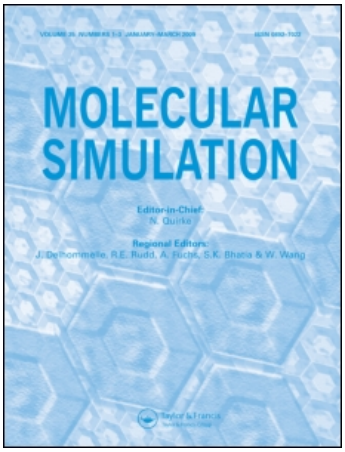

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

DRF90: a polarizable force field

M. Swart[a]; P.Th. van Duijnen[a]
[a] Theoretische Chemie, MSC Rijksuniversiteit Groningen, Groningen, The Netherlands

To cite this Article Swart, M. and van Duijnen, P.Th.(2006) 'DRF90: a polarizable force field', Molecular Simulation, 32: 6, 471 — 484
To link to this Article: DOI: 10.1080/08927020600631270
URL: http://dx.doi.org/10.1080/08927020600631270

Molecular Simulation, Vol. 32, No. 6, 15 May 2006, 471–484

Taylor & Francis
Taylor & Francis Group

DRF90: a polarizable force field

M. SWART*† and P.TH. VAN DUIJNEN

Theoretische Chemie, MSC Rijksuniversiteit Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

(Received July 2005; in final form July 2005)

The direct reaction field (DRF) approach has proven to be a useful tool to investigate the influence of solvents on the quantum/classical behaviour of solute molecules. In this paper, we report the latest extension of this DRF approach, which consists of the gradient of the completely classical energy expressions of this otherwise QM/MM method. They can be used in (completely classical) molecular dynamics (MD) simulations and geometry optimizations, that can be followed by a number of single point QM/MM calculations on configurations obtained in these simulations/optimizations. We report all energy and gradient expressions, and results for a number of interesting (model) systems. They include geometry optimization of the benzene dimer as well as MD simulations of some solvents. The most stable configuration for the benzene dimer is shown to be the parallel-displaced form, which is slightly more stable (0.3 kcal/mol) than the T-shaped dimer.

Keywords: Polarizability; Force field; Geometry optimization; Molecular dynamics simulations

1. Introduction

Recently, we have seen a renewed interest in the development and application of so-called polarizable force fields [1–4]. Although these studies are all concerned with a proper description of polarization effects in chemical systems, the philosophy behind the approaches is substantially different. In the paper by van Duijneveldt *et al.* [4], the separate energy terms have a well-defined physical meaning, which enhances transferability to other molecules. Moreover, the parameters were obtained from *ab initio* calculations (MP2) using large basis sets and corrected for the basis set superposition error (BSSE). On the other hand, Kaminski *et al.* fitted atomic polarizabilities either to the change in electrostatic potential (ESP) at a set of grid points outside the van der Waals surface of the molecule [1] or to three-body energies of five small molecules [2]. The fitting on a grid in their first study seems questionable as it is known from potential derived charge analysis methods that ESP charges depend on the choice for the grid that is used [5–7] and charges of buried atoms can not be obtained with great confidence with these methods, which is likely to cause similar problems here. Also their second paper raises questions: it is unclear which method has been used to obtain the "three-body" energies and whether the results have been corrected for BSSE. More importantly, their "three-body" energies are obtained from intramolecular polarization only, while three-body energies usually concern intermolecular interactions [8], and the polarizabilities will be used to describe intermolecular interactions.

One of the main objections to studies such as those by Kaminski *et al.* [1,2] is that the polarizability enters as a fitting parameter, thereby ignoring the fact that the polarizability is an intrinsic property of the molecule. This is in contrast to the two other papers [3,4], as well as previous studies in our group: our first paper appeared already in the beginning of the 1980s [9], when Thole introduced screening functions to avoid the polarization catastrophe. His model has been revised a few years ago, where a new standard set of atomic polarizabilities has been constructed [10]. These atomic polarizabilities provide molecular polarizabilities with a mean absolute deviation of 3–4%, which is of the same order of magnitude as the experimental uncertainty and gives the same performance [11] as high-level quantum-chemical calculations (time dependent density functional theory with a large TZ2P++basis set using the asymptotically correct van Leeuwen–Baerends potential). Recently, the same model has been applied to amino acid residues [12], where it was shown that the atomic polarizabilities

*Corresponding author. Email: swart@chem.vu.nl
†Current address: Theoretische Chemie, Vrije Universiteit, Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands.

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020600631270

reproduce TD–DFT values with a mean absolute deviation of ca. 1%.

Over the last decades, the old [9] and new [10] sets of atomic polarizabilities have been used extensively within the direct (or discrete [13]) reaction field (DRF) approach [13–17], a method for combined classical and quantum mechanics calculations (QM/MM approach [18]) to study the effect of surroundings on chemical systems that are necessarily described by quantum-chemical methods. The surroundings usually consist of polarizable explicit solvent molecules, while the quantum-chemical system is the one of special interest (solute), which can in principle be treated by any quantum-mechanical method (RHF, MCSCF, CI, DFT). In the past, the method has been used to study among others the effect of the surroundings on proton transfer in papain [19–22], the influence of solvent rearrangement on the "sudden" polarization in alkenes [23–25], dissociation reaction of tert-butyl chloride [26], many-body effects [8] and the solvent effect on the $n \rightarrow \pi^*$ transition of acetone [13,27,28].

In our previous studies we used Metropolis Monte Carlo (MC) simulations [29] (with importance sampling) to study the statistical distribution of the solvent molecules around the solute; molecular dynamics (MD) simulations [29] were not yet possible as the gradients of the energy expressions in the DRF approach were not available. In this paper, we report these gradient expressions so that from now on we cannot only perform MD simulations but also geometry optimizations. We will first describe the polarization effects, report the gradient expressions and then move on the results, where we will first focus on some small systems to give a better understanding of the DRF approach, its characteristics and strengths, and then report the first MD simulations with our polarizable force field.

2. Polarizable force field

For a molecule containing N polarizable points (with polarizabilities α_p), the induced dipole in point b (μ_b) is obtained from the external field E_{external} as well as the dipole field T_{bc} of all other induced dipoles μ_c:

$$\mu_b = \alpha_b \left[E_{\text{external}} + \sum_{c \neq b}^{N} T_{bc} \mu_c \right]. \tag{1}$$

This can be written as a matrix equation:

$$\mathbf{M} = \boldsymbol{\alpha}[\mathbf{E} + \mathbf{TM}] \tag{2}$$

which can be solved directly with the inverse of matrix $\mathbf{A} = \boldsymbol{\alpha}^{-1} - \mathbf{T}$:

$$\mathbf{AM} = \mathbf{E} \rightarrow \mathbf{M} = \mathbf{A}^{-1}\mathbf{E}. \tag{3}$$

The so-called relay matrix $\mathbf{R} = \mathbf{A}^{-1}$ gives the linear response of the molecule to a given external field, and can be reduced to a 3×3 molecular polarizability tensor. Likewise, the relay matrix can be reduced to block-diagonal form for an atom (or a group of atoms), resulting in the effective atomic (group) polarizability tensors, that can subsequently be used for intermolecular interactions:

$$\alpha_b^{\text{eff}} = \sum_{i,j \in b} \mathbf{R}_{ij}. \tag{4}$$

The point dipole model as expressed in equation (1) can lead to infinite polarization by the cooperative (head–tail) interaction between two induced dipoles (the polarization catastrophe). Thole [9] "repaired" this deficiency by screening the dipole field tensor T_{bc} to account for overlapping charge densities (see Appendix A for the screening functions).

The relay matrix can also be used for intermolecular interactions; it is then constructed in the same way as above. The inducing field consists of the eventual external field plus the electrostatic field of all permanent point charges (where atoms b and d are in two different molecules; q_d the permanent charge on atom d; R_{bd} the distance between atoms b and d; $\mathbf{r}_{bd}$ the x, y or z-component of the distance vector; f_E the screening function for the electric field as given in Appendix A):

$$E(r_b) = E_{\text{external}} + \sum_{d \neq b} f_E \frac{q_d}{R_{bd}^3} \mathbf{r}_{bd}. \tag{5}$$

The interaction energy related to the induced dipoles (induction energy; we reserve the term polarization energy for the energy cost needed to induce the dipoles) is obtained as:

$$U_{\text{ind}} = -\frac{1}{2} \sum_b E(r_b)\mu_b = -\frac{1}{2}\mathbf{ERE} \tag{6}$$

and contains both the energy cost to create the dipoles and the beneficial interaction of the dipoles with the inducing fields.

The induced dipoles may be obtained from a matrix-vector multiplication of the relay matrix with the fields of equation (5), or more conveniently, by an iterative procedure to solve equation (2). We generally use the matrix inversion for obtaining effective atomic polarizability tensors, and obtain the induced dipoles with a simple Gauss-Jordan iterative scheme, which in our experience works quite efficiently.

3. Energy and gradient expressions

In the DRF-approach, the interaction energy U_{drf} is composed of bonding terms (U_{MM}), non-bonding terms (U_{est}, U_{dis}, U_{rep}, U_{ind}) and environment terms (U_{fld}, U_{cav})[†]:

$$U_{\text{drf}} = U_{\text{MM}} + U_{\text{est}} + U_{\text{dis}} + U_{\text{rep}} + U_{\text{ind}} + U_{\text{fld}} + U_{\text{cav}}. \tag{7}$$

For the molecular mechanics energy (U_{MM}) any appropriate intramolecular force field (be it MM2/3, [31] AMBER95, [32] CHARMM [33] or OPLS [34]) may be used, or alternatively, the rigid body approximation may be used; the electrostatic energy (U_{est}) has, apart from appropriate screening functions (see Appendix A) the

[†]Details of the energy expressions are available in the supporting information and in ref. [14,30].

usual form; for the dispersion energy ($U_{\rm dis}$) we use the (an)isotropic Slater-Kirkwood expression; for the repulsion energy ($U_{\rm rep}$) we employ an adapted CHARMM expression that may be combined with charge dependent atomic radii from Frecer's model [35] (Appendix B); the expression for the induction energy ($U_{\rm ind}$) is given in equation (5); the field energy ($U_{\rm fld}$) is the energy related to the interaction of permanent dipoles with an external field (if present); the cavitation energy ($U_{\rm cav}$) is the energy needed to create a vacuum within a dielectric continuum in which the system is placed, for which we use Pierotti's formulation [36].

Important issues involved in performing molecular simulations are long-range electrostatics and the box surrounding the molecules to prevent the molecules from evaporating from the system. In MC simulations the latter can be achieved easily by surrounding the system with an imaginary box of any given shape and size, and rejecting any step in which one of the molecules moves outside the box. A problem arises however when switching to MD simulations, since there is no option to "reject" a step within them. Therefore, we have chosen to add a term to the total energy (wall force energy $U_{\rm wall}$, see below), which ensures that the particles stay within the simulation box. For the long-range electrostatics with a dielectric continuum, we use Friedman's image charge method [37] with energy $U^{\dagger}_{\rm img}$. The total energy is therefore given by the sum of $U_{\rm drf}$, $U_{\rm wall}$ and $U_{\rm img}$.

3.1 Wall force

We explored a repulsive energy term that will push molecules back into the simulation box if they are moving outside it, to keep the molecules from evaporating. We have chosen an exponential function as illustrated in figure 1.

As we already have an exponential function that accounts for overlapping charge densities (see Appendix A), we explored using that one also for the wall force. The additional energy term added is then between all real atoms b and imaginary particles w:

$$U_{\rm wall} = \sum_b \frac{a^3}{8\pi} e^{-au} \quad u = \frac{R_{bw}}{(\alpha_b \alpha_w)^{1/6}}. \tag{8}$$

For this expression we need the positions of the imaginary particles (r_w) and their polarizabilities (α_w). The imaginary particle represents a solvent molecule, and therefore we can use for α_w the molecular polarizability of the solvent [10]. We place the imaginary particle at a distance R_w outside the image box, where the distance R_w is the sum of the atomic radius of atom b and the solvent radius ($R_{\rm solvent}$); the latter is obtained from the macroscopic molecular density ($\rho_{\rm macro}$) and mass ($M_{\rm mol}$):

$$R_{\rm solvent} = \left(\frac{3}{4\pi}\frac{M_{\rm mol}}{\rho_{\rm macro}}\right)^{1/3}. \tag{9}$$

Using these values, the imaginary particle has a negligible interaction with an atom within or at the boundary of the box (<0.1 kcal/mol), which will increase smoothly to keep the atoms within the box.

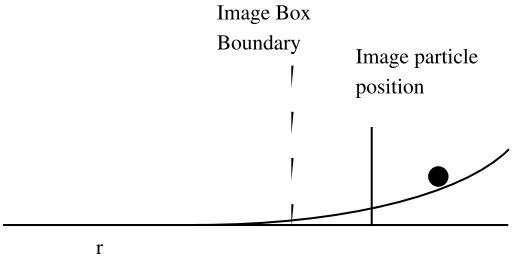

Figure 1. Wall force potential to keep particles from evaporating.

3.2 Gradient expressions

The expressions for the gradient of the classical energies follow naturally from the energy expressions, if the derivative of the energy with respect to the interatomic distance R_{bc} is taken ($\partial U/\partial R_{bc}$). The gradient for one of the two atoms involved is then obtained by multiplication with the partial derivative of R_{bc} with respect to the atomic coordinates ($\partial R_{bc}/\partial r_b$):

$$\begin{aligned}
\frac{\partial U}{\partial r_b} &= \frac{\partial U}{\partial R_{bc}}\frac{\partial R_{bc}}{\partial r_b} \\
\frac{\partial R_{bc}}{\partial r_b} &= \left\{\frac{x_{bc}}{R_{bc}}, \frac{y_{bc}}{R_{bc}}, \frac{z_{bc}}{R_{bc}}\right\} \\
\frac{\partial R_{bc}}{\partial r_c} &= -\frac{\partial R_{bc}}{\partial r_b}.
\end{aligned} \tag{10}$$

The following energy gradient components $i \in (x, y, z)$ are then obtained for respectively, the electrostatic, isotropic dispersion, anisotropic dispersion and repulsion interactions:

$$G^{\rm est}_{bc,i} = -f_E \frac{q_b q_c}{R^3_{bc}} r_{bc,i} \tag{11}$$

$$\begin{aligned}
G^{\rm dis,iso}_{bc,i} = &-\left(\frac{3}{2}f_T^2 - \frac{5}{6}f_T f_E - \frac{5}{2}f_T f_D + \frac{5}{6}f_E f_D\right) \\
&\frac{3}{2}\frac{\alpha_b \alpha_c}{\sqrt{\alpha_b/N^e_b} + \sqrt{\alpha_c/N^e_c}} \frac{6 r_{bc,i}}{R^8_{bc}}
\end{aligned} \tag{12}$$

$$\begin{aligned}
G^{\rm dis,aniso}_{bc,i} = &-\frac{1}{4}\frac{1}{\sqrt{\alpha_b/N^e_b} + \sqrt{\alpha_c/N^e_c}}[WR_{xx} + WR_{xy} \\
&+ WR_{yy} + WR_{xz} + WR_{yz} + WR_{zz}] \\
WR_{jk} = &\frac{(\partial W_{jk}/\partial r_{bc,i})}{1+\delta_{jk}}[\alpha_{jx,b}\alpha_{kx,c} + \alpha_{jy,b}\alpha_{ky,c} + \alpha_{jz,b}\alpha_{kz,c} \\
&+ \alpha_{kx,b}\alpha_{jx,c} + \alpha_{ky,b}\alpha_{jy,c} + \alpha_{kz,b}\alpha_{jz,c}]
\end{aligned} \tag{13}$$

$$\begin{aligned}
G^{{\rm rep},X}_{bc,i} = &-\frac{1}{2}\frac{(R^{vdw}_b + R^{vdw}_c)^6}{R^6_{bc}}\left[G^{{\rm dis},X}_{bc,i} + 6 r_{bc,i} \quad U^{{\rm dis},X}_{bc}/R^2_{bc}\right] \\
&X = {\rm iso, aniso}.
\end{aligned} \tag{14}$$

The screening factors f_E, f_T, f_D are given in Appendix A, as well as the matrix W; α is the atomic polarizability (tensor), $R^{\rm vdw}$ the atomic radius and N^e the number of valence electrons.

For the induction energy, things are a bit more complicated as we need the derivative of an inverse matrix (since $\mathbf{R} = \mathbf{A}^{-1}$):

$$G^{\text{ind}} = -\mathbf{TM} - \frac{1}{2}\mathbf{E}\frac{\partial \mathbf{R}}{\partial \mathbf{r}}\mathbf{E} = -\mathbf{TM} - \frac{1}{2}\mathbf{E}\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{r}}\mathbf{E}. \quad (15)$$

However, as $\mathbf{A}^{-1}\mathbf{A} = 1$, by taking the gradient of both sides of this equation, we obtain:

$$\begin{aligned} \frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{r}}\mathbf{A} + \mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{r}} = 0 \rightarrow \\ \frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{r}}\mathbf{A} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{r}} \rightarrow \\ \frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{r}} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{r}}\mathbf{A}^{-1} = -\mathbf{R}\frac{\partial \mathbf{A}}{\partial \mathbf{r}}\mathbf{R} \\ = \frac{\partial \mathbf{R}}{\partial \mathbf{r}}. \end{aligned} \quad (16)$$

This is again a simple expression, because the fields $\mathbf{E}$ can be multiplied with the relay matrix $\mathbf{R}$ to give the induced dipoles, to give the following final expression for the gradient of the induction:

$$G^{\text{ind}} = -\mathbf{TM} + \frac{1}{2}\mathbf{M}\frac{\partial \mathbf{A}}{\partial \mathrm{r}}\mathbf{M}. \quad (17)$$

4. Computational details

All calculations were performed on a cluster of IBM/RS6000 and Pentium-Linux workstations, where sometimes the parallel version of the DRF90 program was used to speed up the calculations. For the relay equation, a simple iterative scheme was used where the induced dipoles were kept, and re-used as starting point in the next MD/optimization-step. In the relay equations for intermolecular interactions (equations (1)–(3), (6)), all atoms were taken into account explicitly (i.e. no GROUP polarizability was made), as it gives a better description with only a limited increase in CPU-time for small systems. For larger systems (thousands of atoms) this may lead to a relay matrix cq. equation with a dimension of several thousands, in which case it might become the bottleneck. Furthermore, we used anisotropic dispersion/repulsion throughout which comes also with a small increase in CPU-time, but gives a more accurate description.

The monomer geometries of the molecules were obtained from density functional theory calculations with the ADF program [38]. The Becke [39]-Perdew [40] exchange-correlation potential was used in a triple zeta valence plus double polarization (TZ2P) basis set of Slater type orbitals. The atomic charges were taken from the Multipole Derived Charge analysis [7] (MDC-q where appropriate, MDC-d otherwise), which gives charges that reproduce by construction both the atomic and molecular multipoles.

The MD simulations reported in this paper were performed by using Verlet's leap-frog algorithm, in conjunction with a Nosé–Hoover (NVT) thermostat [41] to keep the temperature constant.

5. Results and discussion

5.1 Hydrogenfluoride dimer

The importance of a proper representation of the charge distribution is shown by the HF dimer. Being a diatomic molecule, it is impossible to represent both the dipole moment and the quadrupole [42] moment simultaneously by using fractional charges located at the two atoms only. In fact, by doing so the quadrupole moment is underestimated by 40%. By adding a third point charge in the bond midpoint, both the dipole and quadrupole moment can be represented exactly. The potential and electrostatic field of the additional point charge should then also be taken into account for the electrostatics and induction, respectively, for which we need a polarizability value of the additional point charge QQ. For this, we exploit the Thole screening functions (Appendix A) to get a weighted average of all other atomic polarizabilities, with the distance $r_{QQ,i}$ between the QQ point and atom:

$$\alpha_{QQ} = \sum_i \alpha_i e^{-au^*}; \quad u^* = \frac{r_{QQ,i}}{\alpha_i^{1/3}}. \quad (18)$$

The sum runs over all atoms of the molecule to which the QQ-point belongs, excluding any other QQ-points; in this case therefore just the H- and F-atom.

The geometry of the hydrogen-bonded HF-dimer is described by three geometric variables, under the assumption of rigid HF molecules (figure 2):

- the F–F distance R;
- the angle θ_1 between the acceptor hydrogen atom and the two F-atoms; and
- the angle θ_2 between the donor molecule hydrogen atom and the F-atoms.

Experimentally, the values for these three variables have been estimated at 2.72–2.79 Å, 117 ± 6 and 7–9°, respectively [43]. If we use only two point charges (i.e. represent only the monopole and dipole of the monomer), the optimized dimer geometry differs considerably from experiment with values of 3.022 Å, 133.7 and 23.7° for R, θ_1 and θ_2, respectively. Adding the third point charge (e.g. represent also the quadrupole moment of the monomer), the parameter values are improved significantly: 3.047 Å, 115.0 and 5.1°, respectively. Although we predict a slightly larger F–F distance, the agreement with the experimental data is quite satisfactory.

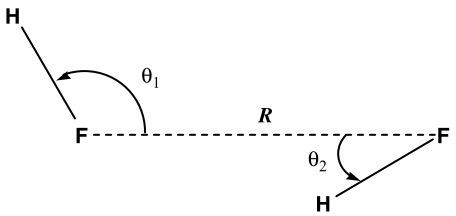

Figure 2. Hydrogen fluoride dimer geometry.

5.2 Benzene dimer

Many investigations have focused on the benzene dimer [44–55], be it from experimental, quantum-chemical or classical-mechanical point of view. The attention for this dimer can be attributed to the fundamental issue at stake: the interactions between large π-electron clouds, or the so-called π-stacking. This is an important property in many fields: organic, polymeric and biochemistry, as well as biology. For a theoretical chemist, the system raises an interesting challenge. Conventional *ab initio* methods show difficulties with the system size, while density functional theory is known to behave less accurately than usual with respect to dispersion energy (the most important energy term for this dimer, as will be shown later on). Furthermore, from experiments it is known that there are probably a few isomers that almost have equal energies. However, no conclusions can be drawn from these experiments, as they implicitly include entropy effects; e.g. although one isomer may have a less favorable energy, it may be more often present in the experiments because there are more opportunities to make that particular isomer. Three important isomers can be distinguished (among others): a parallel (P), parallel-displaced (PD) and a T-form (T), where the PD- and T-form exist in two ways (figure 3). Other groups have considered several more isomers but they are more or less related to either one of these three.

As the first non-vanishing molecular multipole moment of benzene is the quadrupole moment, it is important that the charges describing the charge distribution within the molecule should reproduce it. The multipole derived charges (MDQ-q) [7] by construction represent the computed molecular quadrupole moment (Θ_{zz} -5.59 a.u.), which is only slightly lower than the experimental value of -6.47 a.u. [56] To investigate the importance of this underestimation we used also a set of charges that were uniformly scaled upwards to reproduce the experimental value. Likewise, it is known that the standard DRF-polarizabilities [10] underestimate the mean value of the polarizability of conjugated molecules. Therefore, we also used a set of atomic polarizabilities where we scaled the polarizability of carbon (to 10.5850) to reproduce the experimental polarizability [56]. In total, we therefore have four models of benzene; they are referred to as PQ, where the P stands for the polarizability model (either D for standard DRF or E for the values scaled to reproduce the experimental value) and Q for the charge model (either M for Multipole Derived Charges or E for the scaled "experimental" charges). For these four models, we optimized the geometry of the five isomers of the benzene dimer (figure 3). The optimized intermolecular distances and interaction energies are given in table 1.

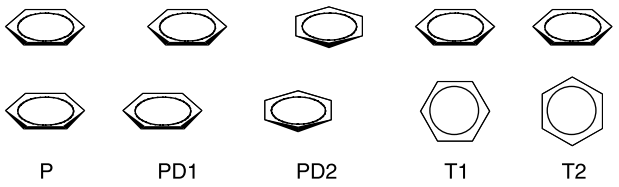

Figure 3. Possible conformations of the benzene dimer.

The four models show similar results, i.e. the parallel-displaced isomer is most stable with the T form less favorable by approximately 0.2 kcal/mol. Moreover, for all four models the interaction energy is close to the experimental value of 2.2 ± 0.3 kcal/mol. Our results are in good agreement with the work of Hobza *et al.* [48–50].

In table 2, we report the energy components of the isomers at their optimized geometries. Dispersion is shown to be the most important component, both in absolute sense as discriminative for relative stability of the isomers. Only for the parallel form (P) is electrostatics playing an important (repulsive) role.

5.3 Simulation of liquids

Finally, we performed MD (NVT) simulations for seven liquids differing in polarity, size, dielectric constant and polarizability, to check the impact of the wall force and the Nosé–Hoover thermostat [41] we are using. We chose to have 200 water molecules and determined from the macroscopic density the size of a spherical box that would be needed for this number of water molecules (21.3 Bohr). The same box size was used for the other solvents, but the number of molecules was adapted such that the macroscopic density of each solvent was retained. The RanGenConf program [57] was used to generate a starting structure and an initial run of 20 ps was performed for equilibration. After this, the production run was performed over 50 ps, where radial distribution functions were made for several atompairs. In both the equilibration and production run, we used a timestep of 1 fs and a

Table 1. Intermolecular distance r (Å) and interaction energies U (kcal/mol) for benzene dimer.

	*DM-model**		*DE-model**		*EM-model**		*EE-model**		*MP2*[†]	
	r	U	r	U	r	U	r	U	r	U
P[‡]	3.65	−1.10	3.73	−0.58	3.64	−1.63	3.69	−1.12	3.90	−0.85
PD1[‡]	4.80	−2.17	4.88	−2.35	4.58	−2.49	4.85	−2.65	3.85	−1.98
PD2[‡]	4.74	−2.24	4.87	−2.42	4.57	−2.56	4.82	−2.73	–	–
T1[‡]	5.00	−2.03	4.97	−2.28	5.00	−2.28	4.97	−2.53	5.10	−1.66
T2[‡]	5.10	−2.03	5.05	−2.34	5.10	−2.26	5.05	−2.59	5.00	−1.89

*D = standard DRF polarizabilities; M = multipole derived charges; E = scaled to reproduce experimental values of molecular quadrupole and/or polarizability. [†] Values from Hobza *et al.* [49]. [‡] Different forms for benzene dimer (see text and figure 3).

Table 2. Energy components (kcal/mol) of the benzene dimer in the EE-model.

	Electrostatics	*Dispersion*	*Repulsion*	*Induction*	*Total*
P*	2.18	−4.76	1.71	−0.25	−1.12
PD1*	−0.52	−3.85	1.91	−0.19	−2.65
PD2*	−0.54	−3.89	1.94	−0.24	−2.73
T1*	−0.67	−3.23	1.60	−0.23	−2.53
T2*	−0.87	−3.09	1.69	−0.32	−2.59

* Different forms for benzene dimer (see text and figure 3).

relaxation time t of 0.1 ps. In table 3, we report the details of the setup of the simulations together with solvent information. The Nosé–Hoover thermostat performs well in obtaining the desired temperature for all solvents.

The diversity of the solvents is reflected in the total energy, which ranges from some −42 kcal/mol for the apolar tetra to −814 kcal/mol for water. The energy components are reported in table 4. For acetonitrile, the largest contributions come from electrostatic, dispersive and repulsive interactions, which are similar in magnitude; induction energy is (apart from water) the largest for all solvents. This can probably be related to its more or less linear shape, which enables the induced dipoles to orient parallel to each other (head-to-tail) which is the most favourable position for two dipoles. The situation is completely different for benzene, whose interactions are almost completely resulting from dispersion and repulsion, similar as was observed for the benzene dimer. Although the polarizability of benzene is more than twice that of acetonitrile, the induction is almost eight times as small. This is due to the larger size of the benzene molecules and the lack of linearity as exhibited by acetonitrile.

Chloroform is the solvent with the smallest total energy in the MD simulations, despite its permanent dipole moment of 1.1 Debye, and its considerable polarizability of 57 Bohr^3 (table 3). However, due to its more or less spherical shape and reasonably large size, it is much harder to pack the molecules nicely into the box. The dioxan molecules on the contrary, having almost identical size and polarizability values, are almost planar and therefore are better able to fit nicely on top of each other. The average intermolecular distance is reduced, thereby leading to an increased repulsion, but at the same time the dispersion is enhanced even more, leading to an overall gain in energy. It is noteworthy that even though the dioxan molecules do not carry a permanent dipole moment, the induction energy is almost 10 kcal/mol larger than for chloroform, which has a permanent moment. This can be again be related to the larger intermolecular distances in chloroform relative to dioxan.

Methanol behaves similarly to acetonitrile, in that it has considerable electrostatic, dispersive and repulsive interactions. However, the induction energy is almost half that of acetonitrile. This can again be attributed to the almost linearity of the latter. Tetra does not carry a substantial electric moment, and as a result does not have appreciable electrostatic or inductive interactions. The major interactions result from dispersion and repulsion.

The interactions of water are outstanding: both the electrostatic and repulsion energies are larger than the values of the other solvents summed together. This is due to the small size of the water molecules, which enables them to come close to each other, and thereby have large beneficial interactions. This is exhibited by the dispersion and induction energies, whose values are even more

Table 3. Macroscopic density ρ (kg/l), dipole μ (Debye), dielectric constant ε, polarizability α (a.u.) and "molecular radius" R (Bohr) of solvents.

Solvent	*# Molecules*	ρ	μ	ε	α	R
Acetonitrile	69	0.786	4.14	37.5	29.63	5.19
Benzene	40	0.876	0.0*	2.3	70.12	6.20
Chloroform	45	1.483	1.06	4.7	57.41	5.99
1,4-dioxan	42	1.033	0.0†	2.2	58.00	6.11
Methanol	89	0.791	1.64	32.6	22.02	4.77
Tetra	37	1.594	0.0	2.2	70.72	6.37
Water	200	0.998	1.85	78.5	9.69	3.64

*Benzene: $\Theta_{zz} = -5.58$ a.u. † 1,4-dioxan: $\Theta_{xx} = -4.31$, $\Theta_{yy} = -2.32$, $\Theta_{zz} = 6.63$ a.u.

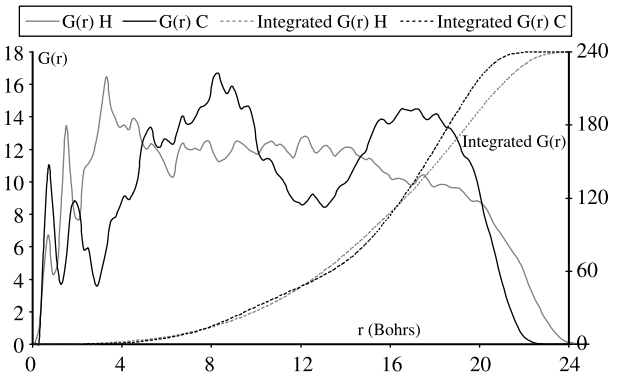

Figure 4. G(r) of C/H-atoms from centre for liquid benzene (at 298 K).

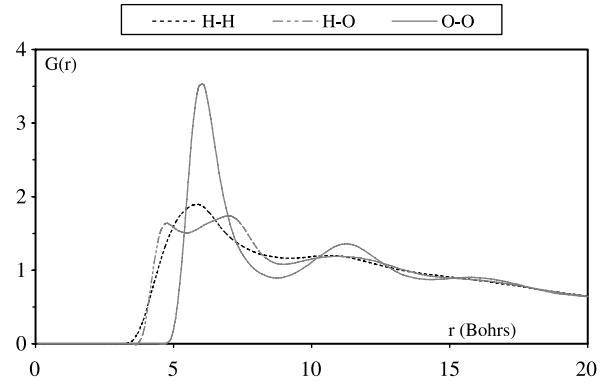

Figure 5. G(r) of several atompairs for water (at 298 K).

Table 4. MD average temperature (K) and energy components* (kcal/mol).

	Acetonitrile	*Benzene*	*Chloroform*	*1,4-dioxan*	*Methanol*	*Tetra*	*Water*
Temperature	298 ± 22	298 ± 33	298 ± 27	298 ± 31	298 ± 21	298 ± 31	298 ± 16
DRF	−404.5	−177.6	−92.1	−186.6	−286.2	−88.3	−1014.9
Elst	−250.7	−8.9	−2.6	−24.7	−133.0	0	−696.0
Disp	−271.1	−296.1	−175.9	−258.0	−259.3	−151.6	−644.2
Rep	180.7	136.0	88.9	107.3	137.7	63.3	498.7
Ind	−63.4	−8.7	−2.5	−11.3	−31.7	0	−173.4
CAV	6.3	8.8	19.7	15.4	17.3	16.9	13.7
Wall	8.0	8.9	19.9	15.6	19.1	16.9	16.9
Fried	−1.7	−0.1	−0.2	−0.2	−1.9	0.0	−3.2
KIN	60.0	30.2	39.7	30.9	75.9	29.0	187.2
TOTAL	−338.1	−138.6	−32.7	−140.3	−193.1	−42.4	−814.0
U_{total}/N	4.9	3.5	0.7	3.3	2.2	1.2	4.1

*$U_{TOTAL} = U_{DRF} + U_{CAV} + U_{KIN}$; $U_{DRF} = U_{elst} + U_{disp} + U_{rep} + U_{ind}$; $U_{CAV} = U_{wall} + U_{friedman}$.

outstanding if the relatively small polarizability of water is taken into account.

To check the impact of the wall force we introduced, we checked the radial distribution function of C/H-atoms from the center of the system in figure 4 as obtained in the simulation of liquid benzene.

As hoped for, the wall force pushes the particles back inside when they try to "leave" the box, and therefore prevents the molecules from evaporating.

The rdf's for several atompairs obtained in the simulation of water is given in figure 5; for instance, not only the O–O pair, but also the H–H and O–H rdf's are included. They show not only the structure of the first solvation shell, but also the second shell is well represented. The long tail of the rdf is less pronounced due to the relatively small box size. Also for the other solvents (see supplementary information), we find the structure of the solvation shells.

6. Conclusions

We report the DRF force field and its gradient of the classical energy expressions within the DRF approach, which enables us to perform geometry optimizations, MD simulations and vibrational frequency analyses (through numerical differentiation of the analytical gradients). All energy and gradient expressions are given together with the MC, MD and optimization implementations.

We have subsequently used them to perform geometry optimizations on some interesting dimers. The most important of these is the benzene dimer, which can occur in many configurations, that can more or less be divided into three classes: parallel, parallel-displaced and the so-called T-shaped configuration. We have used four models to represent the benzene molecules, with charges representing either the molecular quadrupole moment as predicted by density functional theory, or representing the experimental value; and with either the standard DRF-polarizabilities or ones that represent the experimental molecular polarizability. The benzene dimer has been optimized for five configurations using all four models. Although the energy differs between the four models by some 0.5 kcal/mol, the most stable configuration is in all cases a parallel-displaced one, with the T-shaped configuration less stable by 0.1–0.2 kcal/mol. Moreover, in all four models the energy of the most stable configuration is in very good agreement with the experimental value of -2.2 ± 0.3 kcal/mol. Therefore, we conclude that the most stable configuration of the benzene dimer is the parallel-displaced form.

To check the impact of the wall force in MD simulations, we have simulated seven commonly used organic solvents. The Nosé–Hoover thermostat we use to keep the temperature constant results in a NVT-ensemble with correct temperatures. The solvents chosen differ very much in polarity, size and polarizability, which shows up in the energy components, but all simulations run perfectly, including the wall force.

Acknowledgements

Vladimir Frecer is greatly acknowledged for allowing us to report the values for the charge dependent radii prior to publication.

Appendix A. Thole's damping functions

Thole tested several functions [9] to describe charge densities of which one has survived: the exponential one that depends on the reduced distance

$$u = \frac{r_{ij}}{(\alpha_i \alpha_j)^{1/6}} :$$

$$\rho(u) = \frac{a^3}{8\pi} e^{-au}.$$

Associated with this function are screening factors f_v, f_E, f_T, f_D for respectively the electric potential, field, field

gradient and the gradient of the field gradient:

$$v = au$$
$$f_V = 1 - \left(\frac{1}{2}v + 1\right)e^{-v}$$
$$f_E = f_V - \left(\frac{1}{2}v^2 + \frac{1}{2}v\right)e^{-v}$$
$$f_T = f_E - \frac{1}{6}v^3 e^{-v}$$
$$f_D = f_T - \frac{1}{30}v^4 e^{-v}.$$

With these screening functions, the electric potential, field and so on, are given by the following equations (without screening, all factors f_v, f_E, f_T, f_D have the value 1, and the usual formulas are retained):

$$V = \frac{f_V}{r}$$
$$E_i = \frac{-f_E \mathbf{r}_i}{r^3}$$
$$T_{ij} = \frac{(3\mathbf{r}_i\mathbf{r}_j f_T - \delta_{ij} r^2 f_E)}{r^5}$$
$$D_{ijk} = \frac{(3r^2\mathbf{r}_i\delta_{jk}f_T + 3r^2\mathbf{r}_j\delta_{ki}f_T + 3r^2\mathbf{r}_k\delta_{ij}f_T - 15\mathbf{r}_i\mathbf{r}_j\mathbf{r}_k f_D)}{r^7}.$$

The **W**-matrix needed for the anisotropic dispersion energy gradient has the following form:

$$\mathbf{W}_{ij} = \frac{\left(3\mathbf{r}_i\mathbf{r}_j f_W + r^2\delta_{ij}f_E^2\right)}{r^8}$$
$$f_W = 3f_T^2 - 2f_E f_T.$$

Appendix B. Frecer's model of charge dependent radii

The basic scheme for charge dependent radii has been described in Section 2.5 of ref. [58]. Since then, Frecer has adopted a new function (a third-order polynomial) [35] to represent the dependence of the atomic radius on the atomic charge:

$$R_{vdw}(q) = R_0 + R_1 \cdot q + R_2 \cdot q^2 + R_3 \cdot q^3 \qquad \text{(B1)}$$

In the following table, we report the values for the atoms considered (all values inÅ).

Atom	R_0	R_1	R_2	R_3
H	1.520	−1.054	−0.293	0.000
He	1.330	−0.300	−0.129	−0.037
Li	2.210	−0.996	0.003	0.096
Be	2.180	−0.335	−0.103	−0.051
B	2.060	−0.298	−0.007	0.009
C	1.930	−0.234	−0.013	0.008
N	1.800	−0.170	−0.015	0.003
O	1.700	−0.140	−0.014	0.002
F	1.610	−0.096	−0.012	0.004
Ne	1.510	−0.098	−0.005	0.003
Na	2.240	−0.835	−0.009	0.094
Mg	2.420	−0.375	−0.062	−0.039
Al	2.410	−0.365	0.021	0.005
Si	2.330	−0.291	0.007	0.006
P	2.260	−0.275	0.015	0.006
S	2.150	−0.240	0.020	0.004
Cl	2.080	−0.208	0.017	0.001
Ar	1.970	−0.251	0.010	0.017

Supplementary information

The supplementary information consists of four parts:

- the (well-known) classical energy expressions;
- an explanation of the RanGenConf program;
- the radial distribution functions of the simulated solvents; and
- the equations related to the expressions for the Boundary Elements (BEs).

Classical energy expressions

In the DRF-approach, the total energy U_{tot} can be separated in several terms, which are physically feasible and enable a more comprehensive understanding of the interactions between different molecules:

$$U_{\text{tot}} = U_{\text{est}} + U_{\text{dis}} + U_{\text{rep}} + U_{\text{ind}} + U_{\text{fld}} + U_{\text{MM}} + U_{\text{cav}}.$$

The terms are respectively: electrostatic energy (U_{est}), dispersion energy (U_{dis}), repulsion energy (U_{rep}), induction energy (U_{ind}), field energy (U_{fld}), Molecular Mechanics energy (U_{MM}) and cavitation energy (U_{cav}). Not only are these energy terms complementary, they are also independent, i.e. one can be changed without influencing the others. Also no new parametrization is needed, since all terms result from the same set of atomic parameters. This set consists for each atomtype of an atomic polarizability and an atomic radius. For the latter we use the charge dependent values from Frecer's model (see also Appendix B of the paper). In this model the atomic radius increases if the atomic charge is negative (more electrons surrounding the nucleus), for which a third order polynomial is used.

Electrostatic energy. The electrostatic energy results from the (static) interaction between the charge distributions of the molecules, or by using a molecular multipole expansion, between the molecular multipoles. However this is not very convenient for simulations, since it would require many interactions involving many double transformations (from local to global coordinates). Much more appropriate is then to use atomic charges only, since they involve one $1/r$ term and no transformations are needed. The energy is then obtained as:

$$U_{\text{est},ij} = \frac{q_i q_j}{r_{ij}} f_{V,ij}. \qquad \text{(S1)}$$

Here, $f_{V,ij}$ is a screening function for the potential V that accounts for overlapping charge densities (see Appendix A for explicit formulas). These charges should of course represent the molecular multipoles, which makes the choice of charge analysis critical. It is most convenient to use the recently developed Multipole Derived Charge

analysis [37], since this contains the advantages of density functional theory [38] (accurate charge densities within the molecule) and represents the molecular multipole moments (at least up to quadrupole, see also Results for HF-dimer).

Dispersion energy. The dispersion energy results from the instantaneous interaction between two induced dipoles, for which we use the Slater-Kirkwood expression [39]:

$$U_{\mathrm{dis},ij} = -\frac{1}{4}\frac{Tr\left(\boldsymbol{\alpha}_i T_{ij}^2 \boldsymbol{\alpha}_j\right)}{\left(\sqrt{\boldsymbol{\alpha}_i/n_i} + \sqrt{\boldsymbol{\alpha}_j/n_j}\right)} \tag{S2}$$

where α_i is the polarizability of atom i and n_i the number of valence electrons of atom i. This interaction can be between two atomic or group polarizabilities, which may be treated as isotropic or anisotropic. In the former case, equation (S2) reduces to a simpler form:

$$U_{\mathrm{dis},ij}^{\mathrm{iso}} = -\frac{6}{4}\frac{\alpha_i\alpha_j}{\sqrt{\alpha_i/n_i} + \sqrt{\alpha_j/n_j}}\frac{\left(\frac{3}{2}f_{T,ij}^2 - f_{T,ij}f_{E,ij} + \frac{1}{2}f_{E,ij}^2\right)}{r_{ij}^6}. \tag{S3}$$

The $f_{E,ij}$ and $f_{T,ij}$ are again screening functions to account for overlapping charge densities (see Appendix A for explicit formulas), but now for respectively the electric field ($f_{E,ij}$) and the electric field gradient ($f_{T,ij}$). The anisotropic expression for the dispersion energy can be rewritten by carrying out the matrix multiplication of T_{ij}^2 to obtain the matrix **W** (see Appendix A), and writing out the trace of remaining matrix multiplication $Tr(\alpha_i \mathbf{W} \alpha_j)$ explicitly and taking advantage of the symmetric character of the **W**-matrix:

$$U_{\mathrm{dis},ij}^{\mathrm{aniso}} = -\frac{1}{4}\frac{1}{\left(\sqrt{\alpha_i/n_i} + \sqrt{\alpha_j/n_j}\right)} \cdot \begin{bmatrix} \mathbf{W}(\alpha_{xx,i}\alpha_{xx,j} + \alpha_{xy,i}\alpha_{xy,j} + \alpha_{xz,i}\alpha_{xz,j}) + \\ \mathbf{W}_{xy}\left(\alpha_{xx,i}\alpha_{xy,j} + \alpha_{xy,i}\alpha_{xx,j} + \alpha_{xy,i}\alpha_{yy,j} + \alpha_{yy,i}\alpha_{xy,j} + \alpha_{xz,i}\alpha_{yz,j} + \alpha_{yz,i}\alpha_{xz,j}\right) + \\ \mathbf{W}_{yy}\left(\alpha_{xy,i}\alpha_{xy,j} + \alpha_{yy,i}\alpha_{yy,j} + \alpha_{yz,i}\alpha_{yz,j}\right) + \\ \mathbf{W}_{xz}\left(\alpha_{xx,i}\alpha_{xz,j} + \alpha_{xz,i}\alpha_{xx,j} + \alpha_{xy,i}\alpha_{yz,j} + \alpha_{yz,i}\alpha_{xy,j} + \alpha_{xz,i}\alpha_{zz,j} + \alpha_{zz,i}\alpha_{xz,j}\right) + \\ \mathbf{W}_{yz}\left(\alpha_{xy,i}\alpha_{xz,j} + \alpha_{xz,i}\alpha_{xy,j} + \alpha_{yy,i}\alpha_{yz,j} + \alpha_{yz,i}\alpha_{yy,j} + \alpha_{yz,i}\alpha_{zz,j} + \alpha_{zz,i}\alpha_{yz,j}\right) + \\ \mathbf{W}_{zz}\left(\alpha_{xz,i}\alpha_{xz,j} + \alpha_{yz,i}\alpha_{yz,j} + \alpha_{zz,i}\alpha_{zz,j}\right) + \end{bmatrix}. \tag{S4}$$

This reduces the time needed to calculate the energy by a factor of two, making it only slightly more expensive than the isotropic case.

Repulsion energy. For the repulsion energy we use the repulsive potential from the CHARMM [40] force field, although there some differences in order to retain the other parts of the DRF force field. It is derived from the dispersion energy (equation (S2)), but multiplied by a factor depending on the VanderWaals radii of the atoms involved:

$$U_{\mathrm{rep},ij} = -\frac{1}{2}U_{\mathrm{dis},ij}\frac{(R_{vdw,i} + R_{vdw,j})^6}{r_{ij}^6}. \tag{S5}$$

Until recently the distinction between isotropic and anisotropic repulsion had not been made in the direct reaction field approach with only isotropic repulsion as option. But just like the dispersion energy, the repulsion energy can be obtained in an isotropic or anisotropic fashion. Furthermore, we use charge dependent atomic radii using Frecer's model [41] (see also Appendix C). This model uses a polynomial of third order, to describe the change in atomic radii due to the atomic charge. If the atomic charge is positive, it means less electrons are surrounding this atoms, and therefore the radius is smaller. When negative, the radius increases due by the same reasoning.

Induction energy. The induction energy is a direct result from the induced dipoles, and is sometimes referred to as polarization energy. However, in the DRF approach this term is associated with the energy cost needed to polarize the charge distribution (i.e. the wavefunction in QM/MM calculations or in the classical case, the polarizabilities). The induction contains both the energy cost needed to induce the dipoles as well as the interaction of the induced dipoles with the electrostatic field:

$$\begin{aligned} U_{\mathrm{ind}} &= -\frac{1}{2}\mathbf{E}_{vr}\mathbf{R}\mathbf{E}_{zf} = -\frac{1}{2}\mathbf{E}_{vr}\mathbf{M} \\ &= \sum_j -\frac{1}{2}E_{vr}(j)\mu_{\mathrm{ind},j}. \end{aligned} \tag{S6}$$

Here, the **M** stands for the induced dipoles vector, E_{zf} for the electrostatic (inducing) electric field vector, E_{vr} for the electrostatic (reaction) field vector and $E_{vr}(j)$ the electrostatic (reaction) field at point j:

$$E(j) = E_{\mathrm{external}} + \sum_{i\neq j}\frac{q_j}{r_{ij}^3}\mathbf{r}_{ij}. \tag{S7}$$

Optional with this energy is the inclusion of a continuum surrounding the system, which is the reason why we make a distinction between the electrostatic inducing and the reaction field. Without the continuum

present, these fields are the same, but with the continuum they are not. Including the continuum is achieved by using an enveloping surface of Boundary Elements (BEs) that can be either a Connolly surface [42] or preferably by using the GEPOL93 [43–46] routines (solvent accessible surface). This preference is due to the fact that the Connolly surface is created for the starting configuration, and will not change afterwards, while the GEPOL93 surface can be reconstructed during the run and usually results in a smaller number of BEs. The explicit formulas for the interaction of the BEs with the molecules and between the dipoles induced at the BEs with each other are given at pages S10–S11.

External field energy. This energy is related to the direct interaction of the point charges with the external electric field, and is obtained as:

$$U_{\mathrm{fld},i} = q_i\left(x_i E_{\mathrm{external},x} + y_i E_{\mathrm{external},y} + z_i E_{\mathrm{external},z}\right). \quad \text{(S8)}$$

This energy term is related to the interaction of the permanent dipoles (represented by the point charges) with the external field.

Molecular mechanics energy. The MM-energy itself consists of several terms and accounts for the intramolecular interactions. It consists of potentials for bonds, angles, dihedral angles and improper dihedral angles:

$$U_{\mathrm{MM}} = U_{\mathrm{bnd}} + U_{\mathrm{ang}} + U_{\mathrm{dih}} + U_{\mathrm{imp}}. \quad \text{(S9)}$$

For all but the dihedral angles, a harmonic potential is used:

$$U_{x,ij} = \frac{1}{2}k_x(x_{ij} - x_{\mathrm{eq}})^2. \quad \text{(S10)}$$

The proper dihedrals are treated by the following function:

$$U_{\mathrm{dih},i} = k_{\mathrm{dih},i}(1 + \cos(n(\phi - \phi_{\mathrm{eq}}) - \pi)). \quad \text{(S11)}$$

We have added the Amber95 [47] intramolecular force field to the program, but in principle any type can be added and used.

Cavitation energy. When surrounding the system under study with a continuum, this can be achieved in many ways. In our implementation, we use an enveloping surface of Boundary Elements (BEs) which is created with the GEPOL93 routines [43–46]. However, whatever choice is made for the representation, one always has to account for the energy cost needed to create a cavity in the continuum. In the GEPOL93 routines, the BEs are constructed from spheres surrounding the atoms, while the surface area is build by summing the partially exposed BE areas. In the same way, the cavitation energy is obtained by using Pierotti's formulation [48] for each BE, as if it were a full sphere, and scaling it with the surface area of that BE divided by the area of the full sphere:

$$U_{\mathrm{cav}} = \sum_{i}^{N_{\mathrm{BEM}}} \Delta G_{\mathrm{cav}} \frac{a_i}{4\pi r_i^2} \quad \text{(S12)}$$

where r_i is the radius of the sphere surrounding the atom that BE i belongs to.

Wall force. The explicit expressions for the energy/force related to this interaction are given in the Wall force section of the paper.

RanGenConf: generation of a configuration of randomly orientated molecules

Before starting a Monte Carlo or a Molecular Dynamics simulation, one needs to have a starting configuration. One could generate one where one puts the molecules at regular intervals throughout the simulation box, but since they are randomly oriented at finite temperatures, it would be more appropriate if the starting configuration would exhibit the same pattern. Therefore we have made a program (RanGenConf or RGC) that does just this. It puts the solute (if present) at the centre of the simulation box, and adds solvent molecules to the box in random positions. More precisely, every molecule to add is first placed at a random position on the box boundary. Then the nearest molecule is determined, and the new molecule moved in such a way that the two atoms (one from each molecule) that are closest to each other touch each other. Or even overlap each other a bit with a predefined amount. Then it is checked that all atoms (or otherwise the centre of mass of the molecule) stay inside the box. If so, the next molecule is handled, else a new random position is tried with a maximum of 100.000 tries.

One issue remains, and that is how to choose the size of the simulation box. If the molecule mass and the macroscopic density of the solvent and solute are known, the size can be obtained from them. However, if they are not known, the question remains. One way to solve this is to calculate the volume of one solute and one solvent molecule in some way and multiply these by respectively the number of solutes and solvents to get the total volume. This has to be adjusted for the fact that the occupied (molecular) volume is always smaller than the volume that one would obtain by dividing the total volume by the number of molecules (free volume), what we call the VanderWaals ratio. There are two options for calculating the free molecular volumes.

One way is by calculating the occupied volume using the GEPOL93 method [43–46] with Frecer's charge dependent radii [41] (but putting the charge to zero), and surrounding the solute by its VanderWaals surface. For a number of solvents (acetonitrile, benzene, chloroform, 1,4-dioxan, methanol, tetra and water) covering the whole

range of the dielectric constant (i.e. from 2 to 78), a striking similarity for the ratio between the occupied vs. the free volume is found (0.80–0.86; average of 0.82 ± 0.03) with these volumes. With this value for the VanderWaals ratio we therefore obtain a reasonable guess for the macroscopic density.

The other method is by taking the sum of the spherical volumes of all atoms, and subtracting from it part of the overlapping volumes between any pair of atoms. We should only subtract part of the overlapping volume, since it is most likely that the overlapping volume overlaps also with other atoms. Suppose we have a molecule with three atoms, we take the sum of the three spherical volumes of the atoms, and subtract from it the overlapping volume of atoms 1 and 2, 1 and 3, and 2 and 3 (figure S1).

The dark shaded region is the region that would be subtracted three times, while it should be subtracted only once. Therefore, we subtract only part of the overlapping volume between two atoms (75%), which results in occupied volumes for the seven solvents with an average VanderWaals factor of 0.59 ± 0.01. This latter method can be applied directly in the RanGenConf program and shows a smaller deviation than the former, and is therefore the preferred method.

Radial distribution functions of simulated solvents

Figures S2–S8.

Boundary elements expressions

Here we report for completeness sake, the relay equations involved in coupling the polarizabilities and the Boundary Elements (BEs). They involve the matrix, whose inverse is again the relay matrix, but now in its complete form:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathbf{R}} & \mathbf{C}^{\mathbf{R}} \\ \mathbf{B}^{\mathbf{C}} & \mathbf{K} & \mathbf{L} \\ \mathbf{C}^{\mathbf{C}} & \mathbf{M} & \mathbf{N} \end{bmatrix}. \tag{S13}$$

The inclusion of the BEs removes the symmetrical nature of the matrix, which necessitates for the gradient the dipole moments as induced by both the inducing as well as the reaction field (see equation (S15) in the paper). The matrix consists of nine parts, which will be given in more detail. Note that p, q refer to polarizabilities, i, j to BEs and a, b to elements of $\{x,y,z\}$. The **A**-submatrix (like before) couples the polarizabilities within the system:

$$\mathbf{A}_{pq} = \alpha^{-1}\delta_{pq} - T_{pq}(1-\delta_{pq}). \tag{S14}$$

Like before, $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$,

$$E_i^a(\mathbf{r}_j) = -\frac{q_i \mathbf{r}_{ij,a}}{r_{ij}^3}$$

and

$$T_{pq,ab} = \frac{3\mathbf{r}_{pq,a}\mathbf{r}_{pq,b}}{r_{pq}^5} - \frac{\delta_{ab}}{r_{pq}^3}$$

The $\mathbf{B}^{\mathbf{R}}$-matrix consists of the following:

$$\mathbf{B}^{\mathbf{R}}_{\mathbf{r}_{a,p}i} = [\varepsilon(1+\kappa r_{pi})e^{-\kappa r_{pi}} - 1]S_i(-T_{pi}\cdot\mathbf{n}_i)_a - \left\{\varepsilon\kappa^2 e^{-\kappa r_{pi}}\left(\frac{\mathbf{r}_{pi}\cdot\mathbf{n}_i}{r_{pi}^3}\right)S_i\mathbf{r}_{p,a}\right\}. \tag{S15}$$

Here, the S_i stands for the surface area of BE i, $\mathbf{n}_i$ for the normal vector of BE i pointing in the outwards direction, ϵ for the dielectric constant of the continuum, while the part within {} is only included in case of non-zero ionic strength. The same goes for submatrix $\mathbf{C}^{\mathbf{R}}$, which is only present in the case of non-zero ionic strength:

$$\mathbf{C}^{\mathbf{R}}_{\mathbf{r}_{a,p}i} = [(1+\kappa r_{pi})e^{-\kappa r_{pi}} - 1]\frac{\mathbf{r}_{pi,a}}{r_{pi}^3}S_i. \tag{S16}$$

The $\mathbf{B}^{\mathrm{C}}$-matrix is:

$$\mathbf{B}^{\mathrm{C}}_{i\mathbf{r}_{p,a}} = \frac{1}{2\pi(\varepsilon+1)}\frac{\mathbf{r}_{pi,a}}{r_{pi}^3}. \tag{S17}$$

The matrix coupling the BEs consists of:

$$\mathbf{K}_{ij} = \delta_{ij} + (1-\delta_{ij}) \times \frac{\varepsilon(1+\kappa r_{ij})e^{-\kappa r_{ij}} - 1}{2\pi(1+\varepsilon)}S_j\left(\frac{\mathbf{r}_{ij}\cdot\mathbf{n}_j}{r_{ij}^3}\right) \tag{S18}$$

while the **L**-submatrix is again only present for non-zero ionic strength:

$$\mathbf{L}_{ij} = (1-\delta_{ij})\frac{-(1-e^{-\kappa r_{ij}})}{2\pi(1+\varepsilon)}\frac{S_j}{r_{ij}}. \tag{S19}$$

The following three sub-matrices are again only present for non-zero ionic strength (as exemplified by a non-zero κ). The first one is:

$$\mathbf{C}^{\mathbf{C}}_{i\mathbf{r}_{p,a}} = \frac{-\varepsilon}{2\pi(1+\varepsilon)}(-T_{pi}\cdot\mathbf{n}_i)_a. \tag{S20}$$

The **M**-submatrix is formulated as:

$$\mathbf{M}_{ij} = (1-\delta_{ij})\frac{-\varepsilon S_j}{2\pi(1+\varepsilon)} \times \left[(1-(1+Kr_{ij})e^{-Kr_{ij}})((-T_{ij}\cdot\mathbf{n}_j)\cdot\mathbf{n}_i) - K^2 e^{-Kr_{ij}}\left(\frac{\mathbf{r}_{ij}\cdot\mathbf{n}_j}{r_{ij}^3}\right)(\mathbf{r}_{ij}\cdot\mathbf{n}_i)\right]. \tag{S21}$$

Finally, the **N**-matrix is made up from:

$$\mathbf{N}_{ij} = \delta_{ij} + (1-\delta_{ij})\frac{-\varepsilon}{2\pi(1+\varepsilon)}\left[\frac{(1+\kappa r_{ij})}{\varepsilon}e^{-\kappa r_{ij}} - \frac{\mathbf{r}_{ij}\cdot\mathbf{n}_j}{r_{ij}^3}\right] \tag{S22}$$

In the case of zero ionic strength, the **L**- and **M**-submatrices are zero. The **K**- and **N**-matrices then are

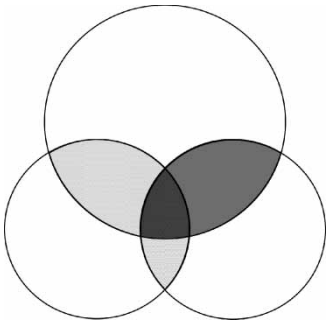

Figure S1. Overlapping spherical atomic volumes.

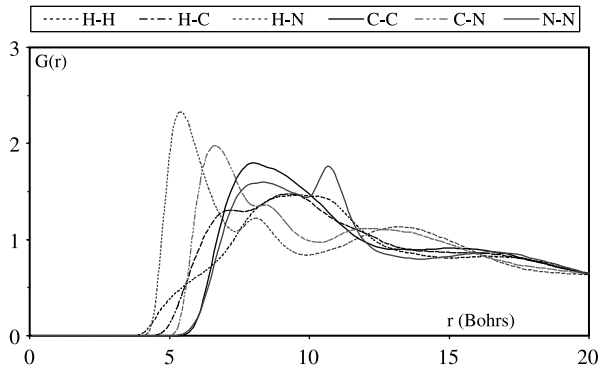

Figure S2. G(r) of several atompairs for liquid acetonitrile (at 298 K).

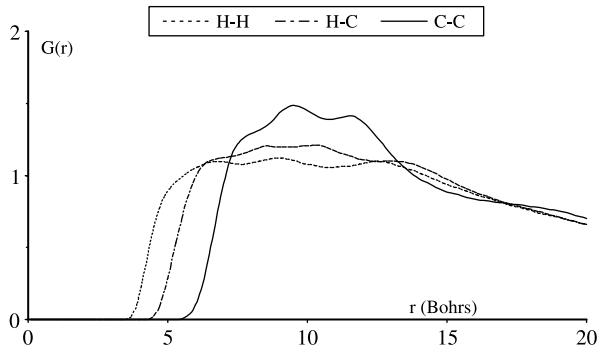

Figure S3. G(r) of several atompairs for liquid benzene (at 298 K).

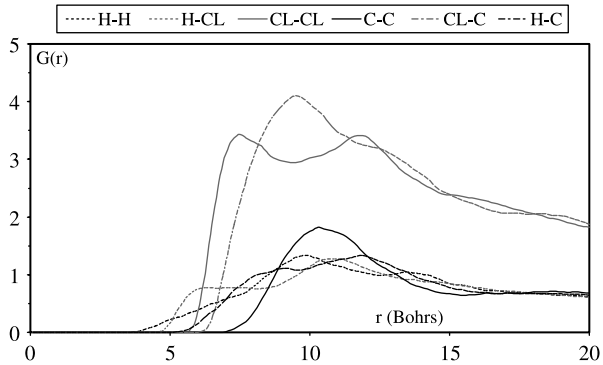

Figure S4. G(r) of several atompairs for liquid chloroform (at 298 K).

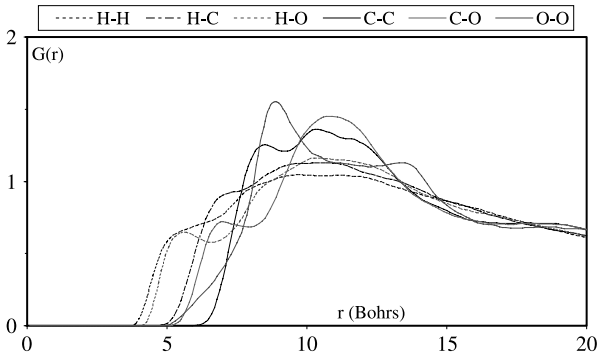

Figure S5. G(r) of several atompairs for liquid dioxan (at 298 K).

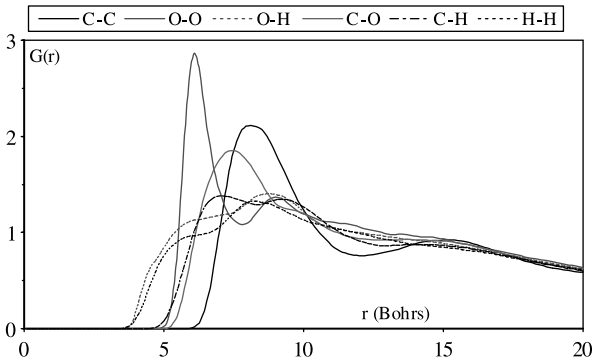

Figure S6. G(r) of several atompairs for liquid methanol (at 298 K).

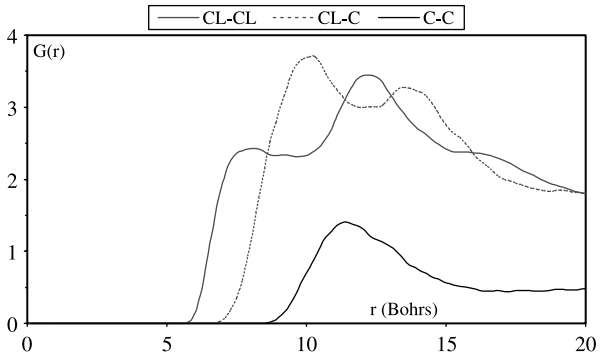

Figure S7. G(r) of several atompairs for liquid tetra (at 298 K).

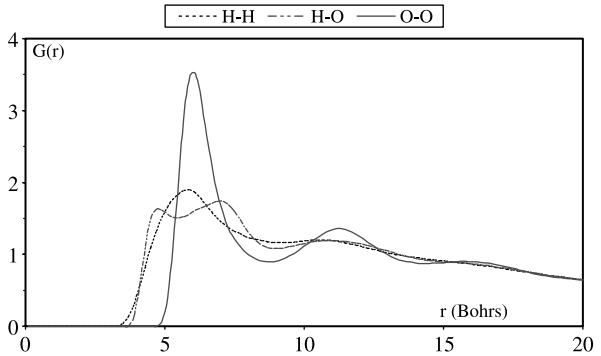

Figure S8. G(r) of several atompairs for liquid water (at 298 K).

related to each other and are two different ways to solve the Poisson equation. The difference between the two is that one uses an inducing potential coupled to a reaction field, while the other uses an inducing field coupled to a reaction potential. The obtained energy is the same in both cases.

These equations are related to the energy. In order to get the gradient, one has to change the potentials V into fields E_a, fields into dipole fields T_{pq}, and so on. The equations have been worked out; they can be straightforwardly obtained from the aforementioned equations. However, they are hardly ever used since only for rather small systems it is worthwhile to include the BEs. Therefore, we do not report explicitly the formulas within this article.

References

[1] G.A. Kaminski, H.A. Stern, B.J. Berne, R.A. Friesner, Y.X.X. Cao, R.B. Murphy, R.H. Zhou, T.A. Halgren. Development of a polarizable force field for proteins via *ab initio* quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.*, **23**, 1515 (2002).

[2] G.A. Kaminski, R.A. Friesner, R.H. Zhou. A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations. *J. Comput. Chem.*, **24**, 267 (2003).

[3] P. Ren, J.W. Ponder. Consistent treatment of inter and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.*, **23**, 1497 (2002).

[4] W.T.M. Mooij, F.B. van Duijneveldt, J.G.C.M. van Duijneveldt-van de Rijdt, B.P. van Eijck. Transferable *ab initio* intermolecular potentials. 1. Derivation from methanol dimer and trimer calculations. *J. Phys. Chem. A*, **103**, 9872 (1999).

[5] L.E. Chirlian, M.M. Francl. Atomic charges derived from electrostatic potentials: a detailed study. *J. Comput. Chem.*, **8**, 894 (1987).

[6] C.M. Breneman, K.B. Wiberg. Determining atom centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, **11**, 361 (1990).

[7] M. Swart, P.Th. van Duijnen, J.G. Snijders. A charge analysis derived from an atomic multipole expansion. *J. Comput. Chem.*, **22**, 79 (2001).

[8] F.C. Grozema, R.W.J. Zijlstra, P.Th. van Duijnen. Many body interactions calculated with the direct reaction field model. *Chem. Phys.*, **246**, 217 (1999).

[9] B.T. Thole. Molecular polarisabilities calculated with a modified dipole interaction. *Chem. Phys.*, **59**, 341 (1981).

[10] P.Th. van Duijnen, M. Swart. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A*, **102**, 2399 (1998).

[11] M. Swart, P.Th. van Duijnen, J.G. Snijders. Mean polarizabilities of organic molecules. A comparison of restricted hartree fock, density functional theory and direct reaction field results. *J. Mol. Struct. (THEOCHEM)*, **458**, 11 (1999).

[12] M. Swart, J.G. Snijders, P.Th. van Duijnen. Polarizabilities of amino acid residues. *J. Comp. Meth. Sci. Engin.*, **4**, 419 (2004).

[13] L. Jensen, P.Th. van Duijnen, J.G. Snijders. A discrete solvent reaction field model within density functional theory. *J. Chem. Phys.*, **118**, 514 (2003).

[14] A.H. de Vries, P.Th. van Duijnen, A.H. Juffer, J.A.C. Rullmann, J.P. Dijkman, H. Merenga, B.T. Thole. Implementation of reaction field methods in quantum chemistry codes. *J. Comput. Chem.*, **16**, 37 (1995).

[15] P.Th. van Duijnen, A.H. Juffer, J.P. Dijkman. Quantum chemistry in the condensed phase: an extended direct reaction field approach. *J. Mol. Struct. (THEOCHEM)*, **260**, 195 (1992).

[16] P.Th. van Duijnen. Embedding in quantum chemistry: the direct reaction field approach. In *New Challenges in Computional Quantum Chemistry*, R. Broer, P.J.C. Aerts, P.S. Bagus (Eds.), pp. 71–84, Dept. of Chemical Physics and Material Science, Groningen (1994).

[17] P.Th. van Duijnen, A.H. de Vries. The direct reaction field force field: a consistent way to connect and combine quantum-chemical and classical descriptions of molecules. *Int. J. Quant. Chem.*, **60**, 1111 (1996).

[18] J. Gao, M.A. Thompson. *Combined Quantum Mechanical and Molecular Mechanical Methods*, ACS, Washington DC (1998).

[19] J.P. Dijkman, P.Th. van Duijnen. Papain in aqueous solution and the role of Asp-158 in the mechanism: an *ab initio* SCF+DRF+BEM study. *Int. J. Quant. Chem., Quant. Biol. Symp.*, **18**, 49 (1991).

[20] P.Th. van Duijnen, B.T. Thole, R. Broer, W.C. Nieuwpoort. Active-site α-helix in papain and the stability of the ion-pair RS^-.ImH^+. *Int. J. Quant. Chem.*, **17**, 651 (1980).

[21] J.A.C. Rullmann, M.N. Bellido, P.Th. van Duijnen. The active site of papain. All-atom study of interactions with protein matrix and solvent. *J. Mol. Biol.*, **206**, 101 (1989).

[22] B.T. Thole, P.Th. van Duijnen, W.G.J. Hol. On the role of the active site α-helix in papain. *Bioph. Chem.*, **9**, 273 (1979).

[23] R.W.J. Zijlstra, P.Th. van Duijnen, A.H. de Vries. Polarization of the excited states of twisted ethylene in a non-symmetrical environment. *Chem. Phys.*, **204**, 439 (1996).

[24] R.W.J. Zijlstra, P.Th. van Duijnen, B.L. Feringa, T. Steffen, K. Duppen, D.A. Wiersma. Excited state dynamics of tetraphenylethene: ultrafast Stoke shift, isomerization and charge separation. *J. Phys. Chem. A*, **101**, 9828 (1997).

[25] R.W.J. Zijlstra, F.C. Grozema, M. Swart, B.L. Feringa, P.Th. van Duijnen. Solvent induced charge separation in the excited states of symmetrical ethylene: a direct reaction field study. *J. Phys. Chem. A*, **105**, 3583 (2001).

[26] P.Th. van Duijnen, F.C. Grozema, M. Swart. Some applications of the direct reaction field approach. *J. Mol. Struct. (THEOCHEM)*, **464**, 193 (1999).

[27] A.H. de Vries, P.Th. van Duijnen. Solvatochromism of the $\pi^* \leftarrow$ n transition of acetone by combined quantum mechanical–classical mechanical calculations. *Int. J. Quant. Chem.*, **57**, 1067 (1996).

[28] F.C. Grozema, P.Th. van Duijnen. Solvent effects on the n to π^* transition of acetone in various solvents. *J. Phys. Chem. A*, **102**, 7984 (1998).

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

[30] A.H. de Vries. Modelling condensed-phase systems. From quantum chemistry to molecular models, PhD thesis Groningen, The Netherlands (1995).

[31] N.L. Allinger, X. Zhou, J. Bergsma. Molecular mechanics parameters. *J. Mol. Struct. (THEOCHEM)*, **312**, 69 (1994).

[32] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179 (1995).

[33] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S.J. Swaminathan, M. Karplus. CHARMM: a program for macromolecular energy, minimization and dynamical calculations. *J. Comput. Chem.*, **4**, 187 (1983).

[34] W.L. Jorgensen, J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic dipeptides and Crambin. *J. Am. Chem. Soc.*, **110**, 1657 (1988).

[35] V. Frecer, (1995), personal communication, 1995.

[36] R.A. Pierotti. A scaled particle theory of aqueous and nonaqueous solutions. *Chem. Rev.*, **76**, 717 (1976).

[37] H.L. Friedman. Image approximation to the reaction field. *Molec. Phys.*, **29**, 1533 (1975).

[38] E.J. Baerends and co-workers. Amsterdam Density Functional (ADF) program, version 2000.02, SCM, Amsterdam (2000).

[39] A.D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **38**, 3098 (1988).

[40] J.P. Perdew. Density-functional approximation for the correlation-energy of the inhomogeneous electron-gas. *Phys. Rev. B*, **33**, 8822 (1986).

[41] S. Toxvaerd. Algorithms for canonical molecular dynamics simulations. *Molec. Phys.*, **72**, 159 (1991).

[42] C.G. Gray, K.E. Gubbins. *Theory of Molecular Fluids. Volume 1: Fundamentals. Appendix D*, Clarendon Press, Oxford (1984).

[43] U. Röthlisberger, M. Parrinello. *Ab initio* molecular dynamics simulation of liquid hydrogen fluoride. *J. Chem. Phys.*, **106**, 4658 (1997).
[44] E. Arunan, H.S. Gutowsky. The rotational spectrum, structure and dynamics of a benzene dimer. *J. Chem. Phys.*, **98**, 4294 (1993).
[45] C. Chipot, R. Jaffe, B. Maigret, D.A. Pearlman, P.A. Kollman. Benzene Dimer: a good model for $\pi-\pi$ interactions in proteins? *J. Am. Chem. Soc.*, **118**, 11217 (1996).
[46] O. Engkvist, P. Hobza, H.L. Selzle, E.W. Schlag. Benzene trimer and benzene tetramer: structures and properties determined by the nonempirical model (NEMO) potential calibrated from the CCSD(T) benzene dimer energies. *J. Chem. Phys.*, **110**, 5758 (1999).
[47] J. Hernandez-Trujillo, M. Costas, A. Vela. Quadrupole interactions in pure nondipolar fluorinated or methylated benzenes and their binary-mixtures. *J. Chem. Soc. Far. Trans.*, **89**, 2441 (1993).
[48] P. Hobza, H.L. Selzle, E.W. Schlag. New structure for the most stable isomer of the benzene dimer. A quantum chemical study. *J. Phys. Chem.*, **97**, 3937 (1993).
[49] P. Hobza, H.L. Selzle, E.W. Schlag. Potential energy surface of the benzene dimer: *ab initio* theoretical study. *J. Am. Chem. Soc.*, **116**, 3500 (1994).
[50] P. Hobza, H.L. Selzle, E.W. Schlag. Structure and properties of benzene-containing molecular clusters: nonempirical *ab initio* calculations and experiments. *Chem. Rev.*, **94**, 1767 (1994).
[51] P. Hobza, H.L. Selzle, E.W. Schlag. Potential energy surface for the benzene dimer. Results of *ab initio* CCSD(T) calculations show two nearly isoenergetic structures: t-shaped and parallel-displaced. *J. Phys. Chem.*, **100**, 18790 (1996).
[52] P. Hobza, V. Spirko, H.L. Selzle, E.W. Schlag. Anti-hydrogen bond in the benzene dimer and other carbon proton donor complexes. *J. Phys. Chem. A*, **102**, 2501 (1998).
[53] R.L. Jaffe, G.D. Smith. A quantum chemistry study of benzene dimer. *J. Chem. Phys.*, **105**, 2780 (1996).
[54] W.L. Jorgensen, D.L. Severance. Aromatic-aromatic interactions: free energy profiles for the benzene dimer in water, chloroform, and liquid benzene. *J. Am. Chem. Soc.*, **112**, 4768 (1990).
[55] V.A. Venturo, P.M. Felker. Intermolecular Raman bands in the ground state of benzene dimer. *J. Chem. Phys.*, **99**, 748 (1993).
[56] M.R. Battaglia, A.D. Buckingham, J.H. Williams. The electric quadrupole moments of benzene and hexafluorobenzene. *Chem. Phys. Lett.*, **78**, 421 (1981).
[57] M. Swart. RanGenConf: a program to generate a random configuration of solvent and solute molecules. Groningen (2002). [Q3]
[58] V. Frecer, S. Miertus. Polarizable continuum model of solvation for biopolymers. *Int. J. Quant. Chem.*, **42**, 1449 (1992).