

# Accurate prediction of bulk properties in hydrogen bonded liquids: amides as case studies

Marina Macchiagodena,<sup>a</sup> Giordano Mancini,<sup>a,b,\*</sup> Marco Pagliai,<sup>a,\*</sup> Vincenzo Barone,<sup>a,b</sup>

In this contribution we show that it is possible to build accurate force fields for small organic molecules allowing the reliable reproduction of a large panel of bulk properties, which are seldom addressed in the same context. Starting from the results obtained in recent studies, we developed a protocol for charge estimation and virtual site generation for the amide class of molecules. The parametrization of electrostatic properties is based on population analysis and orbital localization of quantum mechanical computations rooted into the density functional theory and the polarizable continuum model, without any additional external information. The new protocol, coupled to other recent studies in our group targeted at accurate fitting of internal degrees of freedom, makes available a method for building force fields from scratch (excluding for the moment intermolecular van der Waals interactions) with focus on reproducing the structure and dynamics of hydrogen bonded liquids, yielding results that are in line or better than those delivered by current general force fields. The approach is tested on the demanding series formed by formamide and its two N-methyl derivatives, N-methylformamide and N,N-dimethylformamide. We show that the atomistic structure of the liquids issuing from classical molecular dynamics (MD) simulations employed the new force field is in full agreement with X-ray and neutron diffraction experiments and the corresponding spatial distribution functions are in remarkable agreement with the results of *ab initio* MD simulations. It is noteworthy that the latter result has been never obtained before without using *ad hoc* (and system dependent) scale factors and that, in addition, our parameter free procedure is able to reproduce static dielectric constants over a wide range of values without sacrificing the force field accuracy with respect to other observables. Finally, we are able to explain the trend of static dielectric constants followed by the three amides in terms of properties obtained from the simulations, namely hydrogen bond patterns and reorientational lifetimes.

## Introduction

Faithful reproduction of bulk liquid properties from atomistic simulations hinges on the accuracy of the model describing the interactions between the system's components. Proper boundary conditions are enforced in order to prevent finite size effects (while keeping computational cost within reasonable limits).<sup>1</sup> In the framework of classical Molecular Dynamics (MD) (and Monte Carlo, MC) simulations the reliability of the results is primarily related to the availability of an accurate *Force Field* (FF), which defines a hierarchy of interactions at the atomic and molecular level and a set of functional forms quantifying these interactions.<sup>2</sup> Force Fields are usually based on a set of molecular "templates", whose geometries are (most often) obtained from Quantum Mechanical (QM) calculations; some of the FF parameters may be further tuned in order to better reproduce selected experimental data (such as density, enthalpy of vaporization, isothermal compressibility, heat capacities, surface tension, static dielectric constant, *etc.*,<sup>3,4</sup>). As a matter of fact, FF's can be devised either with the aim of describing in a reasonable way a large panel of systems,<sup>5,6</sup> or with the purpose of reproducing with high accuracy some properties of a reduced (even just a single) number of molecular systems.<sup>7-9</sup>

At present, the most widely adopted FF in molecular mechanics (MM) simulations are geared toward wide, but well defined classes

of chemical systems including proteins and biological membranes,<sup>10-13</sup> or a class of mineral systems,<sup>14,15</sup> polymers, organic liquids *etc.*<sup>3,4</sup> Among the goals that the development of these FF strives to achieve there is (limited) transferability i.e. the possibility of describing a never simulated molecule using model atoms and interactions based on the relation between the atomic groups of the molecule under investigation and other species present in the FF library. Some of the observables used as benchmarks are particularly troublesome; for example, a comparative study carried out on  $\approx 150$  small organic molecules<sup>16</sup> using three very widespread FFs highlighted how properties such as surface tension ( $\gamma$ ) and static dielectric constant ( $\epsilon$ ) (dubbed "the hardest nut to crack"<sup>16</sup>) were often poorly reproduced. Although some of those results were used to address the limits in the benchmark simulations<sup>17</sup> nevertheless the error in reproducing  $\epsilon$  represents a serious limit for the accuracy of MM studies. It is worth to observe that the target of computational biology and medicinal chemistry studies based on simulation or docking techniques (based on this type of FF) is often represented by protein-ligand or nucleic acid-ligand complexes with binding sites immersed in matrices, whose dielectric constants are strongly different from that of bulk water (which is adequately reproduced with contemporary water models).<sup>18,19</sup> Thus,  $\epsilon$  should be included (whenever possible) in the set of reference bulk properties used to validate any FF. The simplest description of electrostatic interactions is based on the Coulomb law and employs fixed point charges (FX) located on nuclei and/or on so called "virtual sites" or "dummy-atoms".<sup>20</sup> Although polarizable force-fields are being increasingly developed, using a variety of approaches such as oscillating dipoles,<sup>21,22</sup> or fluctuating charges (FQ),<sup>23,24</sup> simple point charges are still employed in the most widespread classical force fields, allowing sufficiently reliable results for many applications. It is, of course, true that a polarizable approach may be able to capture some features of

<sup>a</sup> Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, 56126, Pisa, Italy.

<sup>b</sup> Istituto Nazionale di Fisica Nucleare (INFN) sezione di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy.

\* E-mail: [giordano.mancini@sns.it](mailto:giordano.mancini@sns.it); [marco.pagliai@sns.it](mailto:marco.pagliai@sns.it)

Electronic Supplementary Information (ESI) available: Force Fields for formamide, N-methylformamide and N,N-dimethylformamide. Cluster analysis. Autocorrelation functions. Comparison between rigid and flexible formamide models. Atomic charges and VSs positions for water solution simulations. See DOI: 10.1039/x0xx00000x

the studied system that are beyond the “resolution” of its non-polarizable counterpart even when both models are steered towards the same results.<sup>25</sup> However, Wang et al.<sup>19</sup> have recently shown that a properly parameterized fixed charge approach can be very successful in reproducing, for instance, most bulk properties of water, which are by no means simple targets. In addition, from our point of view, fixed and polarizable FFs are not mutually exclusive since we are interested in developing multiscale applications<sup>25–27</sup> in which the outer layers are represented by fixed charge molecules. Several approaches do exist to assign simple point charges to atoms, based for instance on the fitting the QM level electrostatic potential (ESP).<sup>28</sup> In particular, CM5 charges are well suited for MM applications since they are robust with respect to variations due to conformer equilibria and to the specific QM method used in the fitting.<sup>29</sup>

In this contribution we extend the capability of CM5 based FFs using a self-consistent approach, based on the Boys<sup>30</sup> orbital localization procedure. The centroids of the localized Boys orbital have been adopted to assign virtual sites to atoms carrying two lone pairs (such as  $sp^2$  oxygen). The centroids of localized orbitals have been calculated with the approach described by Vidossich et al.<sup>31</sup> In FFs employing fixed charges, atoms and molecules are usually “overpolarized” with respect to results of gas-phase QM calculations in order to account for an “average” polarization of a condensed phase environment. In recent studies,<sup>29,32</sup> Jorgensen and co-workers showed that for CM1A and CM5 charges, empirical scale factors of 1.14 and 1.27 respectively lead to highly improved results for pure liquids and explicit water solutions; they also showed that an average scaling factor of 1.20 introduces tolerable errors, being thus suitable for most situations.<sup>29</sup> However, the degree of success of this approach varies across the benchmark set, yielding results of less than average quality for molecules carrying nitrogen nearby carbonyl groups atoms such as N-methylacetamide and N,N-dimethylacetamide. Our purpose is to show that, although overpolarization is important, a simple scale factor may fail to capture the features of systems in which directional interactions (such as hydrogen bonds) play an important role. Virtual sites (VSs), on the other hand, offer implicit (by splitting and moving further apart point charges) and more fine grained (only some atoms are involved) overpolarization, as shown also recently by Harder et al.<sup>33</sup> in the derivation of the OPLS3 FF.

On these grounds, we tried to derive accurate FFs for molecules representing problematic cases for the scale factor approach. In particular, we selected the homologous series formed by formamide (HCO-NH<sub>2</sub>), N-methylformamide (HCO-NHCH<sub>3</sub> or NMF hereafter) and N,N-dimethylformamide (HCO-N(CH<sub>3</sub>)<sub>2</sub> or DMF hereafter), adding virtual sites to model the lone pairs of the carbonyl oxygen atom. In addition representing difficult case studies *per se* and being formamide a model compound in biological and pre-biological studies, these molecules point out some difficulties underlying parametrization procedures: although they feature a relatively simple, very similar chemical structure, formamide and its N-methyl substitutes have very different properties in terms of liquid structure<sup>34</sup> and macroscopic observables ( $\epsilon$  ranges from 37 in DMF to  $\approx 180$  in NMF).<sup>35</sup> We show that the proper amount of overpolarization is automatically obtained in a self-consistent way if the charges of atoms in the functional group carrying the VS are

adjusted in order to obtain the original dipole moment. Furthermore, we show that performing calculations in implicit solvent<sup>36,37</sup> takes into account most of the needed overpolarization and good results may be obtained with a very small scale factor. We point out that use of VS is not based on additional empirical parameters, but only in an accurate fitting of QM level properties.

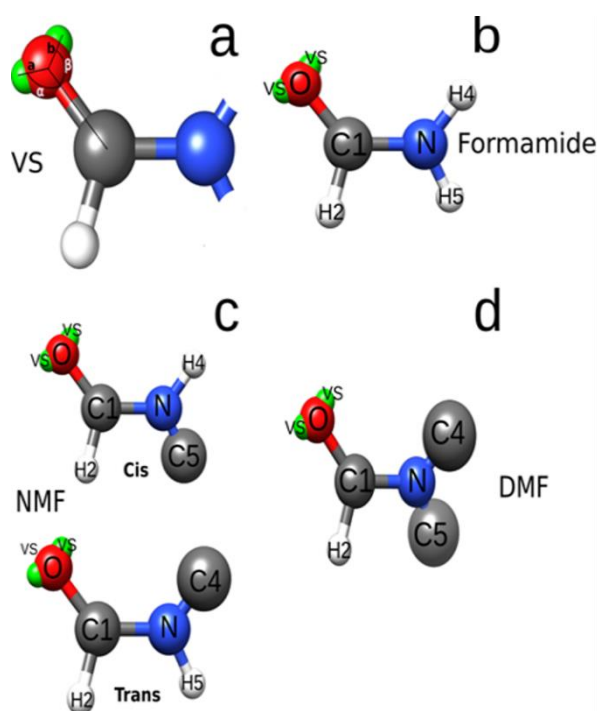
FFs tuned towards the nanoscopic structure<sup>34</sup> or bulk properties<sup>38</sup> for HCO-NX<sub>2</sub> (X=H and/or CH<sub>3</sub>) molecules have been proposed also quite recently but, in addition to being based on empirically derived charges, these models failed to reproduce some of the cited properties. To assess the results obtained with our FF we performed extensive MD simulations, investigating the nanoscopic liquid structure of the three organic liquids as well as the level of reproduction of key thermodynamic properties. We show that our FF is able to accurately reproduce X-ray and neutron diffraction results,<sup>39,40</sup> together with key bulk properties (density, heat of vaporization and static dielectric constant).

The paper is organized as follows: in section Computational methods we review the various steps of the force field generation, including the assignment of point charges and lone pairs, and we explain how the analysis of trajectories and calculation of bulk properties were carried out in order to validate results. Then, section Results illustrates the study on HCO-NH<sub>2</sub> and related compounds. The results of the procedure are discussed in detail in section Discussion, and compared to other recent studies. A few concluding remarks are given in section Conclusions.

## Computational methods

**QM calculations.** All QM calculations needed to build the FF presented in this work were performed using Density Functional Theory (DFT) at the B3LYP/6-31+G(d) level of theory with the Gaussian-09 package.<sup>41</sup> Bulk solvent effects have been taken into account by means of the Conductor-like Polarizable Continuum Model (C-PCM)<sup>37,36</sup> setting the reference solvent and imposing the value of the scaling factor for the sphere radius ( $\alpha$ ) to 1.05. Atomic charges were then estimated using population analysis based on Charge Model 5 (CM5)<sup>42</sup> verifying the basis set effect as reported in Table S1 of Electronic Supporting Information.

The parameters for the FF internal degrees of freedom are derived by fitting optimized energies, gradients and Hessian matrices with the procedure implemented in JOYCE,<sup>43</sup> using the ULYSSES graphical user interface;<sup>44</sup> *ab initio* torsional profiles were built at the B3LYP/6-31+G(d) (C-PCM) level of theory. Lennard-Jones parameters were taken from OPLS.<sup>3,4</sup> Virtual sites were introduced as the centroids of localized molecular orbital positions for the  $sp^2$  oxygen atom using the Boy localization procedure,<sup>30,31</sup> as shown in Figure 1. The VSs charge was obtained as follows: first, the oxygen atom charge was transferred to the VSs; then the charges of all the sites belonging to the carbonyl moiety (including VSs) were adjusted to retain the molecular dipole moment. Therefore the molecular dipole moment in presence and in absence of VSs is the same. The two virtual sites (sites 7 and 8 in Table S2 of Electronic Supporting Information; see also Figure S1) lie in the plane determined by atoms O, C1, H2 and O, C1, N respectively (Figure 1b). The position of VSs during the MD simulations were constrained using fixed distances from the oxygen atom and angles ( $a$ ,  $b$ ,  $\alpha$  and  $\beta$  in Figure 1a and Table 1).



**Figure 1:** Ball and stick representation of the studied molecules, showing the position of virtual sites for the  $sp^2$  oxygen and the atom labeling used in the study. (a) Geometry of virtual sites (all systems). (b) Representation of HCO-NH<sub>2</sub>, (c) Representation of HCO-NHCH<sub>3</sub> (upper panel: cis; lower panel: trans). (d) Representation of HCO-(CH<sub>3</sub>)<sub>2</sub>.

**Table 1:** Values of O-VS distances (a,b) and C-O-VS ( $\alpha$  and  $\beta$ ) angles used to generate the position of VSs in formamide, N-methylformamide (cis and trans conformers), N,N-dimethylformamide molecules.

	a (Å)	b (Å)	$\alpha$ (°)	$\beta$ (°)
HCO-NH <sub>2</sub>	0.312	0.316	111.195	110.518
HCO-NHCH <sub>3</sub> (cis)	0.311	0.317	110.901	110.526
HCO-NHCH <sub>3</sub> (trans)	0.313	0.314	111.107	110.827
HCO-N(CH <sub>3</sub> ) <sub>2</sub>	0.311	0.314	109.838	111.215

**MD simulations.** Classical MD simulations were carried out using GROMACS v. 4.6.5.<sup>45</sup> Methyl groups in NMF and DMF were described using a united atom model. The cross interactions for Lennard-Jones terms are calculated using the Lorentz-Berthelot<sup>46,47</sup> mixing rules and we have excluded intramolecular interactions between atom pairs separated up to three bonds. Simulations were performed on system of 500 molecules in a cubic box with periodic boundary conditions. Systems were initially brought to 0 K with a steepest descent minimization and then heated to 298.15 K a NVT ensemble using the velocity-rescaling method<sup>48</sup> with an integration time step of 0.2 fs

and a coupling constant of 0.1 ps for 10<sup>6</sup> simulation steps (200 ps). The time step and temperature coupling constant were then increased to 2.0 fs and 0.2 ps, respectively, and systems were let to converge to uniform density in a NPT ensemble (using the Parrinello-Rahman barostat and a coupling constant of 1.0 ps) for 10 ns (20 ns for N-methylformamide). Production runs in the NVT<sup>48</sup> ensemble were then carried out and used for analysis using either flexible bonds ( $\delta t=0.2$  fs) or fixing the fastest internal degrees of freedom by means of the LINCS algorithm ( $\delta t=2.0$  fs).<sup>49</sup> The total sampling time for these runs was 100 ns (formamide and N,N-dimethylformamide) or 300 ns (N-methylformamide) depending of the time needed for the total dipole autocorrelation function (see below) to reach a stable value. Electrostatic interactions are treated the using particle-mesh Ewald (PME) method<sup>50</sup> with a grid spacing of 1.2 Å and a spline interpolation of order 4. Furthermore, since the dipole moment fluctuations are sensitive to the boundary conditions applied<sup>51</sup> we have fixed the external dielectric constant to the experimental value for each of the studied system.

**Ab initio MD.** The *ab initio* molecular dynamics (AIMD) simulation was carried out using the CP2K program.<sup>52</sup> We considered a cubic box of size 14.88 Å containing 50 formamide molecules and subjected to periodic boundary conditions. The system was preliminary thermalized at 300.0 K. We performed a simulation of 6 ps in the NVE ensemble with a time step of 0.2 fs. The electronic structure was calculated with density functional theory, utilizing the BLYP<sup>53,54</sup> functional. The TZV2P<sup>55</sup> basis set was used in conjunction with GTH<sup>56,57</sup> pseudopotentials. A plane wave cutoff of 340 Ry was adopted for electron density. All hydrogens have been replaced with deuterium atoms.

**Analysis of trajectories.** Radial (rdf) and spatial (sdf) distribution functions were calculated using TRAVIS package.<sup>58</sup> To analyze hydrogen bond dynamics we preferred the continuous function devised by Pagliai et al.<sup>59,60</sup> over a simpler geometrical criterion. According to this method, each donor-acceptor pair can be considered a hydrogen bond only partly, with a “score” between 0 and 1 given by:

$$f_{\text{HB}}(r, \theta) = g(r; r_e, r_{hw}) \times g(\theta; \theta_e, \theta_{hw}) \quad \text{Eq. 1}$$

Where  $r$  is the acceptor-hydrogen distance,  $\theta$  the acceptor-donor-H angle, and

$$g(x; x_e, x_{hw}) = \begin{cases} 1 & \text{if } x \leq x_e \\ \frac{-(x-x_e)^2}{e^{2x_{hw}^2}} & \text{if } x > x_e \end{cases} \quad \text{Eq. 2}$$

Here  $x_e$ , and  $x_{hw}$  are the maximum and half-width values of either the not normalized radial or angular distribution function, obtained by a fit of the corresponding data. Hydrogen bond lifetimes,  $\tau_{\text{HB}}$ , were calculated using the following definition:<sup>61</sup>

$$\tau_{\text{HB}} = \sum_{n=0}^{\infty} \frac{1}{2} \Delta t [H(t_n) + H(t_n + 1)] \quad \text{Eq. 3}$$

The function  $H(t_n)$  is obtained according to:

$$H(\tau_m) = \sum_{n=m}^{\infty} N(t_{n+1}) / \sum_{n=1}^{\infty} N(t_n) H(0) = 1 \quad \text{Eq. 4}$$

Where  $N(t)$  is an ordered histogram containing the lifetimes of hydrogen bonds. Reorientational autocorrelation functions (ACF),<sup>62</sup> defined as:

$$C_l^\alpha(t) = \langle P_l(u^\alpha(t) \cdot u^\alpha(0)) \rangle \quad \text{Eq. 5}$$

where  $P^l$  is the  $l^{\text{th}}$  rank Legendre polynomial, and  $u^\alpha$  is a unit vector in a given direction at time  $t$ . By fitting  $C_l(t)$  to an exponential function,  $C_l(t) = e^{-t/\tau_l}$ , it is possible to determine the correlation time,  $\tau_l$ , defined as the time needed by a molecule to rotate about  $u$ ; we used  $l=1$  and the vector defined by the C–O carbonyl bond or the normal to the plane spanned by the C–O and C–N vectors.

The total dipole moment autocorrelation function was used for the estimation of  $\epsilon$ , according to the following equation:

$$\epsilon = 1 + \frac{4\pi}{3Vk_B T} (\langle M^2 \rangle \langle M \rangle^2) = \sum_i \mu_i \quad \text{Eq. 6}$$

Where  $\mu_i$  is the molecular dipole moment. To calculate  $\epsilon$  we eliminated the first 10 ns of simulation (20 ns for N-methylformamide).

Heats of vaporization were calculated from the equation:

$$\Delta H_{\text{vap}} = E_{\text{intra}}(\text{g}) - (E_{\text{intra}}(\text{l}) + E_{\text{inter}}(\text{l})) + RT \quad \text{Eq. 7}$$

where  $E_{\text{intra}}$  is the intramolecular energy in either gas (g) phase or the liquid (l) phase and  $E_{\text{inter}}$  represents the intermolecular energy. To calculate the heat of vaporization,  $\Delta H_{\text{vap}}$ , gas-phase simulations of 5 to 10 ns ( $\delta t = 0.2$  fs) have been added. For N-methylformamide, the presence of the conformers (cis/trans) has been taken into account, simulating two gas-phases and considering, for the calculation of the  $\Delta H_{\text{vap}}$ , their relative abundance.

A total of 1000 configurations, with a constant spacing of 1 ps, was sampled from the last part of flexible simulations. These configurations were centered around the first residue and then, taking as a cut-off radius the first minimum of center of mass (COM-COM) radial distribution functions, trajectories were sorted in order to select the first  $n$  (with  $n=18$  in all systems) neighbors of residue one in each system. Clustering was then performed on the resulting selections, applying the GROMOS<sup>63</sup> method using the pairwise Root Mean Square Deviations (RMSD) as a measure of distance between two molecules, after a mass-weighted least square fitting to the first configuration:

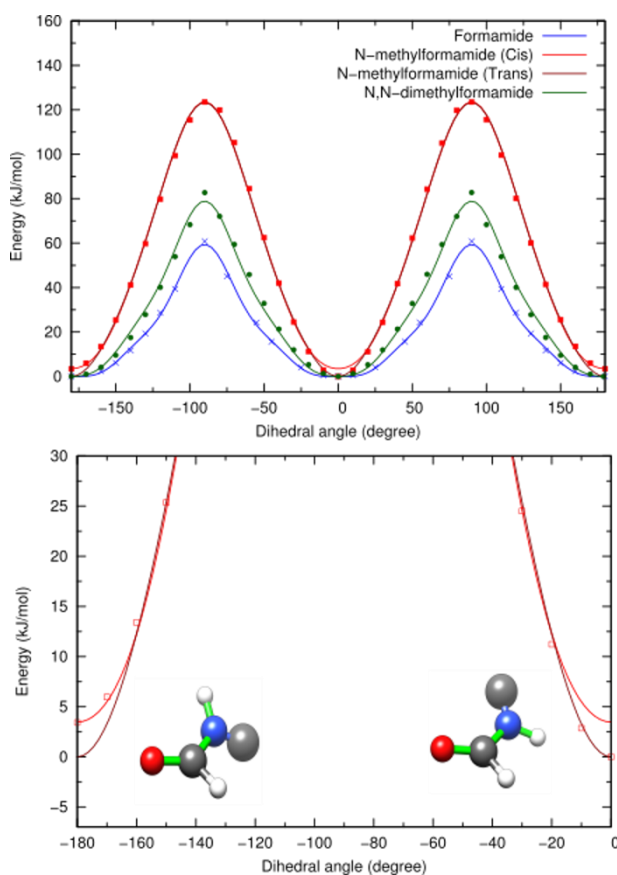
$$\text{RMSD}(t_1, t_0) = \sqrt{\frac{1}{M} \sum_i m_i \|r_i(t_1) - r_i(t_0)\|^2} \quad \text{Eq. 8}$$

Where  $M$  is the sum of atomic masses,  $m_i$  is the atomic mass of  $i$  species and  $t_0$  refers to the first configuration; distance calculations was restricted to the carbonyl and nitrogen atom. A cut-off corresponding to the maximum value observed in the pair-wise distance histogram (Figure S2a of Electronic Supporting Information) was used to assign configurations to clusters.

Analysis were performed using standard GROMACS<sup>45</sup> tools or in-house written code, taking advantage of the NumPy and SciPy libraries.<sup>64</sup> Graphs were prepared with gnuplot,<sup>65</sup> Chimera<sup>66</sup> or VMD.<sup>67</sup>

## Results

**Force field generation.** The first step for building the new FF for the investigated amides was the parametrization of the amine-carbonyl rotation to *ab initio* data. Figure 2 shows the QM torsional profiles and corresponding FF functions for HCO-NH<sub>2</sub> and related compounds calculated using JOYCE; full details about the FF internal degrees of freedom are given in Table S2 in the Electronic Supporting Information.



**Figure 2:** Torsional profile comparison between QM (points) and JOYCE (lines) for formamide (blue), N-methylformamide (light and dark red) and N,N-dimethylformamide (green). Panel (a): complete torsional profiles. (b): detail of NMF profile showing the energy difference arising from the cis/trans asymmetry. The results for cis and trans conformers of NMF are shown with light red and dark red lines, respectively.

**Table 2:** Atomic charges for studied molecules calculated at the B3LYP/6-31+G(d) level (CM5) and used in the simulations (adjusted). C4/C5 refers to united atom methyl group for N-methylformamide and N,N-dimethylformamide.

q(e)								
Atom	HCO-NH <sub>2</sub>		HCO-NHCH <sub>3</sub>				HCO-N(CH <sub>3</sub> ) <sub>2</sub>	
	CM5	Adjusted	CIS		TRANS		CM5	Adjusted
			CM5	Adjusted	CM5	Adjusted		
C1	0.191055	0.154544	0.183870	0.146569	0.185484	0.148973	0.182264	0.146489
H2	0.131902	0.131902	0.127781	0.127781	0.127947	0.127948	0.124366	0.124365
N	-0.566816	-0.566816	-0.429268	-0.429268	-0.430660	-0.430660	-0.298481	-0.298481
H4	0.339036	0.339036	0.339814	0.339814	0.345317	0.345317	--	--
H5	0.345109	0.345109	--	--	--	--	--	--
C4	--	--	--	--	--	--	0.224764	0.224764
C5	--	--	0.230849	0.230849	0.211743	0.211743	0.207582	0.207582
O	-0.440286	0.000000	-0.453046	0.000000	-0.439832	0.000000	-0.440493	0.000000
VS1	0.000000	-0.201887	0.000000	-0.207872	0.000000	-0.201661	0.000000	-0.202359
VS2	0.000000	-0.201887	0.000000	-0.207872	0.000000	-0.201661	0.000000	-0.202359

Then, we proceeded to assign new atomic charges according to the results of the CM5<sup>42</sup> method and finally, virtual sites were placed to mimic the position of the oxygen atom lone-pairs. The initial and adjusted CM5 charges of HCO-NH<sub>2</sub>, HCO-NHCH<sub>3</sub> and HCO-N(CH<sub>3</sub>)<sub>2</sub> atoms are shown in Table 2. Table 3 includes available experimental estimates for the molecular dipole moment for the studied species and the computed value obtained from *ab initio* calculations. The latter show the polarization effect due to the presence of the implicit solvent as compared to gas phase calculations; note that the C-PCM value was used to adjust the atomic charges shown in Table 2. For comparison, in the third column, we have also shown the experimental values for each system. It is worth to observe that HCO-NX<sub>2</sub> molecules present very similar dipole moments but very different dielectric constants (Table 5).

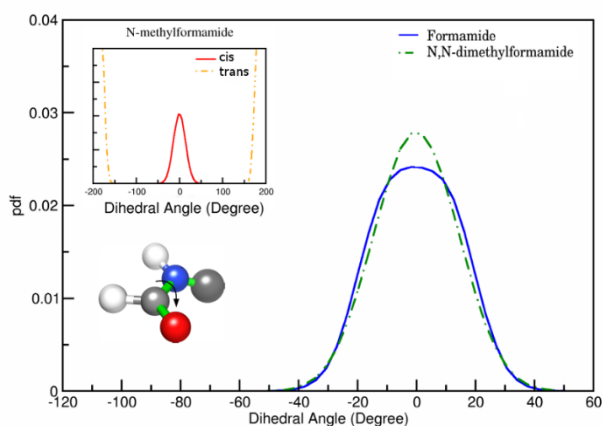
Having obtained a FF for the three molecules we proceeded with MD simulations using the protocol illustrated in section Computational methods. For each molecule two "production runs" were actually carried out using either flexible bonds ( $\delta t=0.2$  fs) or rigid ones ( $\delta t=2.0$  fs). N-methylformamide, exists in both cis and trans form (aminic hydrogen atom respect to the oxygen atom); the less stable cis conformer is estimated to represent between 6 and 10%<sup>68</sup> of molecules in solution and, accordingly, multiple tests were performed varying the relative amount of the two compounds. In particular, we run MD simulations at three different concentrations of trans/cis conformers: 94:6, 92:8 and 90:10; in the following, only the results relative to the 94:6 concentration are discussed.

**Table 3:** Experimental and computed molecular dipole moment  $\mu$  (D) for formamides as obtained from *ab initio* calculations using G09. The Gas-phase and C-PCM columns refer to the absence of presence of solvent contributions.

	$\mu$ (D)		
	Experimental <sup>b</sup>	Computed	
		Gas phase	C-PCM
HCO-NH <sub>2</sub>	3.73±0.07	3.68	4.63
HCO-NHCH <sub>3</sub>	3.83±0.08	3.70 (4.08) <sup>a</sup>	4.78 (5.55) <sup>a</sup>
HCO-N(CH <sub>3</sub> ) <sub>2</sub>	3.82±0.08	3.98	5.19

<sup>a</sup>The value is relative to the trans conformer. For the cis conformer, the dipole moment is reported in brackets. <sup>b</sup>Ref<sup>23</sup>.

**Pure liquid structure.** It is worth to observe that most FF derived for HCO-NX<sub>2</sub> molecules employ planar geometries which fix the carbonyl-amine dihedral angle; here we used a flexible model with *de novo* set of parameters for internal degrees of freedom; thus, we first verified that a correct geometry was generated in the three simulations. Figure 3 shows the histogram of the O-C-N-H(or C) dihedral angles (for both conformers in the case of NMF) obtained from flexible simulations; in all cases a normal distribution of the dihedral angle with an average planar molecule and a standard deviation of about 10° is obtained.



**Figure 3.** Histogram for formamide O-C-N-H dihedral angle (full blue lines) and N,N-dimethylformamide O-C-N-C dihedral angle (green dot-dashed line). In the insert the O-C-N-H histogram for the cis (full red line) and trans (orange dot-dashed line) conformers of N-methylformamide.

To describe the liquid structure, radial distribution functions ( $g(r)$  or rdf hereafter) were calculated from flexible simulations. The  $g(r)$  between the molecules centers of mass (Com), N $\cdots$ O pairs and O $\cdots$ H2 (aldehydic) pairs are shown in Figure 4 (atom labeling Figure 1b,c,d). The rdf right shift between molecular centers of mass, from formamide to N,N-dimethylformamide (from 4.5 Å to 5.5 Å) and broadening of the  $g(r)$  first peak (from 2.9 Å to 3.5 Å) illustrates the decrease in density observed when inserting methyl groups (see Table 5), in agreement with literature data and with a more random distribution of molecules found in N,N-dimethylformamide.<sup>69</sup>

The O $\cdots$ N (middle panel) rdf shows the change in the liquid structure that take place upon loss of carbonyl–amine hydrogen bonds: while the first peak for HCO-NH<sub>2</sub> and HCO-NHCH<sub>3</sub> have comparable position and height, the first peak of HCO-N(CH<sub>3</sub>)<sub>2</sub>, is located further away at 4.5 Å; it is also worth to observe that the latter's position compares well with HCO-NHCH<sub>3</sub> second peak and it is very likely related to the interposition of a methyl group between N and O. The first peak of the HCO-NH<sub>2</sub>  $g_{O\cdots N}(r)$  is centered around 2.9 Å in agreement with experimental value obtained using different techniques.<sup>39,40</sup> N-methylformamide shows a small left shift (first maximum at 2.8 Å) that can be correlated with a more strong hydrogen bond due to stabilizing presence of the electron-donating methyl group and thus to a more basic character of the oxygen atom. The  $g_{O\cdots N}(r)$ , in HCO-NHCH<sub>3</sub> and HCO-NH<sub>2</sub> indicate a strong hydrogen bond which is supported by resonance formulas where the nitrogen atom presents a positive and the oxygen atom a negative charge. The stabilizing effect of N-methyl substitutions on resonance is also confirmed by the increase of torsional energy profile reported in Figure 1.

The rightmost panel shows also that the loss of N-O HBs tends to

increase the relative strength of O $\cdots$ H2 interaction as demonstrated by the sharp and defined rdf first peak of HCO-N(CH<sub>3</sub>)<sub>2</sub> (located at 2.25 Å) as compared to HCO-NHCH<sub>3</sub>, having a much broader peak at 3.0 Å, and to the less defined structure featured by HCO-NH<sub>2</sub>.

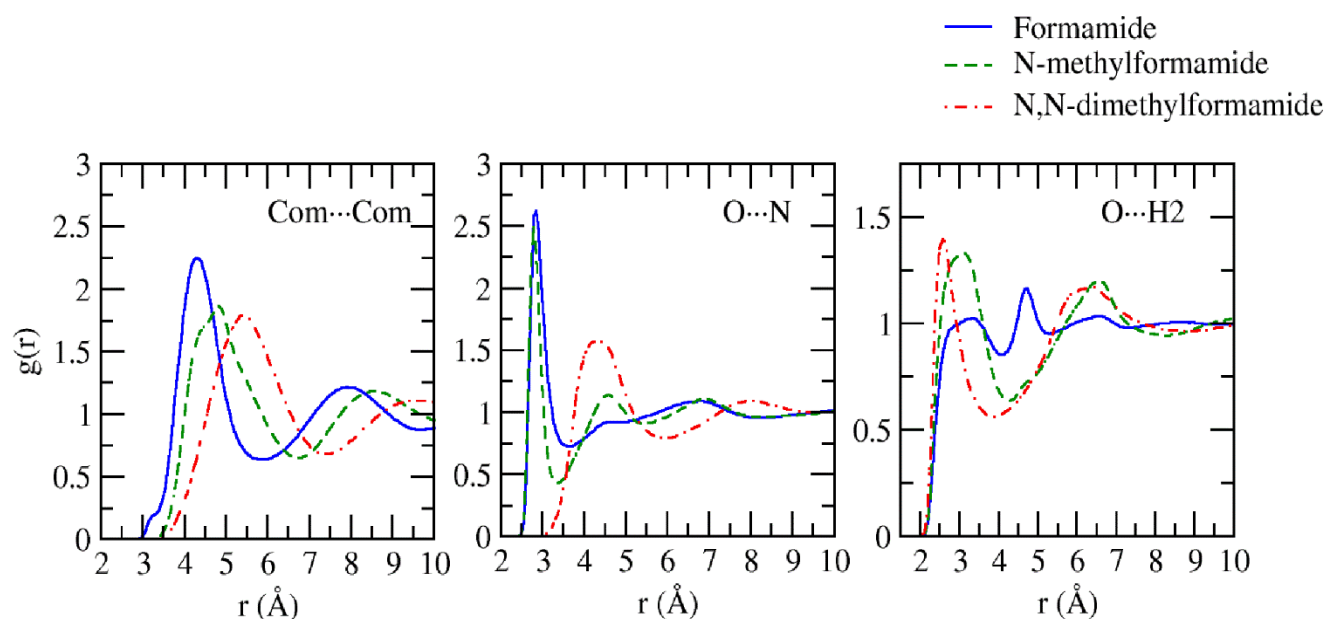
Overall, the results of for all  $g(r)$  show that with our FF the obtained results are very similar to those reported by Jorgensen and Swenson<sup>70</sup> and by Cordeiro,<sup>34</sup> the only difference being the relative height of the  $g(r)$  peaks which are located at the same position.

The  $g_{O\cdots HX}(r)$  (H4/H5 for HCO-NH<sub>2</sub> and HCO-NHCH<sub>3</sub>, H2 for HCO-N(CH<sub>3</sub>)<sub>2</sub>) integral up to the first minimum has been calculated and a coordination numbers of 2.01, 1.02 and 1.29, respectively, have been obtained. To better visualize and to summarize the interactions, spatial distribution functions (sdf) of oxygen atoms in formamide and N-methylformamide are shown in Figure 5: clearly, the main interaction involves the aminic hydrogens.

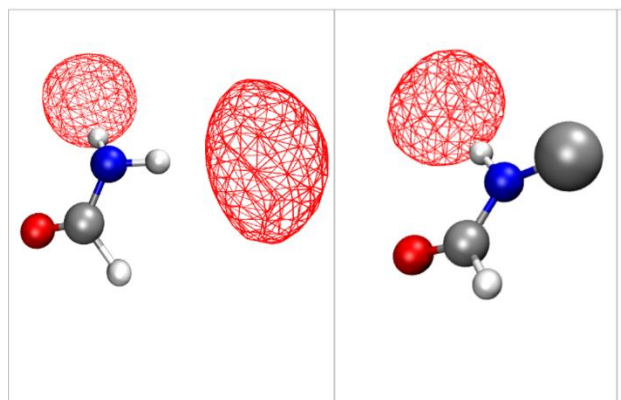
To further understand the HB orientation in the three systems, in Figure 6 the surfaces representing the  $g(r, \theta)$  weighted by  $f_{HB}$  function (see section Computational methods) are shown. From left to right, we reported the results for formamide (column 1), NMF (column 2) and DMF (column 3). From top to bottom we report the  $g(r, \theta)$  functions corresponding to the H4/C4 (row a), H5/C5 (row b) and H2 atoms (row c).

Rows a and b clearly illustrate the strong hydrogen bonds present in HCO-NH<sub>2</sub> and HCO-NHCH<sub>3</sub>, at  $\approx 1.8$  Å and 10° in agreement with the results shown in Figure 4. It is worth to observe the higher intensity presented by the NMF single bond, favoured by the electron-donating effect of the methyl group. Panels 1a and 1b show that the interaction between the two aminic hydrogen and the oxygen atoms results slightly different. The reason is the tiny difference of hydrogen atoms charges (Table 2) that we intentionally choose do not equalize, since they are immersed in a different chemical environment. The results shown in the third row allow to describe in details the improper hydrogen bond between the aldehydic hydrogen and the oxygen atom. It constitutes the only (scrolling the rightmost column 3) strong interaction observed in HCO-N(CH<sub>3</sub>)<sub>2</sub>, presenting modal distance and angular values of 2.6 Å and 12°, respectively. Note that the corresponding distribution for NMF presents a similar shape and modal values but with a lower intensity. Formamide, on the other hand, features a much more disordered distribution centered at 3.4 Å.

To get a deeper understanding of molecules arrangement and of the hydrogen bonding network shape in the three systems, we performed a cluster analysis of trajectories using the first coordination shell of the first residue as the input structure. All details about the clustering procedure are found in section Computational methods. Figure S2a (Electronic Supporting Information) shows the pairwise distance histograms, used in the selection of cut-off ranges and the obtained cluster sizes (Figure S2b); as can be observed the first clusters accounts for a reasonable amount of the total variance, including a minimum of 785 frames; four clusters were obtained in all cases.



**Figure 4.** Radial distribution functions between selected sites for amide simulations. From left to right: Com...Com, O...N and O...H2  $g(r)$  from pure liquid simulations of formamide (full blue lines), N-methylformamide (green dashed lines) and N,N-dimethylformamide (red dot-dashed lines).



**Figure 5:** Isosurface of the oxygen atom around the molecule of formamide (left) and N-methylformamide (right) at an isovalue of  $35 \text{ nm}^{-3}$ .

Figure S3 (Electronic Supporting Information) shows the first clusters obtained for the three studied systems; as can be easily observed, formamide and NMF tend to form hydrogen bond pairs or triplets in which the dipole moment forms a small angle (the considered molecules are approximately in the same plane); on the other hand such conformation is impossible for DMF, where the improper HB between the carbonyl oxygen and hydrogen atoms orients dipole moments in orthogonal arrangement, thus contributing to explain the different dielectric constant featured by three molecular species as opposed to their comparable molecular dipole moments.

**Dynamical properties.** Having investigated the structural properties we proceeded with comparing the dynamics of the studied organic liquids. Table 4 shows HB lifetimes and (rotational) relaxation times (see section Computational methods) as obtained from the last 2 ns of flexible simulations. We calculated the hydrogen bond lifetime for N-H4...O (and N-H5...O for formamide) and C-H2...O interactions, using a threshold of  $1.0\text{E-}4$  to obtain a discrete signal from the  $f_{HB}$

function. In addition, the relaxation time ( $\tau_{\mu}$ ) from the molecular dipole autocorrelation function and the rotational relaxation time of a vector along the carbonyl bond ( $\tau_{CO}$ ) and of the normal to the H-C-O plane ( $\tau^{\perp}$ ) were estimated.

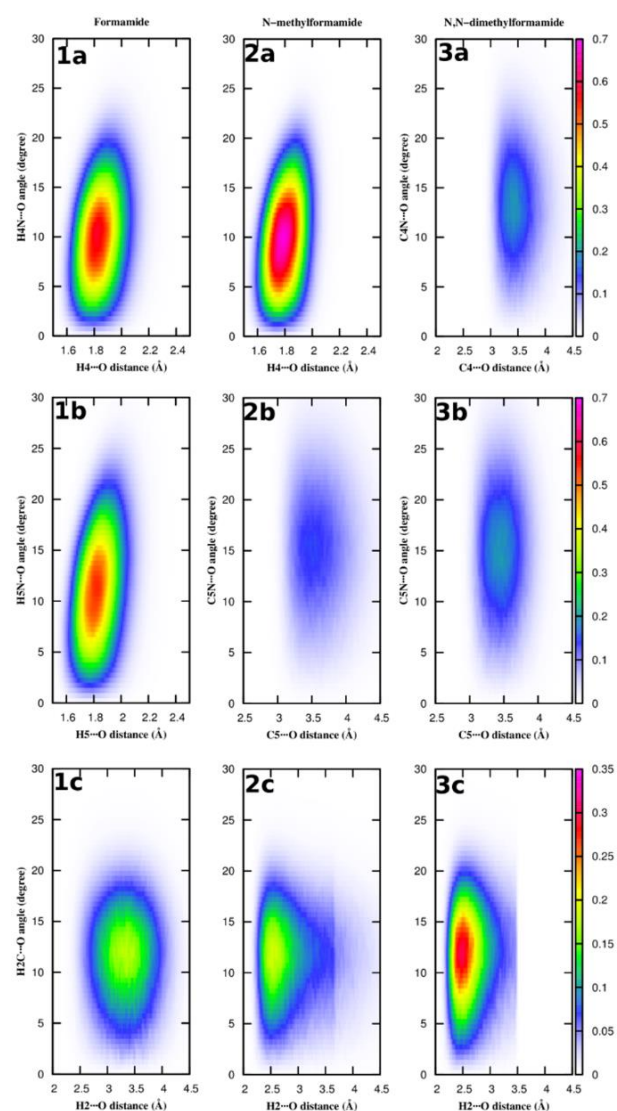
**Table 4:** Dynamical properties: hydrogen bond lifetimes for proper ( $\tau_p$ ) and improper ( $\tau_i$ ) HB, average molecular dipoles ( $\langle\mu\rangle$ ), and relaxation time of the molecular dipole ( $\tau_{\mu}$ ), rotation of the CO vector ( $\tau_{CO}$ ) and rotation of the normal to the molecular plane ( $\tau^{\perp}$ ). With the exception of dipoles (D) all values are in ps.

Molecule/ Properties	$\tau_p$	$\tau_i$	$\langle\mu\rangle$ (D)	$\tau_{\mu}$	$\tau_{CO}$	$\tau^{\perp}$
HCO-NH <sub>2</sub>	0.16	0.02	4.8	16.4	15.7	4.8
HCO-NHCH <sub>3</sub>	0.28	0.04	4.7	58.8	62.0	14.0
HCO-N(CH <sub>3</sub> ) <sub>2</sub>	-	0.05	4.7	10.5	10.0	5.9

As for hydrogen bond lifetimes the results are in line with what could be expected looking at  $g(r, \theta)$  histograms, with NMF having a lifetime for the proper HB roughly two times that of formamide. For the improper HB, we found that DMF had a longer lifetime as compared to NMF, in agreement with the results of Figure 4 and 6. Furthermore for this interaction, we have to consider only the trend in the three systems and not each single value, since the coordinates used to perform the analysis are saved every fs.

A comparable trend is observed with dipole autocorrelation function (see Figure S4 in the Electronic Supporting Information) with HCO-NHCH<sub>3</sub> values decaying about four times more slowly than the other two systems. It is worth to observe how the different structure observed for these highly related compounds is reflected also in

dynamical properties, confirming the soundness of formamide and methyl derivatives as significant test cases.

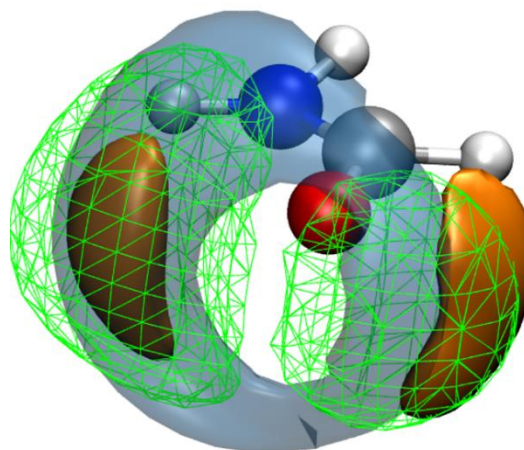


**Figure 6:** The weighted  $g(r, \theta)$  functions for  $\text{HN}\cdots\text{O}$  (or  $\text{CN}\cdots\text{O}$ ) and  $\text{HC}\cdots\text{O}$  interactions for the three systems. From left to right: formamide, NMF and DMF. From top to bottom, H4 (or C4 for  $\text{HCO-N}(\text{CH}_3)_2$ ) and H5 (or C5 for  $\text{HCO-NHCH}_3$  and  $\text{HCO-N}(\text{CH}_3)_2$ ) and H2.

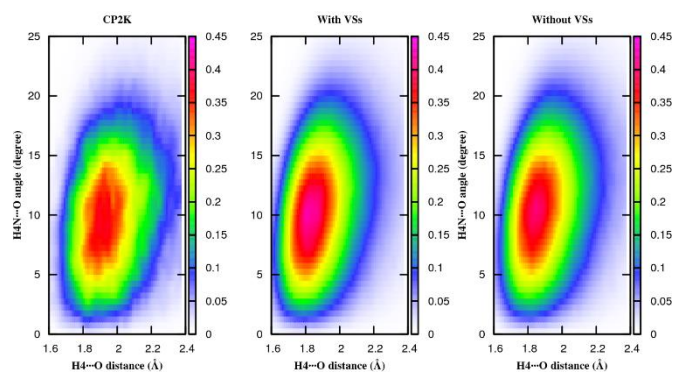
**Comparison with AIMD.** To further assess the reliability of our procedure, we compared the results of the  $\text{HCO-NH}_2$  flexible simulation with those issuing from an AIMD trajectory (see section Computational methods) of liquid formamide and another classical trajectory obtained with CM5 charges without VSs. In Figure 7 the sdf of amine hydrogen atoms with respect to the carbonyl oxygen atom are shown: the model without VSs lacks the directionality in HB that is clearly shown in the *ab initio* and VSs trajectories. This feature can be further appreciated in figure 8 where the weighted  $g(r, \theta)$  function for  $\text{H}_4\text{N}\cdots\text{O}$  is reported.

**Bulk properties.** Having investigated in detail the structure of the three species and demonstrated the agreement between previous computational and experimental studies (and, for formamide, with the AIMD), we tested the ability of the FF to reproduce macroscopic

observables, performing very long MD simulation with a rigid geometry to increase sampling.



**Figure 7:** Isosurface of the amine hydrogen atoms around the molecule of formamide. In orange CP2K results, in green wire frame model with VSs and in blue solid surface model without VSs. The isovalue is of  $25.6 \text{ nm}^{-3}$  for MD simulations and of  $38.5 \text{ nm}^{-3}$  for AIMD simulation.



**Figure 8:** The weighted  $g(r, \theta)$  functions for  $\text{H}_4\text{N}\cdots\text{O}$  interaction in formamide. From left to right: CP2K, model with VSs and model without VSs.

Note that using flexible and rigid models yields the same results (see Figure S5 in the Electronic Supporting Information) in terms of liquid structure but lower time interval between saved configurations in the former allows for a less noisy estimation of relaxation times; on the other hand the estimation of  $\epsilon$  requires very long simulations<sup>16</sup> and thus a rigid model was more convenient.

Considering  $\rho$  and  $\Delta H_{\text{vap}}$  the performance of our FF is in line with the results obtained in previous studies: in some cases our model slightly outperforms previous results (e.g. DMF density) while in other it is slightly worse (e.g.  $\text{HCO-NHCH}_3$  density) but overall the agreement remains within 3% of the experimental value. For the dielectric constant however, the picture is very different: our model reproduces  $\epsilon$  with a relative error that ranging from 0.008 ( $\text{HCO-N}(\text{CH}_3)_2$ ) to 0.11 ( $\text{HCO-NHCH}_3$ ) while OPLS leads to errors about an order of magnitude larger. Indeed, de la Luz et al.<sup>38</sup> managed to obtain slightly better estimates of thermodynamic observables; it must be observed however, such a result was based on after a very careful selection of the charge fitting method (Mulliken charges) and on stepwise rescaling of both atomic charges and Lennard Jones

parameters to obtain sequentially optimal values for  $\epsilon$ , the surface tension  $\gamma$  and  $\rho$ . Because the calculated  $\epsilon$  and  $\rho$  values for HCO-NHCH<sub>3</sub> are slightly worse with respect to the value obtained for the other systems, we performed a further simulation for HCO-NHCH<sub>3</sub> setting the box volume in order to match the experimental density. Under these conditions we have obtained an improved dielectric constant value of  $178 \pm 3$  (relative error 0.02). Goldman and Joslin<sup>73</sup> proposed that it is possible to correlate the presence of hydrogen bonds with the value of  $\epsilon$ , proposing that hydrogen bonded and dipolar aprotic liquids form two different classes of high vs low  $\epsilon$  with respect to an empirical function depending on density, molecular dipole and thermodynamic conditions; in their work they include the molecular species presented here in first (formamide and NMF) or second (DMF) class. Therefore, considering the strength of HB interactions shown in Figure 6 and the lifetimes of Table 5 the highest  $\epsilon$  value should be expected for HCO-NHCH<sub>3</sub> and the lowest for HCO-N(CH<sub>3</sub>)<sub>2</sub> in perfect agreement with the results obtained using our FF.

**Free Energy of Hydration.** One of the benchmark usually adopted to assess the accuracy of a FF is represented by hydration free energy,<sup>29</sup>  $\Delta G_{\text{hyd}}$ , which allows to evaluate the reliability of models to describe Iso interactions in water. A series of MD simulations considering explicit water solvent has been carried out with new boxes formed by one of the studied amide molecules (formamide or N-methylformamide or N,N-dimethylformamide) and 318 TIP3P-FB<sup>19</sup> water molecules. We calculated the new atomic charges and VSs positions using the same procedure described in section Computational methods (the values are reported in Tables S3 and S4 respectively of Electronic Supporting Information). The  $\Delta G_{\text{hyd}}$  were calculated using Bennett acceptance ratio (BAR),<sup>74</sup> a free energy perturbation (FEP) method, performing the simulations with GROMACS v. 4.6.5.

The FEP protocol computes the free energy difference between the solvated molecule and the molecule in the gas phase by constructing a pathway during which the interactions (electrostatic and Van der Waals) of the molecule with the solvent are stepwise decoupled. The coupling parameter  $\lambda$  takes values of 0 (the maximum interaction strength), 0.2, 0.4, 0.6, 0.8, 0.9, and 1 (no interaction). The derivative of the Hamiltonian for each  $\lambda$  value was saved and used to obtain the free energy contribution,  $\Delta G_{\text{hyd}}$ , which is calculated as  $\Delta G_{\text{hyd}} = -(\Delta G_{\text{Coul}} + \Delta G_{\text{vdW}})$ . The hydration free energy for formamide is  $-10.01 \pm 0.43$  kcal/mol, whereas for NMF and DMF we have obtained values of  $-8.95 \pm 0.39$  kcal/mol and  $-6.82 \pm 0.41$  kcal/mol respectively. The agreement with experiment can be considered satisfactory, considering that the measured  $\Delta G_{\text{hyd}}$  for NMF and DMF are  $-10.00$  kcal/mol<sup>29</sup> and  $-7.80$  kcal/mol<sup>29</sup> respectively. We can therefore state that our model, obtained to describe properties and structures of pure liquids, can be confidently extended to the study of aqueous solutions.

## Discussion

In this study a new general protocol for the parametrization of classical force fields has been introduced and applied to formamide and its methyl derivatives with the aim of reproducing with high accuracy the nanoscopic structure of the pure liquids and some key thermodynamic properties. Both intra- and inter-molecular degrees of freedom were taken into account in long time MD simulations under different ensembles. We obtained FF parameters specific for formamide and its derivatives developing a general and robust strategy applicable to other systems.

**Table 5:** Computed values, experimental and literature data for density (kg/m<sup>3</sup>), heat of vaporization (kcal/mol) and static dielectric constant for HCO-NH<sub>2</sub>, HCO-NHCH<sub>3</sub>, HCO-N(CH<sub>3</sub>)<sub>2</sub>.

		HCO-NH <sub>2</sub>	HCO-NHCH <sub>3</sub>	HCO-N(CH <sub>3</sub> ) <sub>2</sub>
$\rho$ (kg/m <sup>3</sup> )	Exp <sup>71,72</sup>	1128.8	999.0	943.3
	Ref <sup>16</sup> (OPLS)	$1121.8 \pm 0.1$	$979.5 \pm 0.1$	$921.9 \pm 0.1$
	Ref <sup>34</sup>	$1100 \pm 3$	$1000 \pm 3$	$930 \pm 3$
	This work	$1138.3 \pm 0.1$	$978.5 \pm 0.2$	$944.9 \pm 0.1$
$\Delta H_{\text{vap}}$ (kcal/mol)	Exp <sup>35</sup>	14.70	13.77	11.10
	Ref <sup>16</sup> (OPLS)	$12.41 \pm 0.02$	$11.91 \pm 0.02$	$11.01 \pm 0.02$
	Ref <sup>34</sup>	$14.79 \pm 0.02$	$13.77 \pm 0.02$	$11.40 \pm 0.03$
	This work	$14.1 \pm 0.1$	$13.3 \pm 0.1$	$10.5 \pm 0.1$
$\epsilon$	Exp <sup>35</sup>	108.9	181.6	37.2
	Ref <sup>16</sup> (OPLS)	51	$18.8 \pm 0.6$	$14.9 \pm 0.5$
	Ref <sup>38</sup>	107.6	-	-
	This work	$109.8 \pm 0.9$	$162 \pm 2$	$35.7 \pm 0.3$

Starting from the proposal put forward in a recent study<sup>29</sup> we tried to find a rationale behind the magnitude of scaling factors for atomic charges and to improve the performances of the model for some troublesome systems. We based the needed correction for overpolarization, with respect to gas-phase charges, on the use of C-PCM and of virtual sites, whose positions and charges were based on “first principle” considerations (the Boys procedure and the QM molecular dipole moment). In this section we briefly review the key results of the present work, underlining its strong and weak points in comparison to previous studies.

#### *Use of virtual sites yields a correct structure*

The first consideration concerns the generation of a liquid structure in agreement with the available experimental data from neutron and X-ray diffraction studies,<sup>39,40</sup> as shown by inspection of radial distribution functions in section Results. To the best of our knowledge, the only available computational study of the complete series formed by formamide, NMF and DMF was performed by Cordeiro<sup>34</sup> with results very similar to those presented here; however, in that case molecular charges had to be rescaled independently for each compound yielding molecular dipole moments significantly higher (between 15% and 20%) than the corresponding gas phase values (e.g., the charge on the N atom changes from -0.83 to -0.57 when going from HCO-NH<sub>2</sub> to HCO-N(CH<sub>3</sub>)<sub>2</sub>) to obtain the right polarization and some other parameters (e.g. the nitrogen and oxygen atom well-depth to the Lennard Jones potential) needed also to be adjusted, at variance with our generalized procedure. As mentioned in section Results, Jorgensen and Swenson<sup>70</sup> also obtained a good match with known data for interatomic distance, although their study did not include N-methylformamide. In a recent work, de la Luz and co-workers<sup>38</sup> present a thorough study of fitting charge methods for the derivation of a FF able reproduce “hard to crack”<sup>16</sup> observables such as dielectric constant; however, they obtain an excessively tight liquid structure for HCO-NH<sub>2</sub>, as evidenced, for instance, by the very short first shell distances of  $g_{ON}(r)$ . Finally, our FF was in agreement with experimental observations, and our procedure using VSs was also able to generate a distribution of hydrogen bonds having the same, non uniform, pattern obtained in AIMD simulation.

#### *Thermodynamic properties are consistently reproduced in the whole series*

In the computational studies cited in the previous sections,<sup>16,34,38,29</sup> several bulk properties were presented as benchmarks attesting the ability (or lack thereof) of the studied classical FF to reproduce thermodynamic observables. Generally speaking, good results were obtained for most properties with the notable exception of the dielectric constant (with the exclusion of the work of de la Luz et al.<sup>38</sup>). However, as mentioned above, this remarkable outcome came at the price of an excessively tight liquid structure. In the present work, we were able to obtain a reasonable liquid structure and, simultaneously, a good agreement with experimental estimations of  $\epsilon$  for the studied molecules, including (albeit with slightly reduced accuracy with respect to the other two species) N-methylformamide. Moreover, we were able to correlate the results obtained for the dielectric constant with the characteristics of hydrogen bond

dynamics, providing further evidence for the consistence of our protocol.

## Conclusions

In this study we have presented a protocol for the derivation of highly accurate and transferable classical Force Fields, to be used in Molecular Dynamics and Monte Carlo simulations. The protocol was focused on deriving effective point charges for molecular species featuring high dipole moment and being able to form, in condensed phase, strong hydrogen bonds. The characteristic features of our approach are the use of virtual sites for describing lone pairs and of C-PCM for tuning charges in different solvents. We showed that the proposed procedure was able to predict correct nanoscopic structures for all the studied systems as compared to previous computational and experimental studies; comparable success was also obtained in the estimation of bulk properties.

## Acknowledgements

This work was supported by Italian MIUR through FIRB program contracts n° RBFR12ETL5 and RBFR13PSB6. The authors thank the SMART@SNS Laboratory technical staff for managing the computing facilities at SNS. MM thanks the support of Scuola Normale Superiore grant “Giovani Ricercatori 2015”.

## References

- 1 Allen M. P.; Tildesley D. J. *Computer Simulation of Liquids*, Clarendon Press/Oxford University Press.; Oxford, England/New York, 1989.
- 2 Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: Harlow, England ; New York, 2001.
- 3 Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- 4 Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. OPLS All-Atom Force Field for Carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955–1970.
- 5 Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- 6 Yakovenko, O. Y.; Li, Y. Y.; Oliferenko, A. A.; Vashchenko, G. M.; Bdzhola, V. G.; Jones, S. J. M. Ab Initio Parameterization of YFF1, a Universal Force Field for Drug-Design Applications. *J. Mol. Model.* **2012**, *18*, 663–673.
- 7 Zazza, C.; Mancini, G.; Brancato, G.; Barone, V. *In Silico* Study of Molecular-Engineered Nanodevices: A Lockable Light-Driven Motor in Dichloromethane Solution. *J. Phys. Chem. Lett.* **2013**, *4*, 3885–3890.
- 8 Chillemi, G.; Coletta, A.; Mancini, G.; Sanna, N.; Desideri, A. An

- Amber Compatible Molecular Mechanics Force Field for the Anticancer Drug Topotecan. *Theor. Chem. Acc.* **2010**, *127*, 293–302.
- 9 Chillemi, G.; Mancini, G.; Sanna, N.; Barone, V.; Longa, S. D.; Benfatto, M.; Pavel, N. V.; D'Angelo, P. Evidence for Sevenfold Coordination in the First Solvation Shell of Hg(II) Aqua Ion. *J. Am. Chem. Soc.* **2007**, *129*, 5430–5436.
  - 10 Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
  - 11 Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
  - 12 Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **1984**, *23*, 1513–1518.
  - 13 Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS Software for Biomolecular Simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
  - 14 Heinz, H.; Lin, T.-J.; Kishore Mishra, R.; Emami, F. S. Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir* **2013**, *29*, 1754–1765.
  - 15 Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G. Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *J. Phys. Chem. B* **2004**, *108*, 1255–1266.
  - 16 Coleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.
  - 17 Zubillaga, R. A.; Labastida, A.; Cruz, B.; Martínez, J. C.; Sánchez, E.; Alejandre, J. Surface Tension of Organic Liquids Using the OPLS/AA Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 1611–1615.
  - 18 Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
  - 19 Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.
  - 20 Berendsen H. J. C.; van Gunsteren W. F. *Molecular Dynamics Simulations: Techniques and Approaches*; Series C135; Barnes A. J. et al., 1984; Vol. Molecular Liquids - Dynamics and Interactions.
  - 21 Duvail, M.; Vitorge, P.; Spezia, R. Building a Polarizable Pair Interaction Potential for lanthanoids(III) in Liquid Water: A Molecular Dynamics Study of Structure and Dynamics of the Whole Series. *J. Chem. Phys.* **2009**, *130*, 104501.
  - 22 Lamoureux, G.; Roux, B. Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 3025.
  - 23 Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *J. Chem. Phys.* **1994**, *101*, 6141.
  - 24 Lipparini, F.; Barone, V. Polarizable Force Fields and Polarizable Continuum Model: A Fluctuating Charges/PCM Approach. 1. Theory and Implementation. *J. Chem. Theory Comput.* **2011**, *7*, 3711–3724.
  - 25 Mancini, G.; Brancato, G.; Barone, V. Combining the Fluctuating Charge Method, Non-Periodic Boundary Conditions and Meta-Dynamics: Aqua Ions as Case Studies. *J. Chem. Theory Comput.* **2014**, *10*, 1150–1163.
  - 26 Brancato, G.; Rega, N.; Barone, V. A Hybrid Explicit/implicit Solvation Method for First-Principle Molecular Dynamics Simulations. *J. Chem. Phys.* **2008**, *128*, 144501.
  - 27 Carnimeo, I.; Cappelli, C.; Barone, V. Analytical Gradients for MP2, Double Hybrid Functionals, and TD-DFT with Polarizable Embedding Described by Fluctuating Charges. *J. Comput. Chem.* **2015**, *36*, 2271–2290.
  - 28 Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
  - 29 Dodda, L. S.; Vilseck, J. Z.; Cutrona, K. J.; Jorgensen, W. L. Evaluation of CM5 Charges for Nonaqueous Condensed-Phase Modeling. *J. Chem. Theory Comput.* **2015**, *11*, 4273–4282.
  - 30 Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.
  - 31 Vidossich, P.; Lledós, A. The Use of Localised Orbitals for the Bonding and Mechanistic Analysis of Organometallic Compounds. *Dalton Trans.* **2014**, *43*, 11145–11151.
  - 32 Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation of CM5 Charges for Condensed-Phase Modeling. *J. Chem. Theory Comput.* **2014**, *10*, 2802–2812.
  - 33 Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
  - 34 Cordeiro, J. M. M. C-H...O and N-H...O Hydrogen Bonds in Liquid Amides Investigated by Monte Carlo Simulation. *Int. J. Quantum Chem.* **1997**, *65*, 709–717.
  - 35 Lide, E. *Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*; Handbook of Chemistry and Physics; CRC Press: Boca Raton, Florida, 1993.
  - 36 Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24*, 669–681.
  - 37 Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
  - 38 de la Luz, A. P.; Méndez-Maldonado, G. A.; Núñez-Rojas, E.;

- Bresme, F.; Alejandre, J. A New Force Field of Formamide and the Effect of the Dielectric Constant on Miscibility. *J. Chem. Theory Comput.* **2015**, *11*, 2792–2800.
- 39 Kálmán E.; Serke I.; Pálincás G.; Zeidler M. D.; Wiesmann F. J.; Bertagnolli H.; Chieux P. The Molecular Structure and Hydrogen Bond Geometry in Liquid Formamide: Electron, Neutron, and X-Ray Diffraction Studies. *Z. Für Naturforsch* **1983**, *38a*, 231–236.
- 40 Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. Intermolecular N–H.O Hydrogen Bonds Assisted by Resonance. Heteroconjugated Systems as Hydrogen-Bond-Strengthening Functional Groups. *Acta Crystallogr. B* **1995**, *51*, 1004–1015.
- 41 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,; M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci,; G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian,; A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada,; M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima,; Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr.,; J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers,; K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand,; K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi,; M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross,; V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann,; O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski,; R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth,; P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels,; O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski,; and D. J. Fox. *Gaussian 09 Revision E.01*.
- 42 Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.
- 43 Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.
- 44 Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: Integrated and User-Friendly Tools for the Parameterization of Intramolecular Force Fields from Quantum Mechanical Data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736.
- 45 Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- 46 Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* **1881**, *248*, 127–136.
- 47 Berthelot D. Sur Le Mélange Des Gaz. *Comptes Rendus Hebd. Séances Académie Sci.* **1898**, *126*, 1703–1855.
- 48 Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- 49 Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- 50 Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.
- 51 Chandramouli, B.; Zazza, C.; Mancini, G.; Brancato, G. Boundary Condition Effects on the Dynamic and Electric Properties of Hydration Layers. *J. Phys. Chem. A* **2015**, *119*, 5465–5475.
- 52 Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVande, J. CP2K: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15–25.
- 53 Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- 54 Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- 55 VandeVondele, J.; Hutter, J. Gaussian Basis Sets for Accurate Calculations on Molecular Systems in Gas and Condensed Phases. *J. Chem. Phys.* **2007**, *127*, 114105.
- 56 Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- 57 Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic Separable Dual-Space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.
- 58 Brehm, M.; Kirchner, B. TRAVIS - A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *J. Chem. Inf. Model.* **2011**, *51*, 2007–2023.
- 59 Pagliai, M.; Muniz-Miranda, F.; Cardini, G.; Righini, R.; Schettino, V. Hydrogen Bond Dynamics of Methyl Acetate in Methanol. *J. Phys. Chem. Lett.* **2010**, *1*, 2951–2955.
- 60 Pagliai, M.; Cardini, G.; Righini, R.; Schettino, V. Hydrogen Bond Dynamics in Liquid Methanol. *J. Chem. Phys.* **2003**, *119*, 6655.
- 61 Candelaresi, M.; Pagliai, M.; Lima, M.; Righini, R. Chemical Equilibrium Probed by Two-Dimensional IR Spectroscopy: Hydrogen Bond Dynamics of Methyl Acetate in Water. *J. Phys. Chem. A* **2009**, *113*, 12783–12790.
- 62 van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. A Systematic Study of Water Models for Molecular Simulation: Derivation of Water Models Optimized for Use with a Reaction Field. *J. Chem. Phys.* **1998**, *108*, 10220.
- 63 Torda, A. E.; van Gunsteren, W. F. Algorithms for Clustering Molecular Dynamics Configurations. *Journal of computational chemistry* **1994**, *15* (12), 1331–1340.
- 64 van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.
- 65 Janert, P. K. *Gnuplot in Action: Understanding Data with Graphs*, 2nd revised edition.; Manning: Greenwich, Conn, 2015.
- 66 Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- 67 Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

- 68 Fantoni, A. C.; Caminati, W. Rotational Spectrum and Ab Initio Calculations of N-Methylformamide. *J. Chem. Soc. Faraday Trans.* **1996**, *92*, 343.
- 69 Ohtaki, H.; Itoh, S.; Yamaguchi, T.; Ishiguro, S.; Rode, B. M. Structure of Liquid N,N-Dimethylformamide Studied by Means of X-Ray Diffraction. *Bull. Chem. Soc. Jpn.* **1983**, *56*, 3406–3409.
- 70 Jorgensen, W. L.; Swenson, C. J. Optimized Intermolecular Potential Functions for Amides and Peptides. Structure and Properties of Liquid Amides. *J. Am. Chem. Soc.* **1985**, *107*, 569–578.
- 71 Marcus, Y. *The Properties of Solvents*; Wiley series in solution chemistry; Wiley: Chichester ; New York, 1998.
- 72 Yaws, C. L. *Yaws Handbook of Thermodynamic Properties for Hydrocarbons and Chemicals*; Knovel: New York, 2009.
- 73 Goldman, S.; Joslin, C. Why Hydrogen-Bonded Liquids Tend to Have High Static Dielectric Constants. *J. Phys. Chem.* **1993**, *97*, 12349–12355.
- 74 Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

## TOC

Accurate tuning of polarization effects and effective virtual sites generation allows to obtain bulk properties in agreement with experimental values.

