskopische Bewegung der Teilchen auf der Basis einer numerischen Lösung der Newtonschen Bewegungsgleichungen beschrieben. Sie generiert eine zeitliche Abfolge von

Systemkonfigurationen, die Trajektorie. Die makroskopischen Observablen ergeben sich als Zeitmittelwerte der generierten Systemkomfigurationen. Sie sind mit statistischen Fehlern behaftet, die sich durch eine Verlängerung der Simulationszeit reduzieren lassen.

Die freie Solvatationsenthalpie ist eine Schlüsselgröße für chemische Vorgänge in flüssiger Phase.[8] Sie bestimmt maßgeblich die Löslichkeit eines Solutemoleküls in einem Lösungsmittel und beeinflusst die Verteilung, Aggregation und Reaktion von Solutes in Lösung. Wird Wasser als Lösungsmittel in biomolekularen Systemen betrachtet,[15] so spielt die freie Hydratationsenthalpie eine entscheidende Rolle für das Verständnis biomolekularer Prozesse wie der Ligandenbindung oder der Struktur und Funktion von Proteinen. Ein viel diskutiertes Phänomen in wässriger Lösung ist die Hydrophobie.[16,17] Sie beschreibt auf der einen Seite die geringe Löslichkeit unpolarer Moleküle in Wasser auf der anderen Seite die Tendenz unpolarer Moleküle oder Gruppen sich in wässriger Phase aneianderzulagern. Hydrophobe Effekte werden anhand der freien Hydratationsenthalpie oder mittels der freien Transferenthalpie zwischen Wasserphase und einer unpolaren Phase gemessen. Daher können Methoden zur Bestimmung der freien Solvatationsenthalpie auch der Erforschung der Hydrophobie dienen.

Molekulare Simulationstechnik kann für die Berechnung freier Enthalpiedifferenzen zwischen zwei thermodynamischen Zuständen eingesetzt werden.[18-20] Ein erfolgreicher Algorithmus für solche freie Enthalpiesimulationen beschreibt die inkrementelle Umwandlung eines Startzustandes in einen Endzustand. Die entsprechende freie Enthalpieänderung erhält man über die thermodynmaische Integration (TI) oder die Störungstheorie (FEP). Die FEP Methode ist leichter zu implementieren, doch erlaubt nur die TI Formel eine zuverlässigere Fehlerabschätzung. Die resultierende freie Enthalpie, die enthalpische und entropische Beiträge enthält, wird von deutlichen statistischen Schwankungen begleitet, die lange Simulationszeiten erfordern. Freie Enthalpiesimulationen sind daher generell eine Rechleistung fordernde Aufgabe, die jede Möglichkeit der Effizienzsteigerung rechtfertigt.

Der Aufwand molekularer Simulationen wird entscheidend durch die Anzahl N der betrachteten atomaren Teilchen sowie der Art der Wechselwirkungspotenziale bestimmt. Werden bei dem Einsatz von Paarpotenzialen alle Wechselwirkungen berücksichtigt, so müssen bei jedem Simulationsschritt N^2 Paarpotenziale berechnet werden.[21] Diese Aufgabe übersteigt schnell die Resourcen aktueller Hochleistungsrechner, auf denen Systemgrößen von $N = 10^3 - 10^6$ sinnvoll behandelt werden können. Für effiziente Simulationsstrategien wird daher die Teilchenzahl so klein wie möglich gewählt, und endliche Simulationsboxen mit cyclischen Randbedingungen werden eingesetzt um ausgedehnte Bulkphasen zu beschreiben.[1] Eine weitergehende Effizienzsteigerung ermöglicht die Einführung eines sphärischen Abschneideradius.[22] Dabei werden die Paarpotenziale außerhalb des Abschneideradius R_c auf Null gesetzt. Um die numerische Stabilität der Simulation zu gewährleisten wirkt zusätzlich eine Shiftfunktion auf das Potential, die auch die erste Ableitung am Abschneideradius gleich Null setzt. Als Ergebnis sinkt der Rechenaufwand bis auf eine $N \cdot R_C^3$ Proportionalität, was besonders bei kleinen Abschneideradien effizient ist. Eine fortschrittliche Technik für die Behandlung weitreichender, elektrostatischer Potenziale ist die Ewald Summation.[23] Sie wurde ursprünglich für die Berechnung der Gittersumme von Ionenkristallen entwickelt und schließlich für die Simulation adaptiert. Die gesamte Coulomb-Summe eines periodischen Gitters aus Simulationsboxen wird teilweise im Realraum und teilweise im reziproken Raum aufsummiert. Der besonders effiziente *smooth particle mesh Ewald* (SPME) Algorithmus[23] erreicht dabei eine $N \cdot \log(N)$ Proportionalität. Im Prinzip bietet die Ewald Summation die beste Näherung des realen Potenzials. Doch selbst die effiziente SPME Methode ist noch deutlich aufwendiger als kleine sphärische Abschneideradien.

Die Ergebnisse molekularer Simulationen werden generell als sinnvoll betrachtet, solange die Kraftfeldparameter akurat angepasst wurden. Doch die effizienten Simulationsstrategien erzeugen systematische Fehler in den berechneten Observablen.[22,24,25] Zwei wesentliche Fehlerquellen werden unterschieden: Die Verwendung einer begrenzten Anzahl von Teilchen in der Simulationsbox erzeugt

einen *finite size* Effekt, während die Näherung des Wechselwirkungspotentzals einen Abschneidefehler hervorruft. Unter sphärischen Abschneidebedingungen mit cyclischen Randbedingungen sind beide Effekte wirksam. *Finite size* Effekte können nur durch eine entsprechende Vergrößerung des Systems vermieden werden, doch Abschneidefehler lassen sich mit geeigneten Simulationsalgorithmen reduzieren oder mittels theoretischer Ansätze abschätzen. Für einige Observablen wurden spezielle Korrekturverfahren vorgeschlagen,[26] doch eine generelle Korrekturmethode für Abschneidefehler ist in der Literatur nicht beschrieben.

Ein alternativer Ansatz für die Berechnung thermodynamischer Eigenschaften fluider Systeme basierend auf der statistischen Theorie von Flüssigkeiten ist die Integralgleichungstheorie (IE Theorie).[27-30] Dabei muss ein Satz nichtliniearer Integralgleichungen iterativ gelöst werden. Unter den Verfahren zum Lösen der Integralgleichungen ist das eindimensionale *reference interaction site* Modell (RISM)[30] mit der *hypernetted chain closure* (HNC) als robuste Methode bekannt. Die numerische Lösung der RISM Gleichungen führt direkt zu den radialen Paarverteilungsfunktionen (g Funktionen) eines Systems. Diese Funktion beschreibt die Wahrscheinlichkeit ein Paar von Atomen bei einem bestimmten Abstand *r* voneinander zu finden. Andere thermodynamische Eigenschaften können über analytische Ausdrücke abgeleitet werden. Die RISM Integralgleichungstheorie kann die selben Kraftfeldfunktionen und -parameter einsetzen wie molekulare Simulationen. Daher sind beide Methoden zwei alternative Wege um thermodynamische Eigenschaften von fluiden Phasen basierend auf atomaren Kraftfeldern zu erhalten. Die RISM Theorie hat den Vorteil für moderat große Moleküle wesentlich schneller zu sein und Ergebnisse frei von statistischen Fehlern zu liefern, wohingegen die Simulation langsam ist und statistische Fehler aufweist. Doch die aktuelle Güte der Näherung der RISM Theorie speziell bei der Beschreibung einer Wasserphase ist für die Berechnung absoluter thermodynamischer Größen noch nicht ausreichend. Ergebnisse aus Simulation und Integralgleichung wurden bereits miteinander verglichen[31] und mögliche Abschneidefehler mit Hilfe der IE Theorie untersucht.[32,33]

Ein neues Konzept ist die direkte Kombination von Simulation und Integralgleichung in hybriden Ansätzen,[34,35] um die Defizite der einzelnen Methoden zu überwinden.

Das Ziel dieser Arbeit war die Entwicklung von Strategien für die Steigerung der Effizienz molekularer Simulationen, speziell freie Enthalpiesimulationen. Das kann durch Einsatz sphärischer Abschneidebedingungen bei möglichst kleinen Abschneideradien erreicht werden, impliziert jedoch erhebliche Abschneidefehler. Zwei Ideen scheinen aussichtsreich diese Abschneideeffekte zu reduzieren. Der erste Ansatz versucht die Fehler durch die Verwendung eines geeignet parametrisierten effektiven Wechselwirkungspotenzials so weit wie möglich zu minimieren. Der zweite Ansatz basiert auf einer Kombination von Simulation und Integralgleichung mit dem Ziel eine generelle Korrekturmethode für Abschneideeffekte zu finden, die nach der Simulation angewendet werden kann. Die erfolgreichste Strategie soll schließlich bei der Simulation absoluter freier Hydratationsenthalpien eingesetzt werden. Gegenstand dieser Arbeit ist somit die Analyse, Minimierung und Korrektur von Abschneideeffekten bei Simulationen mit sphärisch abgeschnittenen Wechselwirkungspotenzialen.

Um die Aufgabe zu bewältigen waren ein leistungsfähiger Molekulardynamik Code als auch ein RISM Integralgleichungsprogramm notwendig. Der MD-Code DLPOLY[36] wurde um die freie Enthalpie Funktionalität erweitert und auf parallelen Hochleistungsrechnern[37,38] implementiert. Der eingesetzte Integralgleichungscode geht zurück auf F. Schmidt und S. M. Kast.[39] Alle RISM Berechnungen wurden von S. M. Kast durchgeführt. Auf Grund der praktischen Bedeutung wurde für die Simulationen die flüssige Wasserphase ausgewählt. Um den Simulationsaufwand zu begrenzen kamen dabei die einfachen, dreiatomigen Wassermodelle TIP3P[40] und SPC[41] zum Einsatz. Als typische sphärische Abschneidemethode wurde die CHARMM Funktion[42] gewählt und die Ewald Summation[23] als Referenz für das vollständige Potential.

Die Abschneideeffekte wurden an den Wassermodellen TIP3P und SPC unter Einsatz der CHARMM Funktion mit verschiedenen Abschneideradien untersucht. Erreicht wurde das durch die Berechnung der Differenz von Observablen simuliert

unter sphärischen Abschneidebedingungen und bei Ewald Bedingungen. Statische Größen wie Dichte, potenzielle Exzessenergie und radiale Paarverteilungsfunktionen, sowie dynamische Eigenschaften wie Diffusionskoeffizienten und dielektrische Eigenschaften wie die dielektrische Konstante wurden gemittelt. In Übereinstimmung mit anderen Autoren zeigten sich alle Observablen als deutlich empfindlich gegenüber Abschneideeffekten.[22,43,44]

Eine Strategie zur Minimierung der Abschneidefehler wurde mit der Optimierung des Wolf Potenzials[45] verfolgt, eines speziellen, effektiven, lang reichweitigen Wechselwirkungspotenzials, das ursprünglich für den Einsatz bei Ionenschmelzen entwickelt wurde. Durch Modifikation der Orginalform konnte die bestehende Inkonsistenz von Kraft- und Potenzialfunktion behoben werden. Nach erfolgreicher Parametrisierung der Dämpfungskonstante für flussiges Wasser zu 0.2 Å bei 9 Å Abschneideradius wurden die Eigenschaften von reiner TIP3P und SPC und Wasserphase berechnet. Die Ergebnisse von potenzieller Energie, Dichte, Diffusionskoeffizient und besonders die dielektrische Konstante lagen sehr nah bei den entsprechenden Ewald Referenzwerten. Durch den sachten Verlauf der Dämpfungsfunktion ergaben sich zudem die radialen Paarverteilungsfunktionen in exzellenter Übereinstimmung mit der Ewald Referenz. Das verbessterte Wolf Potenzial hat sich daher als eine vielversprechende Alternative zur konventionellen CHARMM Funktion erwiesen mit entweder geringeren Abschneideeffekten bei gleichem Abschneideradius oder kleineren Abschneideradien bei vergleichbarem Fehlerniveau. Zudem zeigt es eine deutliche Beschleunigung im Vergleich zur Ewald Referenz.

Eine Korrektur von Abschneideeffekten nach der Simulation konnte mit der RISM Integralgleichungstheorie erreicht werden. Ein direkter Vergleich der radialen Paarverteilungsfunktionen von TIP3P Wasser, zum einen berechnet aus Simulationstrajektorien und zum anderen erstellt mit der RISM Theorie, führte zu der Erkenntnis, dass beide Methoden exakt die gleichen Abschneideeffekte aufweisen, obwohl die absoluten Observablen offensichtlich unterschiedlich waren. Diese Entdeckung etablierte die Idee beide Techniken in einem hybriden Ansatz zu

verbinden und Abschneideeffekte von Simulationen mit Hilfe der Integralgleichungstheorie zu korrigieren. Zwei Wege wurden erkundet um Simulation und Integralgleichung miteinander zu verknüpfen. Der erste Ansatz verwendet aus einer Simulation generierte g Funktionen um spezielle zusätzliche Funktionen anzupassen, die die kurzreichweitige Wechselwirkung der IE-Theorie über die RISM Näherung hinausgehend beschreiben, die *bridge* Funktionen. Diese Methode benötigt relativ glatte g Funktionen, was lange Simulationszeiter erfordert. Der zweite störungstheoretische Ansatz wichtet die simulierten Observablen mit einem Korrekturfaktor aus der RISM Theorie. Dieser Weg ist weitaus flexibler und kann leicht auf andere Observablen übertragen werden. Beide Strategien wurden an Hand des erweiterten TIP3P Wassermodells[46] in Verbindung mit der CHARMM Funktion und dem verbesserten Wolf Potenzial analysiert und radiale Paarverteilungsfunktionen sowie die potenzielle Exzessenergie berechnet. Die RISM korrigierten Observablen der abgeschnittenen Simulationen erreichten innerhalb statistischer Fehlerbalken die jeweiligen Ewald Referenzwerte. In Kombination mit der CHARMM Funktion konnte die RISM Korrektur bis zu einem Abschneideradius von 10 Å erfolgreich eingesetzt werden. Zusammen mit dem verbesserten Wolf Potenzial war die Korrektur sogar bis zu einem Abschneideradius von 6 Å stabil. Als Ergebnis wurde die RISM Korrektur als leistunsstarkes Vorhersage- und Korrekturwerkzeug für Abschneideeffekte analysiert. Zudem hat sich die Integralgleichungstheorie bei der Parametrisierung der Dämpfungskonstante des erweiterten Wolf Potenzials als praktisches Kraftfeldparametrisierungswerkzeug erwiesen. Die RISM Methode benötigt allerdings Lennard-Jones Terme ungleich Null an allen atomaren Zentren und nur thermodynamische Erwartungswerte können behandelt, aber keine dynamischen Größen.

Die neue RISM Korrekturstrategie wurde auf freie Enthalpiesimulationen erweitert. Mittels Molekulardynamik und thermodynamischer Integration (TI/MD) konnte die absolute freie Hydratationsenthalpie von Argon in Wasser und der elektrostatische Anteil des chemischen Zusatzpotenzials von Wasser erfolgreich berechnet werden, wiederum mit dem erweiterten TIP3P Modell.[46] Unter Einsatz der

CHARMM Funktion für beide, Lennard-Jones und Coulomb Terme, wurden die freien Enthalpiewerte für eine Reihe von Abschneideradien bestimmt. Sie wiesen erhebliche Abschneideeffekte bezogen auf die jeweilige Ewald Referenz auf. Mit Ensemble basierten Korrekturmethoden und Kontinuumselektrostatik[47] konnten die Gesamtfehler errechnet und mehrere, separate Fehlerbeiträge identifiziert werden. Die RISM Korrektur war in beiden Fällen, Lennard-Jones und Coulomb Wechselwirkung, in der Lage, die gesamten Abschneidfehler und alle gefundenen Fehlerbeiträge zu reproduzieren. Durch Korrektur auf ein vollständiges Lennard-Jones und Coulomb Potential konnte zum ersten Mal der Wert der freien Hydratationsenthalpie von Argon in Wasser bei vollständigem Potenzial ermittelt werden, was mit Simulationstechniken alleine kaum möglich ist. Die RISM Theorie hat sich daher auch bei freien Enthapliesimulationen als zuverlässige und universelle Korrekturmethode erwiesen.

Die hybride Kombination aus molekularer Simulation und RISM Integralgleichung wurde als eine wertvolle Strategie erkannt, die Grenzen der individuellen Methoden zu überwinden. Einerseits ist die RISM Theorie alleine nicht in der Lage zuverlässige Ergebnisse für eine Wasserphase zu liefern. Dieses Defizit wird einer unzureichenden Modellierung der kurzreichweitigen Wechselwirkung zugeschrieben. Die Fähigkeiten der RISM Methode zur Beschreibung langreichweitiger Wechselwirkungen sind jedoch exzellent. Die Molekulardynamik Simulation andererseits kann theoretisch korrekte Ergebnisse liefern. Doch um den Rechenaufwand auf ein akzeptables Maß zu reduzieren, werden die weitreichenden Wechselwirkungspotenziale abgeschnitten und entsprechende Abschneidefehler impliziert. Die kurzreichweitigen Wechselwirkungen hingegen bleiben exakt. Eine Berechnung der Observablen mittels Simulation bei abgeschnittenen Potenzialen und anschließende RISM Korrektur auf das vollständige Potenzial verbindet nun die Vorteile beider Methoden. Durch die Verwendung kleiner Abschneideradien kann die Simulation massiv beschleunigt werden, wobei die RISM Korrektur immer noch den Wert für das vollständige Potenzial erreicht. Grenzen des Ansatzes ergeben sich durch die Grenzen der einzelnen Methoden. Die RISM Berechnung hat einen großen

Bedarf an Speicher der aktuell bis zu 50 verschiedene Atomsorten im System zulässt. Die Simulationstechnik hingegen ist auf eine große Rechenleistung abgewiesen. Auf modernen, parallelen Supercomputern können heute bis zu 1 Millionen atomare Zentren für einen Zeitraum von einigen Nanosekunden simuliert werden. Beide Methoden werden von der zukünftigen Entwicklung der Computertechnik stark profitieren.

Auf Grund der praktischen Grenzen der Simulationsalgorithmen sind aktuelle Kraftfeldparameter meistens mit Abschneidemethoden parametrisiert worden.[13,14] Die in dieser Arbeit analysierten, deutlichen Abschneideeffekte zeigen, dass sich die mit Abschneidemethoden berechneten Observablen grundsätzlich von den entsprechenden Ewald Referenzwerten unterscheiden. Umgekehrt hat das zur Folge, dass Ewald Simulationen mit Kraftfeldparametern, die an Abschneideverfahren angepasst wurden, falsche Ergebnisse liefern. Für zuverlässige Ewald Simulationen sollten die Kraftfeldparameter vorher an die Ewald Bedingungen angepasst werden. Eine solche Reparametrisierung wurde kürzlich an einigen Wassermodelle vorgenommen.[6,48]

Entsprechende Überlegungen gelten in besonderem Maße für freie Enthalpiesimulationen. Einige Autoren haben absolute freie Enthalpien bisher mit abgeschnittenen Potenzialen ohne jede Korrektur berechnet,[49,50] mit dem Argument, dass die verwendeten Kraftfeldparameter auch unter diesen Bedigungen parametrisiert wurden. Die meisten Autoren jedoch haben eine Korrektur angewendet,[46,51,52] meistens begrenzt auf elektrostatische Beiträge. Beide Strategien sind nun fraglich, wenn man die erheblichen, gefundenen Abschneideeffekte bei den simulierten freien Enthalpiewerten bedenkt. Die absoluten freien Hydratationsenthalpien erwiesen sich als sehr empfindlich gegen jede Art der Potenzialmodifikation. Als einziger verlässlicher Bezugspunkt für eine Parametrisierung und Berechnung freier Enthalpiewerte kann daher nur das vollständige Potenzial gelten. Diese freie Enthalpie bei vollständigem Potenzial kann entweder durch Ewald Simulation mit langreichweitiger Integration der Lennard-Jones Potentiale angenähert oder jetzt komplett über eine abgeschnittene Simulation mit anschließender RISM Korrektur

erhalten werden. Der Vorschlag eines Kraftfeldes mit vollständigen Potenzialen ist ein fundamentales Ergebnis dieser Arbeit.

Ein weiteres Resultat ist die erste Berechnung einer radialen freien Enthalpiedichte aus Molekulardynamik Simulationsdaten. Diese Größe läßt sich zwar leicht aus der RISM Theorie berechnen, ist jedoch bezüglich der Wasserphase fehlerhaft. Diese Funktion wurde jetzt am Beispiel von Argon in Wasser zum ersten Mal aus den Trajektorien einer freien Enthlapiesimulation ermittelt. Die radiale freie Enthalpiefunktion liefert wichtige Informationen über die Reichweite, in der ein Solute eine Lösungsmittelphase beeinflusst. So kann die minimale Größe von Simulationsboxen mit möglichst kleinen *finite size* Effekten abgeschätzt werden. Die Erweiterung der radialen freien Enthalpiefunktion auf eine dreidimensionale freie Enthalpiedichte würde die Berechnung eines quantitativen, lokalen Hydrophobieindexes[53] auf der Basis von Molekulardynamik Simulationen ermöglichen.

Eine Weiterentwicklung des verbesserten Wolf Potenzials ist denkbar sowohl für die Simulation von anderen reinen Flüssigkeiten als auch von gemischten Lösungsmitteln und Solutes in Lösung, mit dem Ziel die universelle Anwendbarkeit des Potenzials zu belegen. Weitere Forschungsarbeit kann bei der Anwendung von Simulation und RISM Korrektur auf ionische Solutes in Lösung geleistet werden. Hier können die Ergebnisse direkt mit den Ergebnissen von Kontinuumselektrostatik Methoden verglichen werden. Eine deutliche Beschleunigung von freien Enthalpiesimulationen sollte durch möglichst kleine Abschneideradien unter Einsatz des verbesserten Wolf Potenzials und anschließender RISM Korrektur möglich sein. Das Konzept eines Kraftfeldes mit vollständigen Potenzialfunktionen als Referenzpunkte für die Parametrisierung und Berechnung von Observablen sollte an ausgewählten Systemen untersucht werden. Der wesentliche Vorteil gegenüber den heute vorhandenen Kraftfeldern ist die Möglichkeit, auch absolute freie Enhtlapiewerte zuverlässig berechnen zu können. Dabei kann sich die RISM Integralgleichung zudem als ein praktisches Kraftfeld Parametrisierungswerkzeug erweisen.

6.6 Literatur

(1) M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

(2) W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem.* **102**, 1020 (1990).

(3) I. Borzsak, P. T. Cummings, *Chem. Phys. Lett.* **300**, 359 (1999).

(4) N. Grishina, V. Buch, *J. Chem. Phys.* **120**, 5217 (2004).

(5) T. Head-Gordon, G. Hura, *Chem. Rev.* **102**, 2651 (2002).

(6) D. J. Price, C. L. Brooks III, *J. Chem. Phys.* **121**, 10096 (2004).

(7) K. M. Kast, J. Brickmann, S. M. Kast, R. S. Berry, *J. Phys. Chem. A* **107**, 5342 (2003).

(8) M. Orozco, F. J. Luque, *Chem. Rev.* **100**, 4187 (2000).

(9) S. A. Best, K. M. Merz,Jr., C. H. Reynolds, *J. Phys. Chem. B* **103**, 714 (1999).

(10) S. Boresch, F. Tettinger, M. Leitgeb, *J. Phys. Chem.* **107**, 9535 (2003).

(11) M. Karplus, G. A. Petsko, *Nature* **347**, 631 (1990).

(12) C. L. Brooks III, D. A. Case, *Chem. Rev.* **93**, 2487 (1993).

(13) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).

(14) A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, M. Karplus, *J. Am. Chem. Soc.* **117**, 11946 (1995).

(15) S. K. Pal, A. H. Zewail, *Chem. Rev.* **104**, 2099 (2004).

(16) L. R. Pratt, A. Pohorille, *Chem. Rev.* **102**, 2671 (2002).

(17) W. Blokzijl, J. B. F. N. Engberts, *Angew. Chem.* **105**, 1610 (1993).

(18) D. L. Beveridge, F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431 (1989).

(19) W. F. van Gunsteren, P. K. Weiner, Ed., *Computer Simulation of Biomolecular Systems* (Escom, Leiden, 1989).

(20) P. Kollmann, *Chem. Rev.* **93**, 2395 (1993).

(21) C. Sagui, T. Darden, *J. Chem. Phys.* **114**, 6578 (2001).

(22) P. J. Steinbach, B. R. Brooks, *J. Comp. Chem.* **15**, 667 (1994).

(23) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).

(24) Y. Y. Sham, A. Warshel, *J. Chem. Phys.* **109**, 7940 (1998).

(25) M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys.* **119**, 9129 (2003).

(26) H. Resat, J. A. McCammon, *J. Chem. Phys.* **108**, 9617 (1998).

(27) J.-P. Hansen, I. R. McDonald, *Theory of simple liquids*, 2nd edn., Academic Press, London, 1991.

(28) D. Chandler and H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

(29) F. Hirata and P. J. Rossky, *Chem. Phys. Lett.* **83**, 329 (1981).

(30) D. Chandler, H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

(31) Q. Du, D. Beglov, B. Roux, *J. Phys. Chem. B* **104**, 796 (2000).

(32) C. L. Brooks III, *J. Chem. Phys.* **86**, 5156 (1987).

(33) C. L. Brooks, B. Montgomery, and M. Karplus, *J. Chem. Phys.* **83**, 5897 (1985).

(34) K. F. Schmidt and S. M. Kast, *J. Phys. Chem. B* **106**, 6289 (2002).

(35) S. M. Kast, *Phys. Chem. Chem. Phys.* **3**, 5087 (2001).

(36) T. R. Forester and W. Smith, *DL_POLY molecular dynamics code*, version 2.14, CCP5 of the EPSRC (1995).

(37) Regatta, IBM Power 4 Cluster, HRZ TU-Darmstadt.

(38) CSC Opteron Cluster, Uni Frankfurt.

(39) K. F. Schmidt, S. M. Kast, *J. Phys. Chem. B* **106**,6289 (2002).

(40) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79** (1983) 926.

(41) H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht 1981, p 331.

(42) A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, M. Karplus, *J. Am. Chem. Soc.* **117**, 11946 (1995).

(43) P. H. Hünenberger, W. F. van Gunsteren, *J. Chem. Phys.* **108**, 6117 (1998).

(44) D. J. Price and C. L. Brooks III, *J. Chem. Phys.* **121**, 10096 (2004).

(45) D. Wolf, P. Keblinski,S. R. Phillpot, J. Eggebrecht,. *J. Chem. Phys.* **110**, 8254 (1999).

(46) W. E. Reiher III, PhD thesis, Harvard University (1985).

(47) H. Resat, J. A. McCammon, *J. Chem. Phys.* **108**, 9617 (1998).

(48) S. W. Rick, *J. Chem. Phys.* **120**, 6085 (2004).

(49) V. Helms, R. C. Wade, *J. Comp. Chem.* **18**, 449 (1997).

(50) M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *J. Chem. Phys.* **119**, 5740 (2003).

(51) C. Chipot, C. Millot, B. Maigret, P. A. Kollman, *J. Chem. Phys.* **101**, 7953 (1994).

(52) M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys*. **119**, 9129 (2003).

(53) R. Jäger, F. Schmidt, B. Schilling, J. Brickmann, *J. Comput.-Aided Mol. Des.*, 14, **631** (2000).

Hilfsmittel

Die molekulardynamischen Simulationen wurden mit dem Programm DL_POLY (Version 2.14, CCP5 am EPSRC, England, 1995) durchgeführt. Für die Realisierung des verbesserten Wolf Potenzials und die Berechnung der freien Enthalpiewerte wurde der Orginalcode (FORTRAN 90) entsprechend erweitert.

Die Simulationen erfolgten auf verschiedenen Hessischen Hochleistungsrechnern:

SNI VPP300/6 am HRZ Darmstadt
IBM Power4 Cluster am HRZ Darmstadt
AMD Opteron Cluster des CSC Frankfurt

Bei der Analyse der umfangreichen Simulationstrajektorien kamen ein eigenes, in C programmiertes Auswertungstool und Mathematica 5.0 zum Einsatz. Der Ablauf wurde über Perl Skripte gesteuert.

Die Berechnungen zur RISM Integralgleichungstheorie gehen zurück auf ein Programm von F. Schmidt und S. M. Kast.

Die Plots in dieser Arbeit wurden mit Gnuplot (Version 3.7) erstellt und der Text mit Microsoft Word 2000.

Lebenslauf

Name:		Bernd Schilling
Geburtsdatum:		06.03.1970
Geburtsort:		Köln
Familienstand:		ledig
Schule:	1976 – 1980	Gemeinschaftsgrundschule Köln-Dellbrück
	1980 – 1989	Städtisches Gymnasium Köln-Mühlheim
Studium:	WS 1990/91 – WS 1996	Chemiestudium an der Technischen Universität Darmstadt
	SS 1994 – WS 1996	Stipendium der Studienstiftung des Deutschen Volkes
	1. Oktober 1996	Diplom-Hauptexamen Diplomarbeit: „Berechnung Freier Hydratationsenthalpien hydrophober Teilchen auf der Basis molekulardynamischer Simulationen“
Promotion:	seit 1. Januar 1997	Arbeitskreis von Prof. Dr. J. Brickmann Institut für Physikalische Chemie TU Darmstadt Wissenschaftlicher Mitarbeiter, DFG
	1999 – 2002	Kollegiat im Graduiertenkolleg „Kinetik und Mechanismen von Ionen-Reaktionen“

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich meine Dissertation selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Darmstadt, 8. August 2005