Scalable improvement of SPME multipolar electrostatics in anisotropic polarizable molecular mechanics using a general short-range penetration correction up to quadrupoles

Christophe Narth, Louis Lagardère, Etienne Polack, Nohad Gresh, Qiantao Wang, David R. Bell, Joshua A. Rackers, Jay W. Ponder, Pengyu Y. Ren, Jean-Philip Piquemal

Scalable improvement of SPME multipolar electrostatics in anisotropic polarizable molecular mechanics using a general short-range penetration correction up to quadrupoles

Christophe Narth,[†] Louis Lagardère,[‡] Étienne Polack,[†,¶] Nohad Gresh,[†,§] Qiantao Wang,[∥] David R. Bell,[∥] Joshua A. Rackers,[⊥] Jay W. Ponder,[⊥] Pengyu Y. Ren,[*,∥] and Jean-Philip Piquemal[*,†]

UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, Chemistry and Biology Nucleo(s)tides and immunology for Therapy (CBNIT), UMR 8601 CNRS, UFR Biomédicale, Paris 75006, France, Department of Biomedical Engineering, The University of Texas at Austin, Texas 78712, and Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, Missouri 63110

E-mail: pren@mail.utexas.edu; jpp@lct.jussieu.fr

Abstract

We propose a general coupling of the Smooth Particle Mesh Ewald (SPME) approach for distributed multipoles to a short-range charge penetration correction modifying the charge-charge, charge-dipole and charge-quadrupole energies. Such an approach significantly improves electrostatics when compared to ab initio values and has been calibrated on Symmetry-Adapted Perturbation Theory (SAPT) reference data. Various neutral molecular dimers have been tested as results on charged systems (metal cation complexes) are provided. Transferability of the correction is adressed in the context of the implementation of the AMOEBA and SIBFA polarizable force fields in the TINKER-HP software. As the choices of the multipolar distribution are discussed, conclusions are drawn for the future penetration-corrected polarizable force fields highlighting the mandatory need of non-spurious procedures for the obtention of well balanced and physically meaningful distributed moments. Finally scalability and parallelism of the short-range corrected SPME approach are addressed, demonstrating that the damping function is computationally affordable and accurate for molecular dynamics simulations of complexe bio- or bioinorganic systems in periodic boundary conditions.

1 Introduction

The field of molecular modeling is growing exponentially. The increase of available crystallographic structures[1–3] over the years has posed a substantial challenge for molecular force field development. For such systems, classical approaches still encounter successes but also reach their limits when describing non-bonded interactions which can only be reliably computed

*To whom correspondence should be addressed
†UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France
‡UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France
¶UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
§Chemistry and Biology Nucleo(s)tides and immunology for Therapy (CBNIT), UMR 8601 CNRS, UFR Biomédicale, Paris 75006, France
‖Department of Biomedical Engineering, The University of Texas at Austin, Texas 78712
⊥Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, Missouri 63110

using quantum chemistry. Year after year, it becomes clear that the gap between quantum and classical approaches is too wide to be ignored, should one attempt to be predictive. First-generation force fields ignore polarization effects and employ a simple electrostatic term while a van der Waals interaction term takes the role of a buffer contribution including all quantum-like effects. To compute interaction energies of high accuracy that could be compared directly to quantum chemistry and/or experiments, one should build force fields that embody all anisotropic effects and are able to separately reproduce all quantum contributions. Among all terms, electrostatics remains prominent.[4,5] Point charges can appropriately capture some long-range electrostatics but, quantum chemistry clearly shows the need to go beyond such an approximation to handle anisotropic interactions. Usually, distributed multipoles are used to account for polarization, but they are devoid of the short-range electrostatic quantum effects, the so-called charge penetration, which occurs when the molecular densities overlap. Various strategies have been employed to include such effects. They range from damping functions[6,7] to neural networks[8] and to newly developed point-distributed electron densities such as the GEM[9,10] potential. Quantum mechanical force fields[11] also address this topic. Clearly, both short- and long-range electrostatics do matter and one would like to treat them in concert. This was done in the context of GEM that uses continuous electrostatics which resort to electron densities. To do so, a generalized Smooth Particle Mesh Ewald approach was introduced which handles Hermite Gaussians thereby including penetration effects. As shown initially by Darden *et al*,[12,13] the PME and SPME strategies provide a scaling improvement enabling to deal with electrostatics in periodic boundary conditions with a $N\log(N)$ complexity as compared to the usual direct space N^2. In this contribution we propose a simple strategy based on a damping function to include short-range penetration within a distributed multipole SPME framework. Its aim is to be applicable to modern polarizable molecular dynamics using force fields such as AMOEBA[14] and SIBFA.[7,15–19] The paper is organized as follows. We first present our choice of charge penetration correction. We then briefly recall the SPME methodology for distributed multipoles before presenting a

newly developed coupled approach. Finally, we present the parametrization procedure, the numerical results and a parallel implementation.

2 Charge Penetration

The question of the atomic charge description has previously been investigated.[7,9,20–25] In classical force fields, the charge distribution is modeled using point charges calibrated on the Molecular Electrostatic Potential (MEP) computed with QM (Quantum Mechanical) methods.[26] Although the anisotropy, is not directly taken into account even in the case of MEP fitting,[27,28] this method is still widely used among various communities as it allows the simulation of very large systems due to its low computational cost.

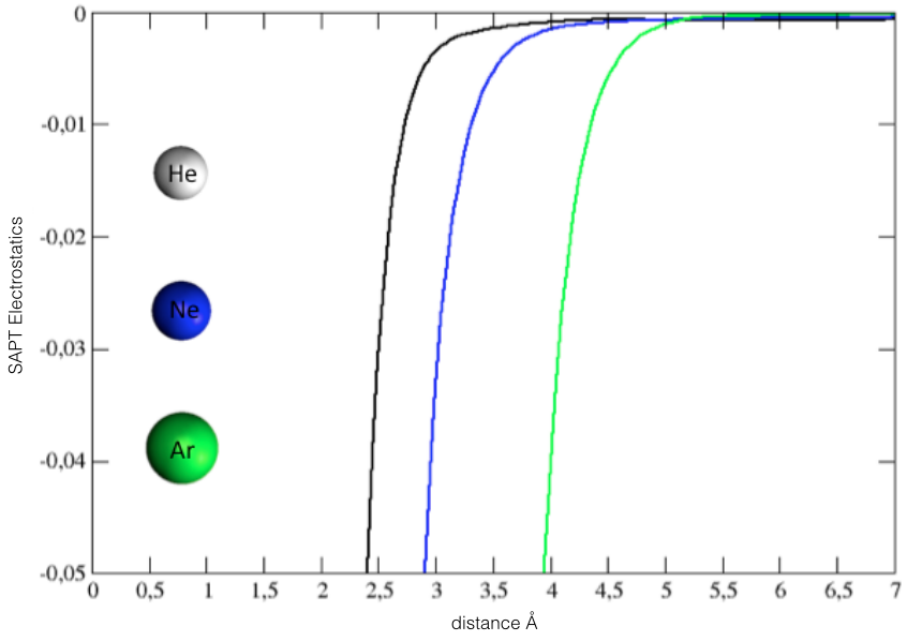
The Distributed Multipole Analysis[29] (DMA) has been shown to be a suitable description of the charge distribution to improve electrostatics. Other distributed multipole methods exist. For example, one of the first full-fledged polarizable force fields, SIBFA,[7,15–18] (Sum of Interactions Between Fragments Ab-initio computed), originally used the multipole partitioning scheme developed by Vigné-Maeder and Claverie.[30,31] Nevertheless, none of these multipole methods properly account for the overlap of molecular densities at short-range. In physical systems, this overlap gives rise to a mainly attractive energy denoted electrostatic penetration. Hence, distributed multipolescan not be rigorously accurate at all ranges because of their classical approximation of electrostatics: theycan not describe the purely quantum mechanical charge penetration effect.

From quantum intermolecular perturbation theory, one can see that charge penetration exhibits an exponential decay being density overlap-dependent.[32]

Charge penetration has been widely studied in QM. For example, Jeziorska *et al* studied Helium dimer intermolecular potential to find a classical description of its potential by means of Energy Decomposition Analysis (EDA), here Symmetry Adapted Perturbation Theory[33,34] (SAPT).

We extended this SAPT study to Neon and Argon dimers. Figure 1 exhibits the electrostatic component for these three rare gas dimers as a function of their interatomic distances. One can notice i) the existence of a non-zero electrostatic energy component at short-range and ii) the confirmation of the exponential decay. This observation and the fact that the rare gas charge is zero, and therefore does not have any classical electrostatic contribution, suggests that short-range effects should also be included. In that context, Piquemal *et al*

Figure 1: Electrostatics contribution of rare gas dimers in kcal/mol



have proposed a damping correction[7,18] grounded in Quantum Chemistry (QC) (see equation 1). Indeed, the electrostatic energy between two monomers (A and B) can be expressed as :

$$E_{\text{elec}} = -2\sum_{i}\sum_{\nu} Z_\nu \int \frac{|\varphi_i(1)|^2}{r_{i\nu}} {dr_1}^3 - 2\sum_{j}\sum_{\mu} Z_\mu \int \frac{|\varphi_j(2)|^2}{r_{j\mu}} {dr_2}^3 + 4\sum_{i}\sum_{\text{J}} \int \frac{|\varphi_i(1)|^2|\varphi_j(2)|^2}{r_{12}} {dr_1}^3 {dr_2}^3 + \sum_{\mu}\sum_{\nu} \frac{Z_\mu Z_\nu}{r_{\mu\nu}} \tag{1}$$

where μ (ν) and φ_i (φ_j) are, respectively, the nucleus and the unperturbed molecular orbitals of monomer A (B) and Z is the atomic number. The first two terms define nucleus-electron

attraction, the third the electron-electron repulsion and the fourth the nucleus-nucleus repulsion. This strongly suggests that mimicking such a term is a key ingredient to correctly model all the components of electrostatics. The electronic integrals explain the appearance of electrostatic exponential decay. As charge penetration is missing, it seems reasonable to consider the outer-shell electrons (electronic valence) in the construction of an overlap-based short-range correction. Indeed, as shown in Figure 1 the atomic size is important, and is closely linked to the magnitude of the penetration effect. One would like to consider van der Waals-like radii as a component of the parametrization. In that context,[18] Piquemal *et al* proposed a short-range correction to the multipolar electrostatics that includes all these features and was originally implemented in SIBFA. The latter modifies the charge-charge, charge-dipole and charge-quadrupole interactions.

The aim of this paper is to include and to reparametrize this correction in the framework of the Smooth Particle Mesh Ewald. The proposed correction modifies the charge-charge (2) and charge-dipole (3). The latter involves an damped charge that is also appicable to correct the charge-quadrupole interactions.[18] These three terms are the dominant ones.[7]

$$\begin{aligned} E^*_{m_i-m_j} =& \frac{1}{r_{ij}}[Z_iZ_j - Z_i(Z_j - q_j)(1 - exp(-\alpha_j r_{ij})) \\ & - Z_j(Z_i - q_i)(1 - exp(-\alpha_i r_{ij})) \\ & + (Z_i - q_i)(Z_j - q_j)(1 - exp(-\beta_i r_{ij})(1 - exp(-\beta_j r_{ij})] \end{aligned} \tag{2}$$

where $\alpha_i = \gamma/r^i_{vdw}$ and $\beta_i = \delta/r^i_{vdw}$, γ and δ are normalization coefficients and r^i_{vdw} the van der Waals radius of atom i. Z_i is the electronic valence of the atom i, q_i the charge (monopole) of atom i.

$$E^*_{m_i-d_j} = \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r^3_{ij}}(Z_i - (Z_i - q_i)(1 - exp(-\eta_i r_{ij})) \tag{3}$$

where $\eta_i = \chi/r^i_{vdw}$, with χ a normalization coefficient.

This charge-charge correction contains the traditional coulombic expression, corresponding to nucleus-nucleus repulsion. The formulation considers two more contributions, nucleus-electron attraction and electron-electron repulsion, where the damping function appears. It describes the depletion of the electronic cloud at short-distance. Following the same idea, the charge-dipole also accounts for this depletion by modulating the charge.[20] It is worth mentioning that at long-range, this term tends to the classical coulombic electrostatic expression. In Molecular Dynamics (MD), conformational exploration of space can lead to very short-interatomic distances where multipoles clearly fail to recover the QM values due to the lack of charge penetration effect. Usually such errors are compensated by parametrization of the van der Waals terms. However, there is no guarantee of a systematic compensation during such MD simulations, the lack of which could lead to errors. In order to apply this methodology to large systems and perform MD, we propose to couple this approach to the Smooth Particle Mesh Ewald[12] (SPME).

2.1 Smooth Particle Mesh Ewald

SPME is based on the Ewald summation[35] which suggests to split the electrostatics into two terms, a short-range E_{dir} and a long-range one E_{rec} and whose complexity is N^2 or even $N^{3/2}$.[36] SPME decreases it to $N \log N$ by using Fast Fourier Transforms (FFT).[35]

SPME can be applied to condensed phase. In that case, the lower complexity leads to particularly large time savings for large systems in Molecular Dynamics. SPME is an efficient and adaptable algorithm, as its formulation for multipoles[37,38] and for the charge penetration correction suggests. In this section, we show how to modify the general SPME formulas in order to include the charge penetration correction.

Let U be a neutral unit cell with N particles in position $(\mathbf{r}_i)_{i\in\{1,\cdots,N\}}$ and respective multipolar moments $(\hat{L}_i)_{i\in\{1,\cdots,N\}}$, $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ its basis and Σ the system made of an infinite number of images of the unit cell.

For any integers n_1, n_2, n_3, m_1, m_2 and m_3 we will note

$$\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad \mathbf{m} = m_1\mathbf{a}_1^* + m_2\mathbf{a}_2^* + m_3\mathbf{a}_3^*,$$

where $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ is the dual basis of $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ (*i.e.* $\mathbf{a}_i^* \cdot \mathbf{a}_j = \delta_{ij}$).

The electrostatic energy of the system is

$$E_{\text{elec}}(\Sigma) = \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} \hat{L}_i\hat{L}_j \left(\frac{1}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|}\right), \tag{4}$$

where the prime means that the terms $i = j$ are not summed up when $\mathbf{n} = \mathbf{0}$. If we consider multipoles up to quadrupoles, then the operators $\hat{L}_i$ can be written as

$$\hat{L}_i(\mathbf{m}) = q_i + \boldsymbol{\mu}_i \cdot \nabla_i + \boldsymbol{Q}_i : \nabla_i\nabla_i, \tag{5}$$

where ∇_i is the gradient at $\mathbf{r}_i$, $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ are the dipoles and the quadrupole moments.

For any $\mathbf{m} \in \mathbb{R}^3$ we define the structure factor

$$S(\mathbf{m}) = \sum_j \tilde{L}_j(\mathbf{m}) \exp(2\pi\mathbf{m} \cdot \mathbf{r}_j) \tag{6}$$

where $\tilde{L}_j$ is the Fourier transform of the operator $\hat{L}_j$, *i.e.*

$$\tilde{L}_j = q_j + 2i\pi\boldsymbol{\mu}_j \cdot \mathbf{m} - (2\pi)^2\boldsymbol{Q}_j : \mathbf{m}\mathbf{m}, \tag{7}$$

If Σ is supposed to be surrounded by a medium with infinite dielectric constant, then it has been shown that for any $\beta > 0$ the sum can be split as[38]

$$E_{\text{elec}}(\Sigma) = E_{\text{dir}} + E_{\text{self}} + E_{\text{rec}}, \tag{8}$$

where

$$E_{\text{dir}} = \sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} \hat{L}_i \hat{L}_j \left(\frac{\operatorname{erfc}\left(\beta|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|\right)}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|} \right), \tag{9}$$

$$E_{\text{self}} = -\frac{\beta}{\sqrt{\pi}} \sum_{j=1}^{N} \left(q_j^2 + \frac{2\beta^2}{3} \left(\boldsymbol{\mu}_j^2 - 2q_j \operatorname{Tr}(\boldsymbol{Q}_j) \right) + \frac{4\beta^4}{5} \left(2\boldsymbol{Q}_j \colon \boldsymbol{Q}_j + \operatorname{Tr}^2(\boldsymbol{Q}_j) \right) \right), \tag{10}$$

and (11)

$$E_{\text{rec}} = \frac{1}{2\pi V} \sum_{\mathbf{m}\neq 0} \frac{\exp(-\pi^2 \mathbf{m}^2/\beta^2)}{\mathbf{m}^2} S(\mathbf{m}) S(-\mathbf{m}). \tag{12}$$

The self energy term[39] E_{self} is a bias due to the Ewald summation and does not have any physical meaning. Both the direct and the reciprocal sums converge rapidly. A 9 Å cutoff is common, allowing a $O(N)$ computational complexity for the direct sum. The B-spline approximation of the structure factor on a grid of dimensions $K_1 \times K_2 \times K_3$ enables the reciprocal energy to be approximated by

$$\tilde{E}_{\text{rec}} = \frac{1}{2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \sum_{k_3=0}^{K_3-1} Q(k_1,k_2,k_3) \cdot (\theta_{\text{rec}} * Q)(k_1,k_2,k_3), \tag{13}$$

where Q is a derivable function obtained after interpolation of the structure factors at the nodes of the grid; and θ_{rec} a pair-potential function independent of particle position. The regularity of Q implies that the force in position i can be approximated by

$$\tilde{\mathbf{F}}_i = -\nabla_i \tilde{E}_r = \frac{1}{2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \sum_{k_3=0}^{K_3-1} \nabla_i Q(k_1,k_2,k_3) \cdot (\theta_{\text{rec}} * Q)(k_1,k_2,k_3). \tag{14}$$

In principle, the short-range correction is only in the direct space and therefore we will show that no modification of the reciprocal sum is required. We will detail our strategy of implementation aimed at an optimal scalability of the corrected electrostatics. As we discussed, the presented implementation in SPME requiries a modification of the direct sum only. To do so, we need to reexpress the damping function by Piquemal et al. in more

convenient form as a simple correction to the direct sum.

2.2 Coupling Smooth Particle Mesh Ewald and a penetration correction

Let us define E^*_{elec} the total modified/damped electrostatic energy of the system Σ.

$$E^*_{\text{elec}}(\Sigma) = \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} E^*_{m_i-m_j} + E^*_{m_i-d_j} + E^*_{m_i-q_j} + E_{d_i-d_j} + E_{d_i-q_j} + E_{q_i-q_j} \tag{15}$$

where $E_{d_i-d_j}$ is the interaction between dipoles of site i and site j ; $E_{m_i-q_j}$, $E_{d_i-q_j}$ and $E_{q_i-q_j}$ are respectively the interaction between monopole/dipole/quadrupole of site i and quadrupole of site j. Let us define E_{pen} the total charge penetration energy,

$$E_{pen} = E^*_{\text{elec}} - E_{\text{elec}} = \sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} E^*_{m_i-m_j} - E_{m_i-m_j} + E^*_{m_i-d_j} - E_{m_i-d_j} \tag{16}$$

Let us define the undamped electrostatic interaction between the monopoles of sites i and j $E_{m_i-m_j}$, let us also define the undamped electrostatic interaction between the dipole of site j and the monopole of site i $E_{m_i-d_j}$.

$$\begin{aligned} E^*_{m_i-m_j} - E_{m_i-m_j} &= (q_i - Z_i)(q_i - Z_j)(\exp(-\beta_i r_{ij}) - 1)(\exp(-\beta_j r_{ij}) - 1) \\ &- (q_j - Z_i)Z_j(\exp(-\alpha_i r_{ij}) - 1) - (q_i - Z_j)Z_i(\exp(-\alpha_j r_{ij}) - 1) - q_i q_j + Z_i Z_j \end{aligned} \tag{17}$$

$$E^*_{m_i-d_j} - E_{m_i-d_j} = \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r^3_{ij}}(Z_i - q_i)exp(-\eta r_{ij}) \tag{18}$$

$$E^*_{\text{elec}} = E_{\text{elec}} + E_{pen} \tag{19}$$

The same strategy has been followed to correct the charge-quadrupole interaction.[18]

As seen in section 2.1, E_{pen} can not be simply included into the SPME framework, as we

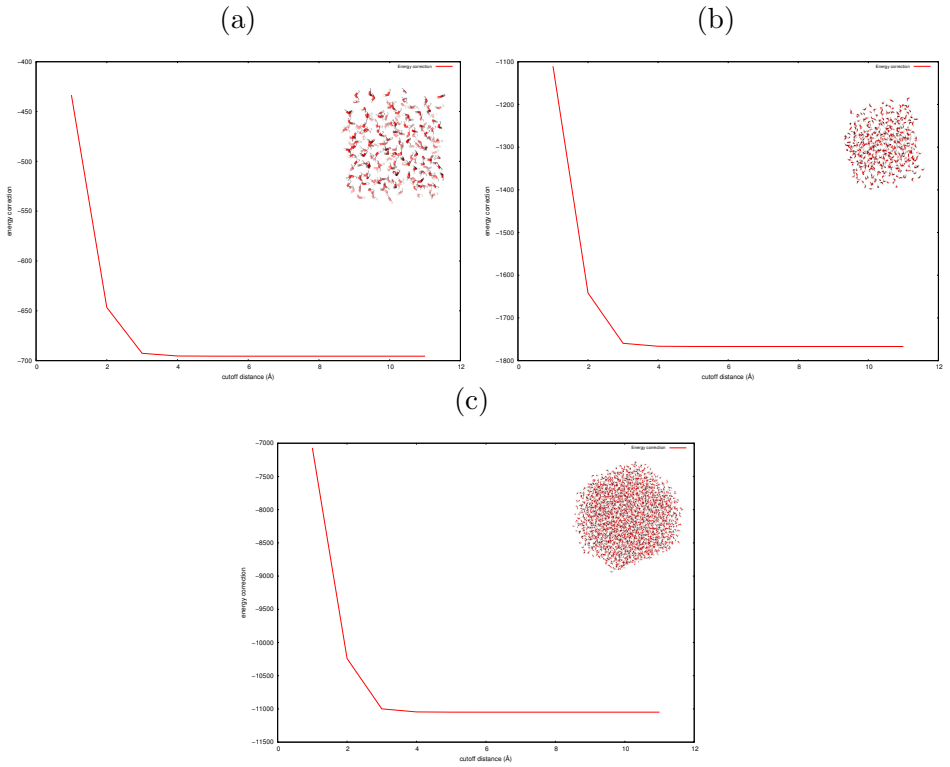
should take a large cutoff for the direct sum to avoid any persistence of penetration outside the direct sum treatment.

Indeed, the charge penetration energy is prominent at short-range and tends quickly to zero at long-range because of its $\exp(-x)/x^n$ (with $n = 1, 3$) form (see figure 4) but not necessarily fast enough to be maintained within the Ewald direct sum cutoff. As such Ewald cutoff is essential for scalability, one would like to separately handle SPME and penetration issues. This justifies the choice of an additional cutoff method dedicated to penetration. Indeed, for a given precision, one can always find a cutoff for penetration that guarantees an error inferior to this threshold. As such an approach generates a discontinuity and non-derivability in the potential function near the cutoff limit. We will couple it with a switching function that ensures derivability of the energy during a MD. Many smoothing functions have been proposed[40] and we chose to use the one defined in Ref. 41 and 42 (equation (20)). Hence, for every distance r, the potential energy is equal to $V(r) \cdot S(r)$ where $V(r)$ is the unswitched function. The interactions are not modified for distances less than an inner cutoff distance R_{in} and are smoothed to zero at the outer cutoff R_{out}. Besides, the continuity of the first derivatives prevents jumps in the energy during MD or minimization, which could lead to non-physical phenomena such as heating.

Such a technique enables to keep the long-range effects and the computational advantage of SPME using FFTs with an optimal Ewald cutoff, while just adding a computationally cheap correction term with its own cutoff (see Figure 2). Besides, numerical simulations of common systems (up to tens of thousands of atoms) show that a cutoff between 6 and 8 Å guarantees an error inferior to 10^{-3} kcal/mol (see Figure 2). Of course, the smoothing function makes possible to choose a more aggressive cutoff of 4 Å(see Figure 2) which will capture most of the penetration interaction in common systems leading to significant computational speed-up. Overall the final cost in term of computational ressources of the penetration correction is inferior to 10 per cent of the total penetration corrected multipolar electrostatics.

$$S(r) = \begin{cases} 1 & r \leq R_{in} \\ \frac{(R_{out}^2 - r^2)^2 (R_{out}^2 + 2r^2 - 3R_{in}^2)}{(R_{out}^2 - R_{in}^2)3} & R_{in} < r \leq R_{out} \\ 0 & r > R_{out} \end{cases} \tag{20}$$

Figure 2: Convergence cutoff study of penetration energy. Energy in kcal/mol.



3 Results

3.1 Parametrization

The electrostatic correction includes four empirical parameters: two for charge-charge interactions (core-electron attraction and electron-electron repulsion), one for charge-dipole interactions, and one for charge-quadrupole. Those were initially adjusted on water dimers.[43–46] As various parametrization strategies are possible, the atomic radii can be used as a pa-

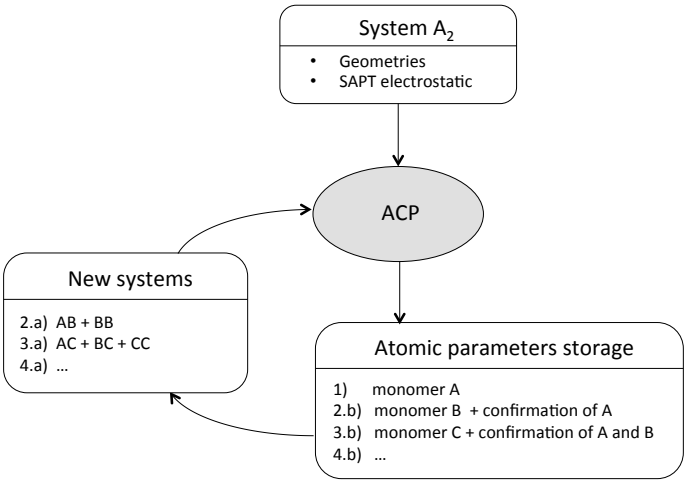
rameter in order to reproduce the charge penetration effect. We apply this methodology to various choices of distributed multipoles. We focused on the GDMA approach by Stone[29] and on the electrostatic potential (ESP)[?] ren2011polarizable fitted multipoles used in the AMOEBA force field. We first follow the initial parametrization strategy developed for SIBFA and then we choose a simpler exponential functional form for the parametrization that does not take into account any atom radii. In this form, the parameters are the previously defined exponents [18] : α_i, β_i and η_i.

We focused our parametrization on a well established panel of intermolecular systems extracted from the S66 dataset as well as on selected configurations of water dimers obtained from high level ab-initio computation[44,47,48] . Several QM intermolecular energy decomposition techniques are now available such as reduced variational space[49] (RVS) or SAPT.[34] SAPT offers the possibility to include all types of intermolecular interactions, as it allows to take into account electronic correlation. Because our upcoming implementation should reproduce all of them in a new anisotropic polarizable force field based on *ab-initio* calculation, we chose to calibrate the electrostatics on SAPT. Note that the idea is to reproduce electrostatic trends at all ranges, with a special focus at short-ranges where charge penetration matters, which could make a perfect match with QM reference values difficult.

The first step of our calibration was to find the best set of parameters for ten different water dimers associated to stationary point on the ab-initio potential energy surface (denoted A on Figure 3). Four of these dimers were chosen (see Figure 4) to refine our results because they represent different chemical situations. Next, a new dimer (denoted B on Figure 3) as well as monomer-water systems (step 2 on Figure 3) can be calibrated. Thus, new parameters can be tested on dimers involving non-calibrated monomers (denoted C on Figure 3). This also allows to validate the previous calibration.

Each system was considered at its lower energy value configuration and at different intermolecular distances. These kind of variation are likely to happen in MD and accounting for them will make our model more robust, as it would be able to sample density overlapping

Figure 3: Flowchart Automated Calibration Procedure (ACP)



regions.

3.2 Computational Details

The parallel implementation was first tested on the SGI UV 2000 supercomputer of the ICS (Institut du Calcul et de la Simulation) which is made of 64 nodes each one of them having two Intel Xeon E5-4650L CPUs with eight cores at 2.6 Ghz and 256 GB of DDR3 DIMM RAM. The nodes are able to communicate within a global shared memory system thanks to the SGI NUMALINK interconnect. We also tested our MPI implementation on the Stampede supercomputer of the TACC (Texas Advanced Computing Center) whose architecture is more standard as it consists, for the part we did our tests on, in 6400 compute nodes with two intel Xeon E5-2680 CPUs with eight cores at 2.7Ghz, one Intel Xeon Phi SE10P coprocessor and 32 GB of DDR3 DIMM RAM. These nodes communicate within a 56 GB/s InfiniBand network. The following numerical results are all based on computations made on this supercomputer. SAPT calculations have been performed using PSI4[50] at third order for water dimers and

second for the others, using aug-cc-pTZ basis sets. All GDMA multipoles were computed from density at the MP2(full)/aug-cc-pTZ.

3.3 Numerical Results

Table 1: Comparison of the pure multipolar electrostatic energy (in kcal/mol) water dimers. GDMA^i describes the multipolar sites scheme, $i = 1$ atoms only and $i = 2$ atoms and bonds.

index	GDMA^2	GDMA^1	ESP	SAPT
1	-6.91	-6.73	-6.04	-7.96
2	-5.22	-5.15	-5.28	-6.72
3	-4.88	-4.85	-5.22	-6.52
4	-4.72	-4.82	-4.63	-6.61
5	-3.55	-3.76	-4.26	-5.71
6	-3.19	-3.44	-4.20	-5.41
7	-3.52	-3.58	-3.15	-4.84
8	-1.28	-1.26	-0.94	-1.49
9	-3.87	-3.79	-3.15	-4.71
10	-2.24	-2.21	-2.03	-2.71

Table 2: Comparison of the penetration corrected electrostatics (in kcal/mol) water dimers. GDMA^i describes the multipolar sites scheme, $i = 1$ atoms only and $i = 2$ atoms and bonds.

index	cp-GDMA^2	cp-GDMA^1	cp-ESP	Ref. 25	SAPT
1	-8.96	-8.18	-7.87	-6.9	-7.96
2	-7.32	-6.64	-6.94	-6.3	-6.72
3	-7.03	-6.36	-6.86	-6.5	-6.52
4	-6.48	-6.58	-6.10	-6.3	-6.61
5	-5.61	-5.78	-5.65	-6.0	-5.71
6	-5.40	-5.58	-5.56	-6.2	-5.41
7	-4.38	-4.71	-3.96	-4.5	-4.84
8	-1.20	-1.43	-1.19	-1.1	-1.49
9	-5.02	-4.84	-3.90	-4.9	-4.71
10	-2.87	-2.82	-2.48	-3.6	-2.71

Tables 1 and 2 gather various results for ten selected orientation of the water dimer[44] including pure multipolar interactions (Table 1) and damped multipolar ones (Table 2). Such configurations are clearly difficult as they embody various types of orientations involving very

different anisotropic orbital interactions. Columns 2 to 4 of table 1 describe the values of the GDMA, atoms and bond midpoints then atoms only and ESP fitted multipoles energies. In the first approximation, compared to reference SAPT and to atoms only GDMA multipoles, one would think that having more centers improve electrostatics. Overall, the ESP fit strategy seems to improve the agreement with SAPT results. The addition of more centers has been already described in the literature by Claverie and Stone[16,29] as a way to capture higher order effects, atoms plus bond midpoints being equivalent to the addition of octupoles within a distributed multipoles analysis. Such an observation might need, however, to be reconsidered. Indeed, the derivation of multipoles is usually said to be non-empirical, however, choices exist within the GDMA algorithm including the choice of the partition approach, the choices of the hydrogen radius etc. Here we tried various settings and kept the following values for the GDMA approach (space partitioning : switching 2, Hydrogen radius 0,325 Å). Choices also exist for the ESP fitted multipoles as, within the procedure, the charge is initially derived from GDMA and is kept frozen whereas dipole and quadrupole moments are reoptimized to minimize the differences with respect to the ab initio electrostatic potential. Overall, with all possible choices, a continuum of multipoles set can be derived with strong energetical differences. Which one should be chosen and are the ESP fitted multipoles really better. The application of the damping function to distributed multipoles casts a new light on such an issue. Clearly, the atoms plus bond midpoints GDMA and ESP fitted multipoles corrected results (see table 2) are improved by damping but surprisingly the best agreement is found for the damping of atoms only GDMA multipoles. A clear-cut explanation can be found by a careful analysis of the multipolar values alone (table 1) compared to SAPT. Both GDMA (atoms plus midpoints) and ESP multipoles energies are too high. Indeed, even, with the best choice of settings, some GDMA configurations energies are almost equal the full quantum SAPT energies that embodies penetration and therefore that should be larger. In a sense, the apparently good behavior of the multipolar energies is artificial and is a consequence of a spurious fit that only minimizes the ESP potential without providing

an adequate physical picture of the distribution of moments on centers. It is of course, exacerbated by the use of damping functions that add another correction to GDMA analysis. In other words, if the multipolar energy of a given configuration is, by chance, spuriously correct, the damping correction will only destroy the fictitious agreement. For large basis sets, the advantage of having a larger number of centers tend to generate lower quality distributed multipoles because of the difficult handling of diffuse function.[29] Damping indeed requires physically meaningful multipoles with a consistent representation of the different angular momenta. For ESP fitted multipoles, no charge redistribution is provided for the reoptimization of the dipole and quadrupole moments. It generates a better fit for some configurations but also partially reduces anisotropy. Indeed, moments are related to atomic orbital symmetries. If one component is overfitted, the others are clearly less accurately or spuriously described. Overall, whatever, the strategy to parametrize the damping function, one has to start to analyze the multipolar energy behavior compared to SAPT: both sets of energies should be clearly different as multipoles do not embody penetration energy. Any attempt to damp overfitted energies will lead to a loss of accuracy when damping is added, especially if one studies various anisotropic interactions. If well-balanced multipoles are extracted, the damping is then increasingly effective as the order of correction tend towards charge-quadrupole (see appendix). If overfitted multipoles are chosen, one can still strongly reduce the error with SAPT but a charge-charge correction should be prefered as higher-order corrections will be perturbed by the spurious anisotropy of the charge distribution. It may lead to double countings of the penetration effect for configurations where multipoles are already overestimated. Overall, on the 10 dimer configurations, the best choice is based on GDMA multipoles up to quadrupole fitted on atoms only and corrected by a damping function going up to the quadrupole level.

Besides the application to a SIBFA-like protocol, we also applied this methodology to treat charge penetration within AMOEBA framework, which use only atom centered ESP multipoles on various systems. As we discussed, ESP multipoles do take not full advantages of higher

order penetration correction as the charge-dipole damping appears already weaker than in the case of well behaved multipoles.

In all these cases, we have significantly improved the AMOEBA results without modifying the existing sets of multipoles (amoeba09).

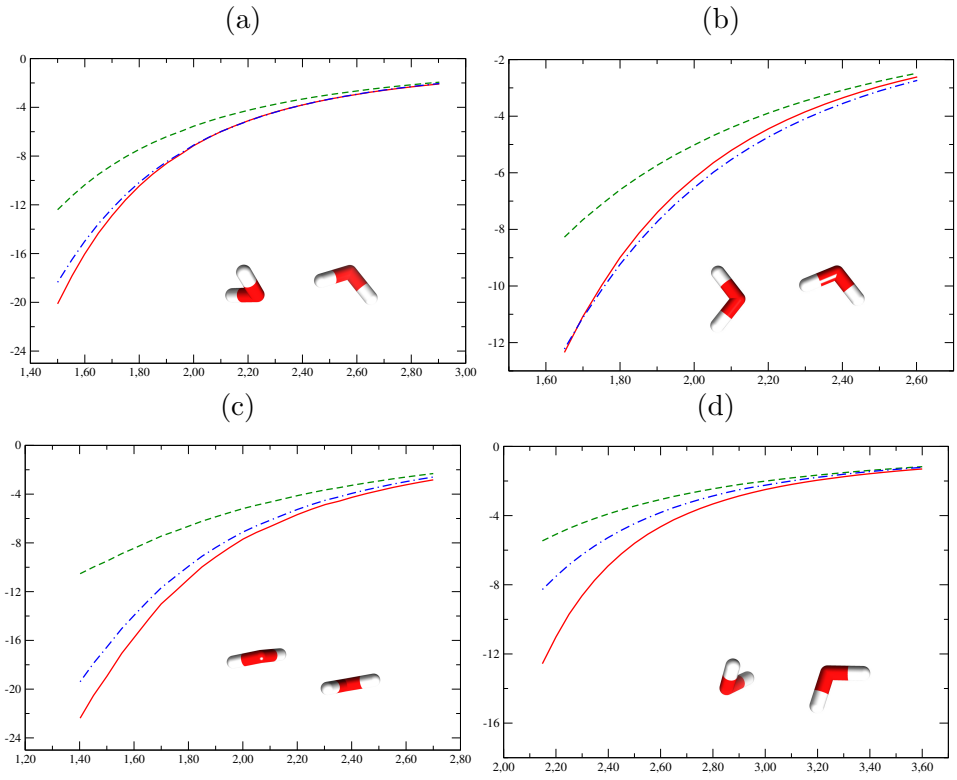
For the previously discussed water dimers, the maximum energy variation between the first six configurations, corresponding to the lowest energies, the charge-charge correction reproduced well the behavior of SAPT energy variations between configurations (energy difference between the most and the less stable configuration : 2.5 for SAPT) whereas other corrections appear strongly less anisotropic (0.9 kcal/mol for Ref. 25). Even for this simplest level of correction, penetration effects are well recovered. Indeed, figure 4 represents the electrostatic energy between two water molecules as a function of their intermolecular distance in Angströms. The first two geometries (fig.4a and 4b) correspond to strong H-bonds while the third (fig. 4c) involves two H-bonds and the fourth (fig. 4d) weak interactions.

Table 3: RMSE for distance scans of water dimers (see text) in kcal/mol (reference SAPT)

water dimer	CP-AMOEBA	AMOEBA
1	0.51	2.81
3	0.25	1.66
4	1.22	4.86
7	1.39	1.88
S66	0.69	2,21

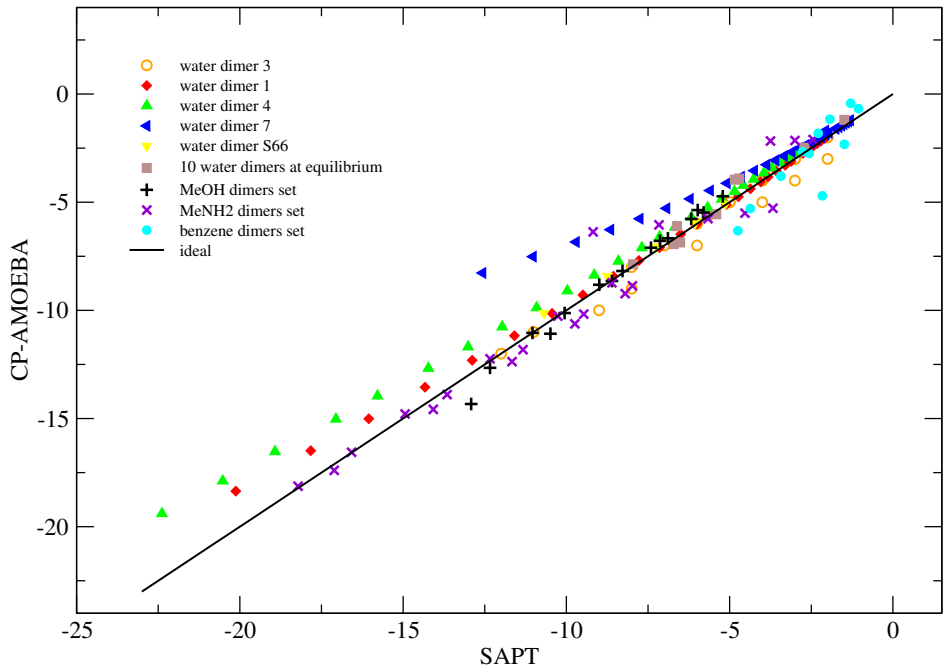
We also report the root mean square error for intermolecular distance scans (from 1.4 to 2.7 A) of selected water dimers (S66 is an additional water geometry extracted from S66 set) in Table 3 which confirms the improvement of the AMOEBA model in different configurations and the robustness of the correction at very short-range. The small improvement between the RMSE's regarding water dimer 7 (Figure 4d) could be explained by its being a high lying energy dimer obtained only at the CCSDT level.[43] Indeed, for this kind of system, the charge penetration correction remains subtle because of the importance of electronic correlation and of the difficulty to model interaction involving the use of very diffuse basis function.

Figure 4: Water dimers electrostatic profile as function of intermolecular distance in kcal/mol. SAPT is the red line. AMOEBA the green dashed line. cp-AMOEBA the blue dashed dot line.



We continue our investigation on other chemical functions: methanol (MeOH) and methyl amine (MeNH2) (see appendix Tables 5 and 6 respectively). Finally, we investigate again benzene (see appendix table 7, 8 and 9). In all these cases, we improved the electrostatics using AMOEBA multipoles and particularly for benzene dimers. The approximation only breaks at very -unphysical- short-range. We report all our results in Figure 4, a plot of the results obtained with the correction against SAPT reference values. This graph includes all the presented molecules and their distance scans from near equilibrium up to short-distances.

Figure 5: Correlation between SAPT and the penetration correction in kcal/mol within binding range for various dimers (see legend)



The electrostatic interactions energies of some of these complexes are for distances far below equilibrium and for these penetration effects can be very significant. As one can see from tables (appendix), the penetration correction always improved the results compared to SAPT, even for the cases where ESP multipolar energies were really off (see benzene results), with the wrong interaction sign in all cases.

To conclude, we extended our tests to charge system, studying complexes of metal cations

with water including K^+, Na^+, Mg^{2+},Ca^{2+} and Zn^{2+}.

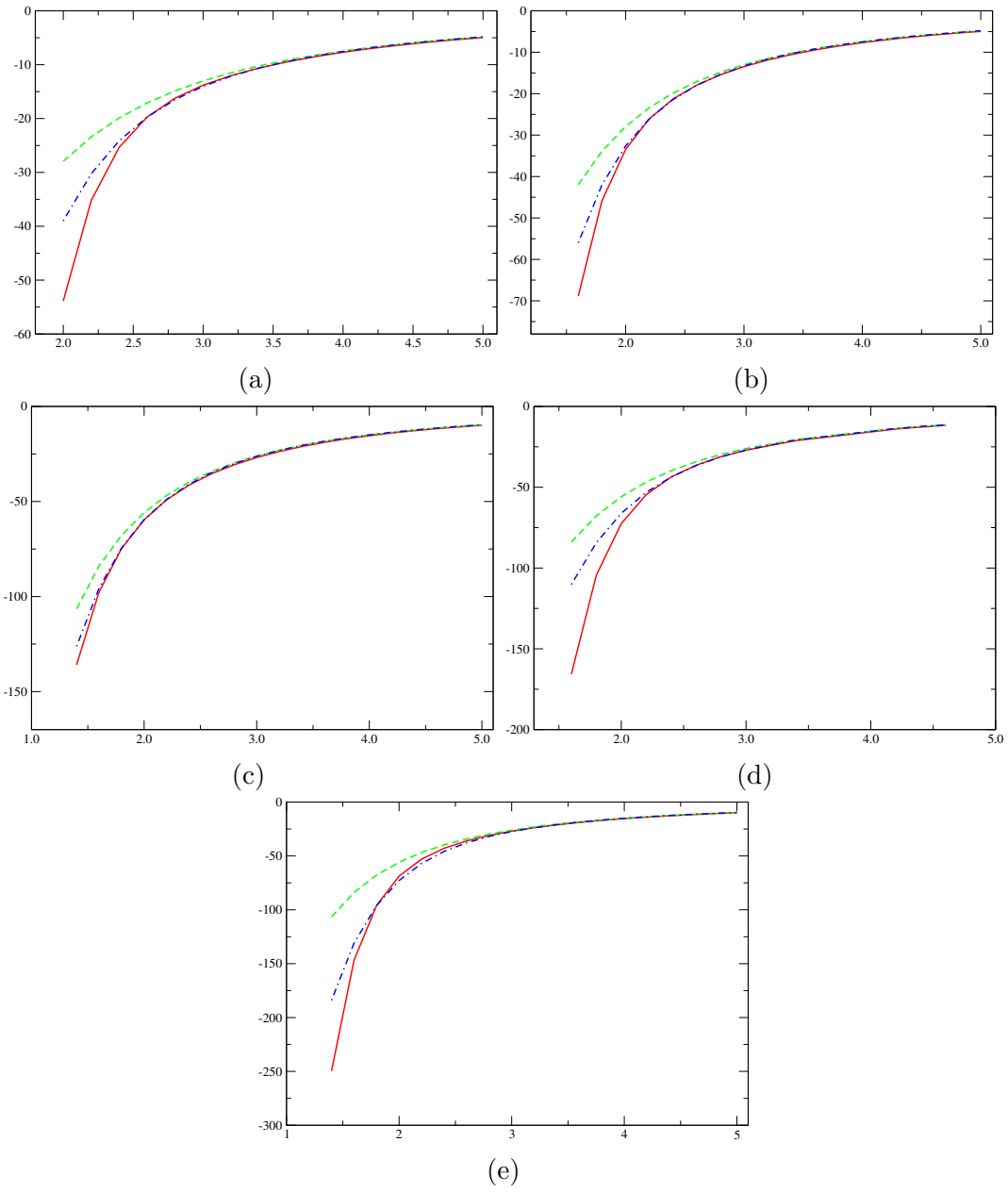


Figure 6: Cation-water complexes : electrostatic variation as function of cation-O distance in kcal/mol. SAPT is the red line, AMOEBA the green dashed line and the penetration correction blue dashed-dot line. a) K^+ water complex. b) Na^+ water complex. c) Mg^{2+} water complex. d) Ca^{2+} water complex. e) Zn^{2+} water complex.

As expected from the SIBFA previous work, penetration effects are a key element to model metals. The simple addition of the charge-charge correction allows a very close agreement with reference SAPT datas.

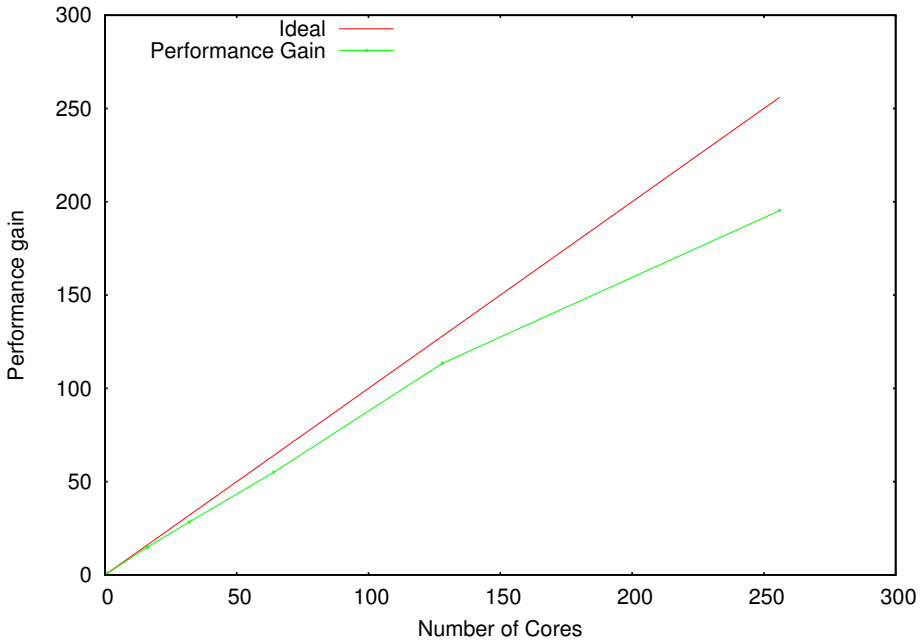
3.3.1 Parallel Implementation

The loops involved in the computation of the correction term can be easily parallelized with a shared memory (OpenMP), distributed memory (MPI) or hybrid (OpenMP/MPI) paradigm. The additional routines are included in a parallel version of the SPME that we developed in the upcoming TINKER-HP software[51–54] : the atom sites and the corresponding energy and forces are distributed among processes using a spatial decomposition.[?] Furthermore, as the cutoff used in the correction term (between 6 and 8 Å) is usually smaller than the real space cutoff (see Figure 2) used in the SPME, no additional cross-processor communication of the atomic positions between each time step is necessary. The parallel scaling of the computation of the correction using MPI is showed on graph 7 for a water box of 60 000 atoms in a cubic unit cell where each edge is 84 Angströms long. Our tests showed that a pure MPI implementation is faster than a hybrid one before reaching a scaling plateau, but once this plateau is reached, the hybrid implementation allows to make use of more cores and to improve the best timing. The best timings obtained to compute both the electrostatic forces using SPME, and the correction, using the pure MPI implementation are shown in table 4, as well as the best timings obtained with the hybrid implementation (with 2 OpenMP threads per MPI process).

Table 4: Absolute best timings (in seconds) and number of cores used (in parentheses) for the computation of the electrostatic forces with SPME and of the correction in our implementation.

	Pure MPI	Hybrid OpenMP/MPI
SPME	0.070 (256)	0.040 (512)
CP-AMOEBA	0.025 (256)	0.015 (512)

Figure 7: Parallel scaling for a water box of 60 000 atoms



With a cutoff of 8 Å, which is beyond convergence, the total penetration cost is only 25 per cent in both cases. Of course, as we discussed, such a cutoff can be chosen more aggressi vely. A more aggressive cutoff of 4 angströms (see figure 2) captured 99 per cent of penetration interaction and led to strong decrease of the computation cost which became lower than 10 per cent.

4 Conclusions

We have studied the general coupling of an empirical short-range charge penetration correction to the SPME. The corrected SPME approach allows a computationally efficient and accurate improvement of electrostatics in polarizable simulations by correcting the short-range charge-charge, charge-dipole and charge-quadrupole terms. The method has been implemented in the TINKER-HP software and the additional cost of the short-range damping function shown to be small as the global electrostatics implementation exhibits a favorable parallel scaling at the MPI and OPEN-MP/MPI hybrid levels. A separate cutoff for the pen-

etration term coupled to a switching function has been implemented in order to maintain the possibility of a performance optimization of SPME thanks to the separate Ewald cutoff. Initial tests have been performed in the context of the AMOEBA and SIBFA force fields (ESP and GDMA approaches respectively). Substantial improvements were noted in the agreement of standard non-corrected multipolar interaction energies (from Stone multipoles[29]) compared to SAPT ab-initio reference data. As we discussed the effects of spurious fit of distributed multipoles, the need of comparing atom-centered GDMA to SAPT is highlighted as one wants to avoid overfitted multipolar energies. When distributed moments are physically meaningful, the correction exhibits a systematic improvement when adding charge-charge, then charge-dipole and charge-quadrupole corrections. For some sets of multipoles, like the ESP one, that exhibit some fictitious inclusion of penetration the correction should be limited to the charge-charge correction only. For these types of multipoles, specific parametrization procedure has been recently proposed in reference[55] based on ESP derivation of the damping parameters.

Finally it is important to point out that, at the end, an error compensation between the damping function and the set of multipoles has to be found. Others damping function exist and will be studied in order to find the best agreement according to the different force field models (Rackers *et al*, in preparation).

This opens two possible strategies as one can focus on the sole optimization of the charge-charge term to obtain a significant improvement of atom centered multipolar electrostatics at a very low computational cost. Strategies in that direction are currently studied in the context of AMOEBA-like potentials.

A second strategy includes higher order damping corrections (charge dipole and beyond), as presented in the SIBFA potential. Such an implementation is therefore the starting point of SIBFA portage into TINKER-HP. Prospective applications in biochemistry, pharmacology and nuclear science[22,56–58] should benefit from the inclusion of charge penetration in PBC molecular dynamics with improved representation of short-range contributions exchange-

repulsion and charge transfer and long-range polarization and dispersion. To conclude, it is important to point out that future inclusion of such correction in polarizable molecular dynamics will require either to reparametrize the van der Waals contribution for the case of AMOEBA-like potentials or to use the penetration-corrected electrostatics in conjunction with an explicit exchange-repulsion contribution like in SIBFA.

Acknowledgements

We wish to thanks Marie-France Couret for technical support. This work was also supported in part by French state funds managed by CALSIMLAB and the ANR within the Investissements d'Avenir program under reference ANR-11-IDEX-0004-02. Qiantao Wang, David R. Bell and Pengyu Ren are grateful for support by the Robert A. Welch Foundation (F-1691) and the National Institutes of Health (GM106137), as well as the high performance computing resources provided by TACC and XSEDE (TG-MCB100057). Funding from French CNRS through PICS grant between UPMC and UT Austin is acknowldeged. Jean-Philip Piquemal and Étienne Polack acknowledge funding from the Délégation Générale de l'Armement (DGA), Maîtrise NRBC, from the French Ministry of Defense.

References

(1) Allen, F. H. *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380–388.

(2) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer Jr, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Archives of biochemistry and biophysics* **1978**, *185*, 584–591.

(3) UniProt Consortium, *Nucleic acids research* **2014**, *42*, D191–D198.

(4) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular cell biology*; WH Freeman, 2000.

(5) Pratt, C. W.; Voet, D.; Voet, J. G. *Fundamentals of biochemistry: life at the molecular level*; John Wiley & Sons, New York, NY, 2005.

(6) Slipchenko, L. V.; Gordon, M. S. *Molecular Physics* **2009**, *107*, 999–1016.

(7) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *The Journal of Physical Chemistry A* **2003**, *107*, 10353–10359.

(8) Handley, C. M.; Popelier, P. L. *The Journal of Physical Chemistry A* **2010**, *114*, 3371–3383.

(9) Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. *Journal of chemical theory and computation* **2009**, *6*, 190–202.

(10) Chaudret, R.; Gresh, N.; Narth, C.; Lagardère, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J.-P. *The Journal of Physical Chemistry A* **2014**,

(11) Giese, T. J.; Chen, H.; Dissanayake, T.; Giambaşu, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T.-S.; Panteva, M. T.; Radak, B. K. *Journal of chemical theory and computation* **2013**, *9*, 1417–1427.

(12) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of chemical physics* **1995**, *103*, 8577–8593.

(13) Darden, T.; York, D.; Pedersen, L. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

(14) Ren, P.; Wu, C.; Ponder, J. W. *Journal of chemical theory and computation* **2011**, *7*, 3143–3161.

(15) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *Journal of chemical theory and computation* **2007**, *3*, 1960–1986.

(16) Gresh, N.; Claverie, P.; Pullman, A. *Theoretica chimica acta* **1984**, *66*, 1–20.

(17) Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. *Journal of computational chemistry* **2003**, *24*, 1963–1970.

(18) Piquemal, J.-P.; Chevreau, H.; Gresh, N. *Journal of Chemical Theory and Computation* **2007**, *3*, 824–837.

(19) Devereux, M.; Gresh, N.; Piquemal, J.-P.; Meuwly, M. *Journal of computational chemistry* **2014**, *35*, 1577–1591.

(20) Cisneros, G.; Tholander, S.; Parisel, O.; Darden, T.; Elking, D.; Perera, L.; Piquemal, J.-P. *International journal of quantum chemistry* **2008**, *108*, 1905–1912.

(21) Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. *The Journal of Physical Chemistry A* **2008**, *112*, 268–280.

(22) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *The Journal of chemical physics* **2006**, *124*, 104101.

(23) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *The Journal of Chemical Physics* **2000**, *112*, 7300–7306.

(24) Stone, A. J. *The Journal of Physical Chemistry A* **2011**, *115*, 7017–7027.

(25) Wang, B.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2014**,

(26) Momany, F. A. *The Journal of Physical Chemistry* **1978**, *82*, 592–601.

(27) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. *Chemical reviews* **2013**, *114*, 779–814.

(28) Wheatley, R. J.; Mitchell, J. B. *Journal of computational chemistry* **1994**, *15*, 1187–1198.

(29) Stone, A. *Chemical Physics Letters* **1981**, *83*, 233–239.

(30) Vigné-Maeder, F.; Claverie, P. *The Journal of chemical physics* **1988**, *88*, 4934–4948.

(31) Claverie, P. Contribution à l'étude des interactions intermoléculaires. Ph.D. thesis, UPMC, 1973.

(32) Hohenstein, E. G.; Duan, J.; Sherrill, C. D. *Journal of the American Chemical Society* **2011**, *133*, 13244–13247.

(33) Jeziorska, M.; Cencek, W.; Patkowski, K.; Jeziorski, B.; Szalewicz, K. *The Journal of chemical physics* **2007**, *127*, 124303–124303.

(34) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chemical Reviews* **1994**, *94*, 1887–1930.

(35) Ewald, P. P. *Annalen der Physik* **1921**, *369*, 253–287.

(36) Perram, J. W.; Petersen, H. G.; De Leeuw, S. W. *Molecular Physics* **1988**, *65*, 875–893.

(37) Sagui, C.; Pedersen, L. G.; Darden, T. A. *The Journal of chemical physics* **2004**, *120*, 73–87.

(38) Smith, E. R. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **1981**, *375*, 475–505.

(39) Smith, W. *INFORMATION NEWSLETTER FOR COMPUTER SIMULATION OF CONDENSED PHASES* **1998**, 15–25.

(40) Norberg, J.; Nilsson, L. *Biophysical journal* **2000**, *79*, 1537–1553.

(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of computational chemistry* **1983**, *4*, 187–217.

(42) Loncharich, R. J.; Brooks, B. R. *Proteins: Structure, Function, and Bioinformatics* **1989**, *6*, 32–45.

(43) Reinhardt, P.; Piquemal, J.-P. *International Journal of Quantum Chemistry* **2009**, *109*, 3259–3267.

(44) van Duijneveldt-van de Rijdt, J.; Mooij, W.; van Duijneveldt, F. *Physical Chemistry Chemical Physics* **2003**, *5*, 1169–1180.

(45) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer III, H. F.; Radom, L. *The Journal of chemical physics* **1990**, *92*, 1240–1247.

(46) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer III, H. F.; Quack, M. *The Journal of chemical physics* **2002**, *116*, 690–701.

(47) Rezác, J.; Riley, K. E.; Hobza, P. *Journal of chemical theory and computation* **2011**, *7*, 2427–2438.

(48) Goerigk, L.; Kruse, H.; Grimme, S. *ChemPhysChem* **2011**, *12*, 3421–3433.

(49) Stevens, W. J.; Fink, W. H. *Chemical physics letters* **1987**, *139*, 15–22.

(50) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 556–565.

(51) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. *Journal of Chemical Theory and Computation* **2013**, *9*, 3637–3648.

(52) Lipparini, F.; Scalmani, G.; Lagardère, L.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. *The Journal of chemical physics* **2014**, *141*, 184108.

(53) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. *Journal of Chemical Theory and Computation* **2014**, *10*, 1638–1651.

(54) Lagardère, L.; Lipparini, F.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. **submitted**,

(55) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. *Journal of Chemical Theory and Computation* **2015**,

(56) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. *Journal of chemical theory and computation* **2010**, *6*, 2059–2070.

(57) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P. *Journal of chemical theory and computation* **2012**, *8*, 1314–1324.

(58) Marjolin, A.; Gourlaouen, C.; Clavaguéra, C.; Ren, P. Y.; Piquemal, J.-P.; Dognon, J.-P. *Journal of molecular modeling* **2014**, *20*, 1–7.

A Appendix

A.1 Forces

$$\begin{aligned}\frac{\partial(E^*_{m_i-m_j}-E_{m_i-m_j})}{\partial r_i^\alpha} =& \frac{(r_j-r_i)^\alpha}{r_{ij}^\alpha}(Z_i(Z_j-q_j)exp(-\alpha_j r_{ij})(\alpha_j+\frac{1}{r_{ij}}) \\ &+ Z_j(Z_i-q_i)exp(-\alpha_i r_{ij})(\alpha_i+\frac{1}{r_{ij}}) \\ &+ (Z_i-q_i)(Z_j-q_j)(exp(-(\beta_i+\beta_j)r_{ij})(\beta_i+\beta_j+\frac{1}{r_{ij}}) \\ &- exp(-\beta_i r_{ij})(\beta_i+\frac{1}{r_{ij}}) - exp(-\beta_j r_{ij})(\beta_j+\frac{1}{r_{ij}}))\end{aligned} \tag{21}$$

$$\frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial r_i^\alpha} = \frac{(Z_i-q_i)exp(-\eta r_{ij})}{r_{ij}^3}(\mu_j^\alpha + (r_j-r_i)^\alpha \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r_{ij}}(\eta+\frac{3}{r_{ij}})) \tag{22}$$

A.2 Torques

Because the energy correction term contains permanent dipole moments that are usually defined in local frames binded to neighboring atoms, the derivatives of this term with respect to the positions of these atoms have to be taken into account. These are often improperly refered to as 'torques'. Suppose that the k-th atom is used to define the local frame in which the dipole moment $\boldsymbol{\mu}_j$ is written, then the corresponding derivative is the following :

$$\frac{d(E^*_{m_i-d_j}-E_{m_i-d_j})}{dr_k^\alpha} = \frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial r_k^\alpha} + < \frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial \boldsymbol{\mu}_j}, \frac{\partial \boldsymbol{\mu}_j}{\partial r_k^\alpha} > \tag{23}$$

$$\frac{d(E^*_{m_i-d_j}-E_{m_i-d_j})}{dr_k^\alpha} = \frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial r_k^\alpha} + < (Z_i-q_i)exp(-\eta r_{ij})\frac{\boldsymbol{r}_{ji}}{r_{ij}^3}, \frac{\partial \boldsymbol{\mu}_j}{\partial r_k^\alpha} > \tag{24}$$

where the derivatives of $\boldsymbol{\mu}_j$ involve the derivatives of the derivatives of the rotation matrix used to go from the local to the global frame

A.3 Data

Tables 5 and 6 report electrostatics using two different distributed multipolar sites scheme : (i) atoms plus bonds (table 5). (ii) atoms only (table 6). Space Partitonning and hydrogen radii were also tested. sw$i$-rad$j$ denotes the space partitioning choice (switch i in GDMA) and j is the hydrogen radii.

Table 5: multipolar electrostatics using moments on bonds and atoms in kcal/mol

index	sw0-rad325	sw0-rad65	sw2-rad325	sw2-rad65	sw4-rad325	sw4-rad65
1	-6.18	-6.50	-6.91	-8.14	-6.96	-8.21
2	-5.19	-5.47	-5.22	-7.07	-5.23	-7.11
3	-5.05	-5.33	-4.88	-6.93	-4.88	-6.96
4	-5.13	-4.99	-4.72	-4.63	-4.69	-4.63
5	-4.51	-4.41	-3.55	-3.98	-3.48	-3.95
6	-4.32	-4.26	-3.19	-3.89	-3.11	-3.85
7	-3.59	-3.46	-3.51	-3.06	-3.50	-3.05
8	-1.04	-1.03	-1.27	-1.09	-1.28	-1.10
9	-3.32	-3.29	-3.87	-3.54	-3.89	-3.57
10	-1.96	-1.91	-2.24	-1.97	-2.25	-1.98

Table 6: multipolar electrostatics using moments on atoms only in kcal/mol

index	sw0-rad325	sw0-rad65	sw2-rad325	sw2-rad65	sw4-rad325	sw4-rad65
1	-6.02	-6.54	-6.73	-7.94	-6.75	-7.93
2	-5.12	-5.51	-5.15	-6.92	-5.16	-6.90
3	-5.01	-5.36	-4.85	-6.80	-4.86	-6.77
4	-5.31	-5.01	-4.82	-4.66	-4.81	-4.68
5	-4.81	-4.44	-3.76	-4.05	-3.74	-4.04
6	-4.67	-4.28	-3.44	-3.96	-3.42	-3.95
7	-3.49	-3.44	-3.58	-3.08	-3.58	-3.09
8	-0.93	-1.04	-1.26	-1.10	-1.26	-1.11
9	-3.06	-3.28	-3.79	-3.53	-3.80	-3.56
10	-1.87	-1.92	-2.21	-2.00	-2.21	-2.02

Table 7: multipolar electrostatics and penetration correction decomposed

SAPT	GDMA[1]	E_{pen}	E_{pen} charge-charge	E_{pen} charge-dipole	E_{pen} charge-quadrupole
-8.18	-6.73	-1.44	-1.22	-0.11	0.12
-6.64	-5.15	-1.48	-1.15	-0.14	0.20
-6.36	-4.85	-1.51	-1.14	-0.15	0.23
-6.58	-4.82	-1.76	-1.57	-0.03	0.15
-5.78	-3.76	-2.02	-1.65	-0.08	0.29
-5.58	-3.44	-2.14	-1.70	-0.10	0.34
-4.71	-3.58	-1.13	-1.13	0.03	0.03
-1.43	-1.26	-0.18	-0.25	0.04	-0.04
-4.84	-3.79	-1.04	-1.00	-0.03	0.02
-2.82	-2.21	-0.61	-0.58	-0.02	0.01

We have retained the S66[47] dimer notations. Tables reports electrostatic energy in kcal/mol.

Table 8: CP-amoeba parameters

fragments	atoms	α	β	η
water	H	3.5	2.9	1.7
	O	4.0	4.9	4.7
methanol	H(-H3C)	3.5	2.6	4.7
	C(-CH3)	5.8	4.9	2.5
	O	6.4	4.5	0.7
	H (-O)	4.8	2.9	1.7
methylamine	H(-H3C)	3.5	2.7	4.7
	C(-CH3)	4.8	5.4	2.5
	N	4.8	4.8	2.7
	H (-N)	2.5	2.9	1.7
benzene	C	2.8	2.5	2.8
	H	3.5	2.3	2.5

Table 9: Distance scans of S66 water orientations

dimers	SAPT	AMOEBA	CP-AMOEBA
2699_01WaterWater090	-10.687	-7.469	-10.136
2700_01WaterWater095	-8.739	-6.407	-8.423
2701_01WaterWater100	-7.231	-5.542	-7.058
2702_01WaterWater105	-6.052	-4.826	-5.961
2703_01WaterWater110	-5.124	-4.230	-5.078

Table 10: Distance scans of various methanol dimers orientations

dimers	SAPT	AMOEBA	CP-AMOEBA
2707_02WaterMeOH090	-12.337	-7.530	-12.656
2708_02WaterMeOH095	-10.053	-6.282	-10.124
2709_02WaterMeOH100	-8.280	-5.274	-8.180
2710_02WaterMeOH105	-6.891	-4.447	-6.667
2711_02WaterMeOH110	-5.799	-3.764	-5.480
2731_05MeOHMeOH090	-12.915	-6.222	-14.330
2732_05MeOHMeOH095	-10.488	-5.047	-11.082
2733_05MeOHMeOH100	-8.605	-4.096	-8.644
2734_05MeOHMeOH105	-7.133	-3.314	-6.788
2735_05MeOHMeOH110	-5.979	-2.667	-5.364
2755_08MeOHWater090	-11.037	-6.1497	-11.042
2756_08MeOHWater095	-8.991	-5.163	-8.815
2757_08MeOHWater100	-7.410	-4.356	-7.102
2758_08MeOHWater105	-6.177	-3.686	-5.770
2759_08MeOHWater110	-5.211	-3.128	-4.728

Table 11: Distance scans of various methylamine dimers orientations

dimers	SAPT	AMOEBA	CP-AMOEBA
2715_03WaterMeNH2090	-16.578	-14.156	-16.557
2716_03WaterMeNH2095	-13.653	-12.178	-13.894
2717_03WaterMeNH2100	-11.328	-10.595	-1.817
2718_03WaterMeNH2105	-9.468	-9.311	-10.175
2719_03WaterMeNH2110	-7.978	-8.261	-8.867
2739_06MeOHMeNH2090	-18.216	-12.793	-18.129
2740_06MeOHMeNH2095	-14.938	-10.881	-14.790
2741_06MeOHMeNH2100	-12.337	-9.354	-12.240
2742_06MeOHMeNH2105	-10.262	-8.117	-10.267
2743_06MeOHMeNH2110	-8.605	-7.107	-8.727
2763_09MeNH2MeOH090	-6.139	-3.797	-2.151
2764_09MeNH2MeOH095	-4.758	-3.364	-2.177
2765_09MeNH2MeOH100	-3.748	-3.011	-2.172
2766_09MeNH2MeOH105	-3.002	-2.719	-2.145
2767_09MeNH2MeOH110	-2.444	-2.475	-2.101
2771_10MeNH2MeNH2090	-9.181	-7.864	-6.372
2772_10MeNH2MeNH2095	-7.163	-7.199	-6.048
2773_10MeNH2MeNH2100	-5.661	-6.655	-5.761
2774_10MeNH2MeNH2105	-4.530	-6.207	-5.505
2775_10MeNH2MeNH2110	-3.675	-5.835	-5.278
2787_12MeNH2Water090	-17.110	-13.859	-17.396
2788_12MeNH2Water095	-14.075	-11.953	-14.576
2789_12MeNH2Water100	-11.665	-10.426	-12.369
2790_12MeNH2Water105	-9.740	-9.185	-10.621
2791_12MeNH2Water110	-8.201	-8.169	-9.226

Table 12: Distances scans of Benzene water values

dimers	SAPT	AMOEBA	CP-AMOEBA
3123_54BenzeneWaterOHpi090	-4.370	5.499	-5.296
3124_54BenzeneWaterOHpi095	-3.432	5.486	-3.795
3125_54BenzeneWaterOHpi100	-2.769	5.536	-2.668
3126_54BenzeneWaterOHpi105	-2.288	5.620	-1.815
3127_54BenzeneWaterOHpi110	-1.932	5.721	-1.166

Table 13: Distances scans of Benzene π stacking

dimers	SAPT	AMOEBA	CP-AMOEBA
2883_24BenzeneBenzenepipi090	-8.480	15.690	-11.851
2884_24BenzeneBenzenepipi095	-4.749	15.556	-6.321
2885_24BenzeneBenzenepipi100	-2.564	15.430	-2.743
2886_24BenzeneBenzenepipi105	-1.292	15.314	-0.428
2887_24BenzeneBenzenepipi110	-0.560	15.208	1.639

Table 14: Benzene T-shaped

dimers	SAPT	AMOEBA	CP-AMOEBA
3067_47BenzeneBenzeneTS090	-4.928	13.934	-13.167
3068_47BenzeneBenzeneTS095	-3.224	13.936	-8.150
3069_47BenzeneBenzeneTS100	-2.160	13.956	-4.706
3070_47BenzeneBenzeneTS105	-1.486	13.985	-2.324
3071_47BenzeneBenzeneTS110	-1.055	14.018	-0.673