

Journal Pre-proof

Self-consistent electrostatic embedding for liquid phase polarization

Miguel Jorge, José R.B. Gomes, Andrew W. Milne



PII: S0167-7322(20)35613-0

DOI: <https://doi.org/10.1016/j.molliq.2020.114550>

Reference: MOLLIQ 114550

To appear in: *Journal of Molecular Liquids*

Received date: 26 August 2020

Revised date: 7 October 2020

Accepted date: 9 October 2020

Please cite this article as: M. Jorge, J.R.B. Gomes and A.W. Milne, Self-consistent electrostatic embedding for liquid phase polarization, *Journal of Molecular Liquids* (2018), <https://doi.org/10.1016/j.molliq.2020.114550>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2018 Published by Elsevier.

Self-Consistent Electrostatic Embedding for Liquid Phase Polarization

Miguel Jorge^{1,}, José R. B. Gomes², Andrew W. Milne¹*

¹ Department of Chemical and Process Engineering, University of Strathclyde, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom

² CICECO – Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, Aveiro, Portugal

* Email: miguel.jorge@strath.ac.uk

Abstract

While it is well known that molecules can be strongly polarized when transferred from the gas phase to a polar liquid, quantifying polarization effects explicitly using either experiment or theory has remained elusive. In this paper, we present a new QM/MM method involving a self-consistent calculation of the liquid state dipole moments, that is able to yield realistic, accurate estimates of the multipole moments of molecules in the liquid state. As a proof-of-concept, we apply our Self-Consistent Electrostatic Embedding (SCEE) method to the widely studied system of pure water. The method gives molecular dipole moments that are significantly enhanced with respect to the isolated gas-phase molecule and that are consistent with the best current experimental estimate of this property. While previous QM/MM calculations on the same system systematically underestimate the liquid dipole moment, those predictions become consistent with our own when several shortcomings are accounted for in an approximate way. Furthermore, sampling liquid configurations using several (but not all) fixed-charge force fields yields results that are consistent with sampling from a classical polarizable model. We then extract several contributions to the polarization energy (i.e. the change in energy when transferring a molecule from the gas to the liquid phase) and show that the distortion correction is cancelled out by the purely electronic contribution to the polarization energy. This insight is very important from the point of view of force-field development, since it allows us to unequivocally quantify the two missing energy terms in classical non-polarizable models. This provides a way to systematically improve predictions of phase-change energies (e.g. enthalpy of vaporization, hydration free energies) from such force-fields by correcting for the missing polarization effects.

Keywords: Water; Dipole moment; Quantum Mechanics/Molecular Mechanics; Ab initio; Force field; Polarization

1. Introduction

When a polar molecule like water is transferred from the gas to the liquid phase, it is strongly polarized by the electric field of the surrounding molecules. This causes a distortion of both the molecular geometry, in the sense that the atomic positions are displaced relative to one another, and the electronic cloud. This intramolecular distortion is unfavorable, resulting in a positive contribution to the free energy of the system. However, it is more than compensated by an enhancement of the intermolecular interactions with the surrounding molecules, caused by the polarization – i.e., the distorted polarized molecules interact more strongly with each other than if they had retained their gas-phase structure and electron distribution. As a result, the ensemble of distorted molecules has a lower overall free energy and the molecular dipole moment (and, most likely, other higher order multipole moments) is enhanced in the liquid phase when compared to the isolated molecule. This picture has been widely accepted and generally understood at least since the pioneering theoretical work by Born [1], Kirkwood [2], Onsager [3] and others (see [4] and references therein). However, actually quantifying the degree of dipole moment enhancement is fraught with difficulties, mainly because the electronic properties of individual molecules cannot be easily decoupled from those of the surrounding molecules in the liquid. In this paper, we propose a new theoretical method, based on the Quantum Mechanics/Molecular Mechanics (QM/MM) approach, that is able to calculate the dipole moment and polarization energies of individual molecules in the liquid phase.

The relatively high dipole moment of water is strongly connected to its unusual thermodynamic properties, including some of the so-called “water anomalies” [5-7]. For example, it contributes significantly to the large cohesive energy of liquid water, leading to its anomalously high boiling point. The dipole moment is also intimately linked to water’s excellent ability to solvate polar compounds and to dissociate electrolytes. Moreover, and of particular interest from the point of view of force-field development, knowledge of the liquid-phase dipole moment can be used to derive polarization corrections for phase-change properties, e.g. the enthalpy of vaporization and the solvation free energy [8, 9], and for electronic properties such as the dielectric constant [10, 11]. It can also inform the assignment of point charges to non-polarizable models so as to approximately capture the potential energy surface (PES) of a polarized liquid phase [12]. For example, we have recently shown that a non-polarizable force field that includes polarization corrections at the parameterization stage can predict solvation free energies of alcohols in solvents of different polarity [13].

Generalizing these recent developments requires a method that can accurately predict the dipole moment and polarization energy contributions of individual molecules in the liquid phase in a computationally expedient way. This is the main aim of the present work.

An illustration of the conceptual difficulties in determining the dipole moment of water in the liquid phase is the fact that for several decades the value referred to as the “experimental liquid dipole” was actually based on a theoretical calculation (not an experiment) on ice Ih (not on liquid water) [14]. To the best of our knowledge, the only experimental estimate of this quantity to date was reported by Badyal et al. [15]. By analyzing neutron and X-ray diffraction experiments in liquid water, they estimated the degree of interatomic charge transfer, and hence the molecular dipole moment, obtaining a value of 2.9 D with a large uncertainty of ± 0.6 D. This represents an enhancement of about 1 D over the isolated gas-phase dipole of 1.855 D [16]. Shortly after, Gubskaya and Kusalik [17] obtained a very similar value of 2.95 D based on a mean field approximation that required both experimental (refractive index), *ab initio* (electronic response properties) and classical simulation data (local fields and field gradients) as inputs.

From a theoretical point of view accurate determination of the liquid dipole moment ideally requires the following ingredients: 1) the molecule of interest (we shall call it the “central” molecule for simplicity) should be polarized by the electric field of the surrounding molecules; 2) the surrounding molecules should themselves be polarized by the presence of their neighbors, including the central molecule; 3) the structure of the liquid, in terms of molecular arrangements, must be correctly described with respect to the experimental structure; 4) an adequate statistical sample of liquid-phase fluctuations at the relevant temperature should be carried out; 5) the electronic properties of the central molecule must be decoupled from those of adjacent molecules. In principle, one possibility would be to use a full QM description of several large clusters of water molecules representative of the liquid phase and then determine the average molecular dipole moment in the ground state. Gregory et al. [18] attempted to achieve this by a combination of QM calculations and experiments on water clusters up to the hexamer. They identified several (typically three) low-energy arrangements for each cluster size, calculated their ground state energies and obtained the dipole moments as a Boltzmann-weighted average over those different arrangements. They observed that the dipole moment of individual water molecules in those clusters increases asymptotically to a value of 2.72 D as the cluster size increases. This approach satisfies Criteria 1 and 2, but quickly runs into problems in Criteria 3 and 4 for clusters beyond the

hexamer, since it becomes impossible to probe the immense number of possible structural arrangements. Furthermore, as the authors themselves remarked, the local structure of liquid water is based on a three-dimensional hydrogen-bonded network that is quite distinct from the predominantly cyclic structures of small clusters [18]. Finally, the calculation of individual molecular properties from the wave function of a super-molecule (Criterion 5) is far from trivial.

An alternative approach has been to use *ab initio* molecular dynamics (AIMD) to simulate a periodic box of liquid water. The AIMD approach [19] invokes the Born-Oppenheimer approximation and treats the nuclear degrees of freedom classically while treating the electronic degrees of freedom through an electronic structure approach, most often based on density functional theory (DFT). This allows one to effectively perform a statistical-mechanical sampling of the system at the relevant temperature, thus addressing Criterion 4 as well as 1 and 2. In principle, Criterion 5 should also be fulfilled, but in practice many DFT exchange-correlation functionals give a rather poor description of the structural and thermodynamic properties of liquid water (see [20] for a detailed review). In fact, while early studies [21-26] made use of functionals based on the generalized-gradient approximation (GGA) like BLYP [27, 28] or PBE [29], due to computational and methodological limitations at the time later studies [29-34] started to employ hybrid functionals (e.g. PBE0 [35]) and approaches to account for dispersion interactions (e.g., Grimme corrections [36]), which have led to a marked improvement in describing the correct structure of liquid water [20]. Furthermore, fulfilling Criterion 5 is also not easy in AIMD, as the resulting dipole moment is known to depend strongly on the way the electron density is partitioned between different molecules [37]. The first AIMD study of liquid water [21] used a simple spherical cut-off scheme that yielded a dipole moment of 2.66 D. A later study [22] used a Bader-type approach [38] to partition the electron density, arriving at a dipole moment of 2.43 D. In the same year, Silvestrelli and Parrinello [23, 24] applied the method of maximally localized Wannier functions (MLWF) [39] to isolate the contribution of each molecule, and obtained a much larger dipole moment of 3.00 D. The fact that these three simulations were carried out at nearly the same conditions and system sizes emphasizes the ambiguity in determining molecular dipole moments from super-molecular wave functions [40]. Nevertheless, the MLWF approach was shown to provide a robust way of carrying out this partitioning [41] and has been practically universally adopted in subsequent AIMD calculations of liquid dipole moments. Interestingly, most calculations using MLWF, hybrid

functionals and dispersion corrections yield dipole moments that are quite close to the experimental estimate of 2.9 D [30-34].

At the opposite extreme from a full QM description of the system are simulations using classical molecular models. The most popular of these are fixed-charge non-polarizable models, as they combine low computational cost with a fairly accurate description of the structural and thermodynamic properties of liquid water [42, 43]. Achieving this, however, relies on a careful calibration of the intermolecular interaction parameters, in particular the point charges. These are selected so that the effects of polarization in the liquid are captured in a mean-field sense, leading to dipoles that are higher than that of the isolated molecule but smaller than experimental and QM estimates of the liquid phase dipole – the dipole moments of most fixed-charge water models in current use lie between 2.18 and 2.48 D [9, 42, 44]. In fact, it has been suggested that point charges that are roughly halfway between the gas and the liquid are best able to capture the average permanent and induced electrostatic interactions in the liquid state [8, 12, 45, 46]. As such, fixed-charge models cannot provide an accurate prediction of dipole moments in the liquid phase. In classical polarizable models, on the other hand, the charge distribution of each molecule can respond on-the-fly to changes in the surrounding electric field during the simulation. Naturally, this comes at a price in terms of computational cost, with polarizable models typically increasing simulation times by a factor of 3-10 [47]. However, they do provide the liquid dipole moment as an observable property that can be averaged during a simulation. Liquid dipole moments estimated from classical polarizable models vary widely, being mostly within the range of 2.5 to 3.1 D [42, 48-50], which is broadly consistent with the experimental and AIMD estimates. The problem with this approach is that the resulting dipole is strongly dependent on the parameterization strategy and target properties.

An approach that has been quite popular over the years is QM/MM, which relies, as the name implies, on a combination of the quantum mechanical and classical approaches. In this method, the part of the system that is of most interest is treated at the QM level, while the rest of the system is treated with molecular mechanics (MM). A key choice in this approach is how to implement the coupling between the two parts of the system, i.e., how to describe the interactions between the QM and MM subsystems. Following Senn and Thiel [51], we can subdivide QM/MM into three categories, depending on how the electrostatic interactions between the QM and MM parts are handled. In mechanical embedding, all interactions are treated classically, so that the QM wave function is not affected directly by the presence of

the MM molecules. This category is of no interest in the context of this paper, since the resulting dipole moment would just be the same as for an isolated molecule. In electrostatic embedding (EE), the QM wave function is polarized by the presence of the point charges of the MM molecules but the latter are kept fixed, typically at the values of the underlying classical model. In polarized embedding (PE), the MM model is polarizable, so that it is itself able to respond to the presence of the QM part and of other MM molecules. Typically, this requires a self-consistent iteration until the QM and MM charge distributions converge. The main advantage of QM/MM, compared to the full QM or AIMD approaches, is that satisfying Criterion 5 is straightforward – if a single molecule is treated at the QM level and the rest classically, its wave function is isolated from the surrounding. The EE method satisfies Criterion 1, but not 2, while PE in principle satisfies both criteria. Criterion 3, however, depends on the quality of the underlying MM model, while Criterion 4 depends strongly on the details of the implementation of the method, as we will discuss below.

There have been several QM/MM studies that calculated dipole moments of liquid water, using both EE [52-63] and PE [64-79] approaches, and these will be analyzed in detail later in the paper. In general, EE leads to lower dipole moments than PE, as would be expected, and both tend to be somewhat lower than the AIMD estimates described above. We will show that this discrepancy is most likely due to shortcomings of the previous QM/MM methods when calculating the dipole moment in the liquid phase, mainly related to difficulties in fulfilling Criteria 3 and 5 (as well as 2 for the EE approach). We further propose a new approach, which we term Self-Consistent Electrostatic Embedding (SCEE), that is able to eliminate these shortcomings and provide an estimate of the dipole moment of liquid water that is coherent with experiment and other theoretical approaches. Furthermore, the SCEE method is easy to implement, computationally expedient and generally applicable to molecular liquids. In this paper, we will focus on liquid water to present extensive methodological tests of our approach, comparing it to previous QM/MM methods. In subsequent publications, we plan to apply SCEE to determine the dipole moment of several pure liquids, as well as extend it to solutions involving small organic molecules. The paper is organized as follows. In Section 2, we first describe the key idea behind the SCEE method and then report the computational details for the QM and MM calculations. In Section 3, we begin by presenting a detailed analysis of the parameters of the method, followed by a comparison of different classical models for the MM part. We compare our results to

previous studies and finish with an analysis of the different contributions to the polarization energy.

2. Methodology

2.1 Self-Consistent Electrostatic Embedding Approach

We carried out QM/MM calculations on a large number of water clusters extracted from classical molecular dynamics (MD) simulations. The theoretical basis of the QM/MM method has been described in great detail elsewhere (see, e.g. [51] and references therein), so here we will focus only on aspects that are specific to our SCEE approach. In a typical QM/MM calculation, the Hamiltonian of the system (H_{Tot}) is split into three components:

$$H_{\text{Tot}} = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM/MM}} \quad (1)$$

where H_{QM} corresponds to the Hamiltonian of the QM system (in our case, the central water molecule), H_{MM} corresponds to the purely classical part of the system (in our case, the interaction energy between the MM point charges), and $H_{\text{QM/MM}}$ is the energy arising from the coupling between the QM and MM parts of the system. It is important to note that in this work only electrostatic interactions (and polarization) were included in the $H_{\text{QM/MM}}$ term, since we are primarily interested in the calculation of electronic properties and electrostatic energy contributions. To obtain realistic values for the total interaction energy, repulsion and dispersion interactions would have to be included as well, which is normally achieved through a Lennard-Jones (LJ) potential. We further note that there is currently no universally accepted method for determining the optimal LJ parameters for use in QM/MM calculations [51].

In typical EE calculations, the QM wave function is perturbed by the presence of the MM point charges, i.e. it is polarized by the surrounding field. However, the charges of each MM molecule are kept fixed at the same value, usually corresponding to the parameters of the classical MM force field used (e.g., TIP3P). In contrast, in PE calculations each of the surrounding molecules will have a different dipole moment determined by the underlying polarizable force field. Furthermore, the MM dipoles are adjusted self-consistently during the calculation, so that they can themselves be polarized by the QM molecule. Naturally, such calculations are much more computationally expensive than in the EE method. In our SCEE approach, we try to strike a balance between the increased realism of PE and the

computational speed of EE. In essence, we carry out a small number of EE calculations for each configuration, using different values of the surrounding fixed point charges, and calculate the dipole moment of the central QM molecule as a function of the dipole moment of the surrounding MM molecules. We then find the point at which the QM and MM dipoles are equal, which corresponds to the self-consistent dipole moment of liquid water. This can be achieved through an iteration loop, as depicted in the diagram of Figure 1a. However, we will show in section 3.1 that the converged dipole moment can also be obtained using a simple fitting scheme from only three EE calculations (Figure 1b), thus avoiding the time-consuming iteration process.

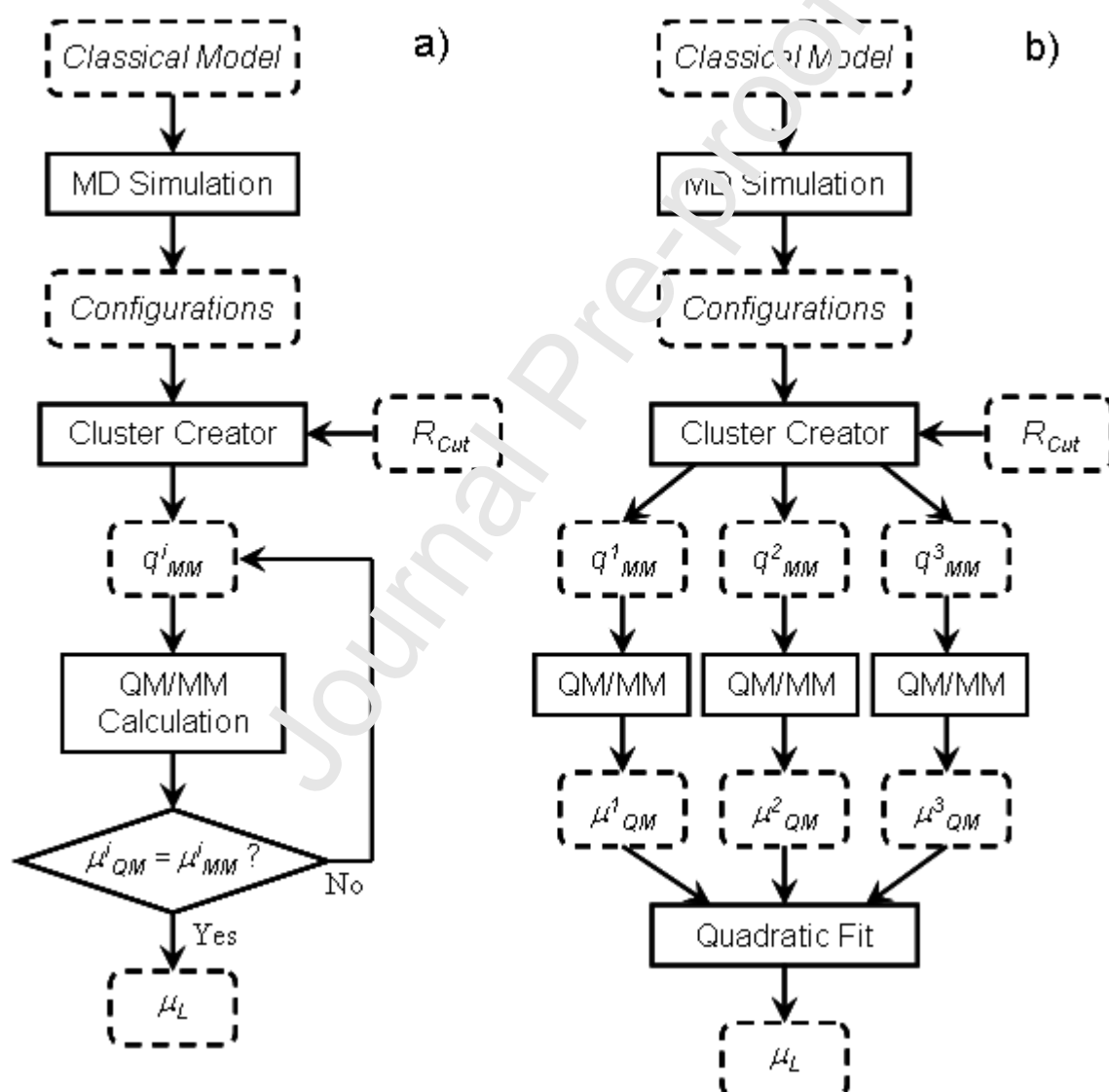


Figure 1 – Schematic flowcharts depicting the SCEE procedure: (a) using the initial iterative approach; (b) using the faster quadratic fit to three individual QM/MM results. The dashed boxes denote inputs to each particular procedure (rectangles).

The first step in the SCEE procedure is to harvest several configurations for pure liquid water from MD simulations. We sampled 200 configurations from each MD simulation using different classical non-polarizable fixed-charge models, namely the 3-site models TIP3P [80] and SPC/E [81], as well as the more recent 4-site models TIP4P2005 [82] and TIP4P-FB [83]. The potential parameters for each of these models are given in the original publications and have been compiled in our recent publication [9], where we also provide input files for all the MD simulations analyzed here. One important limitation of empirical fixed-charge models is that the dipole moment of water in the liquid state is assigned a dipole moment that is the same for all molecules, whereas in reality it will vary depending on the local environment experienced by each molecule. To test the effect of this approximation in the MM part of the system, we have also used configurations from MD simulations using the polarizable SWM4-DP model [48]. We note that these classical models have different levels of performance when describing structural and thermodynamic properties of liquid water, and we refer the readers to previous benchmark comparisons available in the literature [42, 43]. As we discuss below, these different levels of performance can have an effect on the outcome of the SCEE calculations.

For each configuration, one molecule was randomly selected to act as the central molecule, which was treated at the QM level in subsequent DFT calculations. Subsequently, a spherical cluster was cut out of the periodic box so that all water molecules outside a pre-specified cut-off radius (R_{cut}) were discarded. The point charges of the surrounding molecules were assigned values to match a predetermined value for the water dipole moment – e.g., in the procedure depicted in Figure 1a, this was initially the dipole moment of the corresponding classical model, but subsequently assumed other values as the iteration progressed. This cluster generation procedure was carried out using in-house Fortran codes that are provided in Supporting Information. We considered several values of R_{cut} to test the convergence of the results with the number of surrounding water molecules. Unless otherwise noted, the cut-off scheme used the central oxygen atom as a basis – i.e., we retained all surrounding water molecules that had their oxygen atoms within R_{cut} of the oxygen atom of the central molecule.

At the end of each QM/MM calculation, the dipole moment was extracted from the output file and the process was either iterated until convergence of the QM and MM dipoles (Figure 1a), or the QM dipole moments were used in a fitting procedure (Figure 1b) that will be explained in detail later. Once each configuration was optimized, we performed several single-point energy calculations using the fully polarized wave function in order to determine

the relevant energy contributions. For each of those energies, we applied the same fitting approach as in the SCEE calculations, as described in the diagram of Figure S1 (see Supplementary Material). Specifically, we obtained the energy of the full QM/MM system after optimization (E_{Tot}) and then determined the electrostatic interaction energy per QM molecule (E_{Int}) from

$$E_{Int} = \frac{1}{2}(E_{Tot} - E_{QM} - E_{MM}) \quad (2)$$

where E_{QM} was obtained from the first SCF step of a single-point calculation of the polarized QM wave function in vacuum, while E_{MM} is the “self-energy” of the point charges reported in the Gaussian 09 output file. The distortion contribution to the polarization energy (E_{Dist}) [82, 105, 106], also sometimes called the “polarization self-energy”, was obtained from

$$E_{Dist} = E_{QM} - E_{Vac} \quad (3)$$

where E_{Vac} corresponds to the energy of a fully optimized single water molecule in vacuum. We also calculated the total polarization energy (E_{Pol}), as defined by Orozco and co-workers [107], from

$$E_{Pol} = E_{Tot} - E_{Gas} \quad (4)$$

where E_{Gas} corresponds to the total energy of a system composed of the un-polarized QM wave function of the central molecule and the surrounding polarized MM molecules. To obtain this energy, we started from the SCEE-optimized wave function of the central molecule in each configuration, optimized it in vacuum, then replaced all the MM charges corresponding to that original configuration and extracted E_{Gas} from the first SCF step of a new QM/MM calculation. Once E_{Pol} is known, we can calculate the so-called stabilization energy (E_{Stab}) [107, 108], which represents the favorable interaction of the central molecule with the surrounding environment that counteracts the distortion component, from

$$E_{Stab} = E_{Pol} - E_{Dist} \quad (5)$$

Finally, the purely electronic contribution to the polarization energy (E_{Elec}) [8] is given by

$$E_{Elec} = E_{\epsilon_{\infty}} - E_{QM} \quad (4)$$

where $E_{\epsilon_{\infty}}$ was obtained from the first step of a single-point SCRF calculation starting from the polarized QM wave function surrounded by a continuum with dielectric constant equal to

the high-frequency dielectric constant of water. For these single-point energy calculations, we used the SCIPCM model [109], as implemented in Gaussian 09, with a value of $\epsilon_\infty = 1.766$, determined from the square of the experimental index of refraction at the Sodium D-line frequency [11]. This energy term describes in an approximate way the interaction between the polarized wave function and the electronic degrees of freedom of the surrounding liquid [8]. The SCIPCM was preferred over other SCRF implementations because the solute cavity is determined self-consistently without the need to specify any parameters, whereas in most other approaches the energy would strongly depend on the cavity size and shape.

Before we describe the computational details for the calculation reported herein, it is useful to revisit how our new SCEE method compares to the typical EE and PE approaches. SCEE is closer to PE in the sense that the dipole moment of surrounding MM molecules is able to respond self-consistently to the presence of the central QM molecule, i.e. it is polarized, whereas in EE the surrounding electric field is fixed. However, SCEE is closer to EE in assuming that the dipole moments of the surrounding molecules are all the same, whereas in PE each surrounding molecule will have a different dipole moment, determined self-consistently. Furthermore, in SCRF we assume that the structure of the liquid (specifically, the local interatomic correlations) is well described by the MM model from which the configurations are harvested. This assumption is shared with EE, but only partly with PE – in the latter, if a self-consistent iterative process is used, which is not always the case, the structure of the polarizable model can adjust to the presence of the QM molecule. Finally, in the SCEE approach the energy of the QM/MM system is minimized at each step of the iteration, but convergence is determined by the value of the dipole moment. In contrast, both EE and PE use the total energy of the QM/MM system to determine convergence. We will discuss the relevance and validity of these assumptions later in the paper.

2.2 Computational Details

MD simulations were performed using GROMACS version 5.1.2 [84, 85] using the Verlet leap-frog algorithm [86] to integrate the equations of motion. A time step of 2 fs was used for the non-polarizable models, while 1 fs was used for the polarizable model. Each periodic cubic box contained approximately 900 water molecules, so that its length was approximately 3 nm. Except where noted, simulations were run in the NpT ensemble, with temperature controlled at 298 K by a Nose-Hoover thermostat [87, 88] with a coupling constant of 2 ps, and pressure controlled at 1 bar by a Parrinello-Rahman barostat [89] with a

coupling constant of 1 ps and a compressibility of $4.5 \times 10^{-5} \text{ m}^3/\text{bar}$. The water molecules were kept rigid by applying the SETTLE constraint algorithm [90]. A cut-off of 1.2 nm was used for the Lennard-Jones potential, with long-range dispersion corrections added to both energy and pressure, while long-range electrostatic interactions were accounted for by using the particle-mesh Ewald method [91]. All MD simulations were run for 5 ns, with 200 configurations being harvested from the last 4 ns of each run. We note that the spacing of 20 ps between each successive configuration ensures that they are sufficiently uncorrelated to allow efficient statistical sampling, as discussed in detail by Coutinho et al. [62].

The QM/MM calculations were performed with Gaussian 09 [92]. We tested several DFT exchange-correlation functionals – B3LYP [28, 93], CAM-B3LYP [94], M05 [95], M06-2X [96], PBE0 [35], ω B97X-D [97], BMK [98], and B2PLYP [99] – and compared the results to those obtained using the Hartree-Fock (HF) and second-order Moller-Plesset perturbation theory (MP2) approaches. It is well known that large basis sets that include diffuse functions are needed to provide accurate values for the electronic properties of molecules, even in the gas phase [100, 101]. In this work, we considered the 6-311G++(3df,3pd), 6-311G+(3df,3pd), and 6-311G(3df,3pd) split valence basis sets of Pople et al. [102] (henceforth abbreviated as PZ++, PTZ+ and PTZ, respectively), as well as Dunning's correlation-consistent basis sets aug-cc-VXZ (X = D, T, Q) [103] (henceforth abbreviated as DDZ, DTZ and DQZ, respectively). For each configuration, the geometry of the central molecule was optimized while maintaining the oxygen atom at the center of the cluster – i.e., the oxygen was kept fixed, while both hydrogen atoms were allowed to relax with tight convergence criteria. However, the results from the initial step (i.e., corresponding to a single-point calculation without optimization) were also analyzed to assess the effect of geometry relaxation. Scripts for processing Gaussian calculations are provided in Supporting Information.

In a subsequent set of calculations, we surrounded each QM/MM cluster with a uniform continuum described by the self-consistent reaction field (SCRF) approach. For these calculations, we used the IEFPCM method [104] with the default parameters from Gaussian 09 and the “solvent=water” keyword. This means that the dielectric constant of the surrounding continuum takes the experimental value of 78.4 for water. For these “SCEE/SCRF” calculations, the cavity radius for the dielectric continuum was made equal to R_{cut} . However, in this case we only kept surrounding MM molecules when *all* their sites (i.e. oxygens, hydrogens and dummy sites, when applicable) were within the cut-off sphere to

avoid the presence of point charges within the continuum region. We recall that in SCRF calculations, mutual polarization between the central QM region and the surrounding continuum are taken into account through an iterative procedure. As we will discuss below, these calculations turn out to be much more computationally expensive than simple EE.

3. Results and discussion

3.1 Implementing and testing the SCEE method

As described above, the aim of our SCEE method is to find the dipole moment that is self-consistent with respect to the QM/MM procedure, i.e. the value for which the dipole of the QM and MM molecules is the same. The most natural way to do this is to carry out an iterative process (Figure 1a) with the following steps: i) start with an initial guess for the dipole moment and assign point charges to the MM molecules that reproduce that dipole moment; ii) optimize the QM molecule under the effect of the electric field of the MM molecules, and calculate its dipole moment from the polarized wave function; iii) assign new charges to the MM molecules that reproduce the value of the QM dipole moment and repeat the process until convergence. Figure 2 shows the outcome of this iterative process for a single MD configuration of the SPC/E model starting from two different initial guesses: i) the dipole moment of the SPC/E model (i.e., 2.35 D); ii) a large dipole moment of 3.33 D. In the first case, the calculated QM dipole is higher than the initial MM dipole and the calculation converges to 2.725 D after 7 iterations. In the second case, the initial dipole is too high, and the calculation also converges to the same value after 7 iterations. In Figure 2b, we can see that there is still a small difference between the two converged dipoles, but we stopped the procedure after the deviations were smaller than the third decimal point. Already at this stage, the impact of using the SCEE approach instead of standard EE becomes apparent – the dipole moment obtained with EE corresponds to the value at the end of the first iteration of Path 1, i.e. 2.609 D, which is 0.116 D lower than the converged SCEE value. As we will see below, this effect is even more pronounced after averaging over a large number of MD configurations.

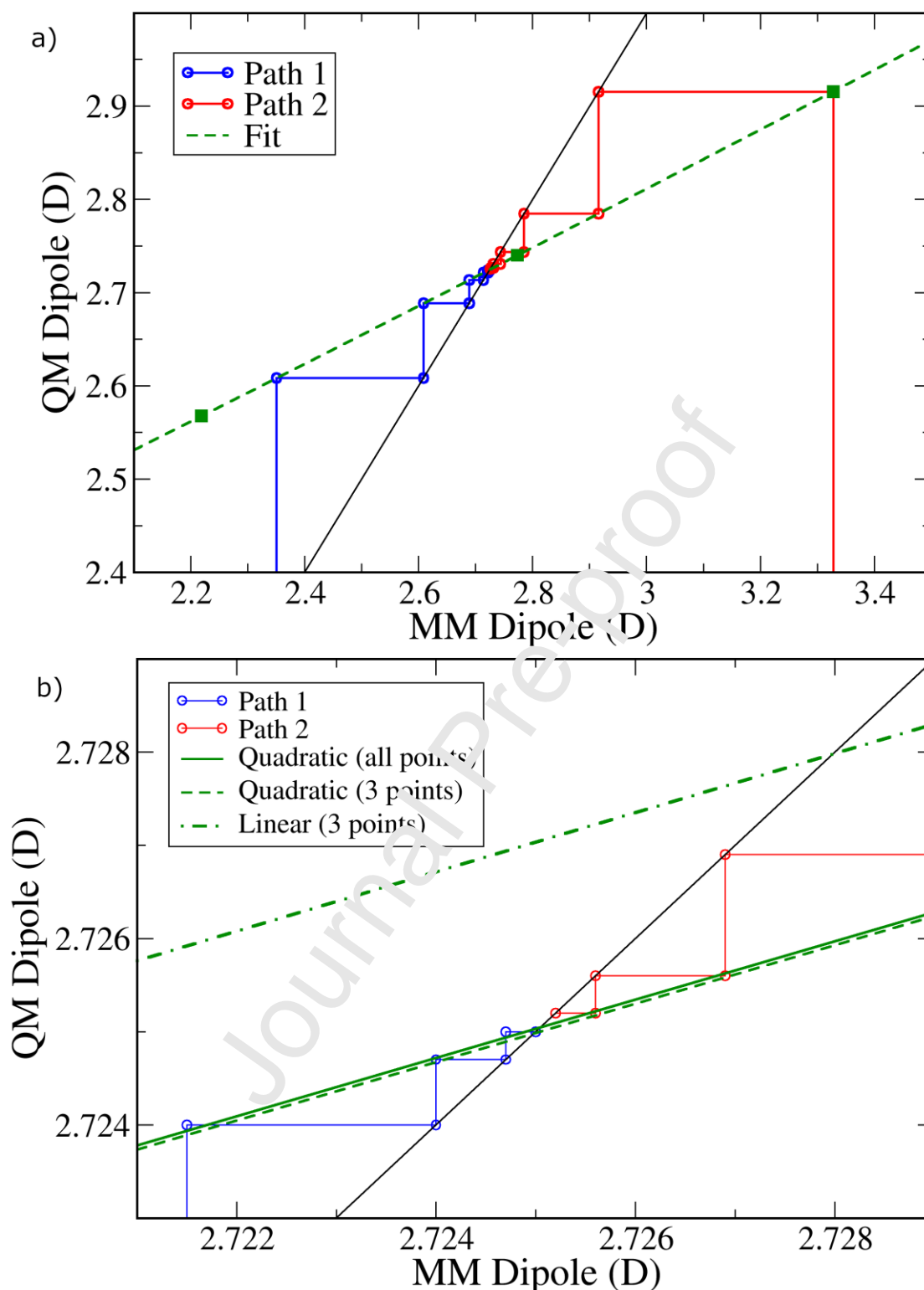


Figure 2 – (a) Comparison between the iterative SCEE procedure starting from either the SPC/E dipole moment (Path 1: blue line) or from a value of 3.33 D (Path 2: red line), and a quadratic fit (dashed green line) over three individual EE calculations (green squares). Panel (b) shows a close-up in the vicinity of the converged dipole and compares also the results of a linear fit over the same three EE points (dashed-dotted green line) and of a quadratic fit over all EE points obtained during the iterations (full green line). In both panels, the black line represents the diagonal for which the QM and MM dipoles are identical. Calculations were done on a single MD configuration for the SPC/E model using the B3LYP/PTZ++ level of theory with optimization of the hydrogen atoms.

After analyzing the results, it was clear that the dependence between the QM and MM dipole moments was nearly linear. We note that this can also be inferred from the data in Table 1 of Kongsted et al. [74]. Upon closer inspection, however, it was found that the deviations from linearity were non-negligible, so we decided to apply a second-order polynomial fit to the data. When all the data points obtained from the two iterative procedures were incorporated in the fit, the full green line in Figure 2b was obtained. As can be expected, this line intercepts the diagonal at a dipole value that is consistent with the outcome of both iterations, within the numerical tolerance (procedure in Figure 1b). More importantly, the results from a quadratic fit to only 3 points, evenly spread over the range of interest (see squares and dashed line in Figure 2a) leads to nearly indistinguishable results. This allows for significant savings in computer time, given that only 3 instead of 7 EE calculations are required. Moreover, it makes it easier to automate the calculation, since the same 3 points can be used for all configurations; in contrast, each configuration would need to be iterated separately if an iterative procedure was employed. We also tested the possibility of applying a linear fit, instead of a quadratic fit, over the same 3 points (dashed-dotted line in Figure 2b). The linear fit led to a small (0.003 D) overestimation of the dipole moment with respect to the full iterative procedure for this particular configuration. We carried out a statistical comparison between the two fitting procedures over 200 SPC/E water configurations (see section 3.2 for details of configuration averaging), and found that the result shown in Figure 2b is representative – in 185 out of 200 configurations, the linear fit overestimated the dipole moment, and the average signed deviation was +0.005 D. This is only 0.2% of the average liquid dipole moment, which is well within the statistical uncertainty of the SCEE method. However, the slight overestimation is systematic, and therefore, due to the negligible difference in computational cost between the two fitting methods, we have employed the quadratic fit in the remaining analysis.

At this point, it is worth comparing the SCEE approach proposed here to a similar idea employed by Tu and Laaksonen [71]. The main difference is that in their approach, it was the point charges that were iterated to convergence, whereas we have used the dipole moments. We repeated the iteration (Path 1) of Figure 2 but using the approach of Tu and Laaksonen instead – i.e., a new set of point charges was computed after the first QM optimization, and these charges were used for the MM molecules in the subsequent step. We used both the ESP method [110] as applied by Tu and Laaksonen, and the more recent CHelpG method [111] to calculate the QM point charges. The QM dipole moments after

convergence of the charges were 2.690 D for CHelpG and 2.695 D for ESP, which were both somewhat lower than the value of 2.725 D from our SCEE procedure. The difference is due to the fact that ESP-derived charges are unable to fully describe the underlying electron density. Given that a certain degree of approximation is required when connecting the QM with the MM Hamiltonians, we find it much more reasonable to ensure that the dipole moment is consistent in those two descriptions.

The next thing we consider is the convergence of the SCEE dipole with the size of the system, namely the number of MM water molecules. For that purpose, we generated clusters of increasing R_{cut} , ranging from 0.4 to 1.5 nm, taken from the same MD configuration. In the smallest cluster, only the first solvation shell is included, while the largest cluster includes more than 400 surrounding molecules. In Figure 3, we show the results of this test for a single configuration obtained with the SPC/E model. It can be seen that the results for 0.85 nm and beyond are not far from the converged value, but full convergence, in the sense that the variation is restricted to the third decimal point, is only achieved for $R_{\text{cut}} = 1.2$ nm. We note that Coutinho et al. [62], in their EE calculations of liquid water, observed that the dipole moment was approaching convergence at 1.15 nm, in agreement with our analysis. Similarly, Osted et al. [77] showed that a cut-off radius of at least 1.1 nm was needed for convergence. The fact that small water clusters lead to underestimated dipole moments has implications in the comparison of our results with those of previous studies, as discussed in section 3.3. We have also observed that the computational time of the QM/MM calculation is practically insensitive to the number of MM charges, at least in the Gaussian 09 implementation. As such, we have used a value of $R_{\text{cut}} = 1.5$ nm in all subsequent calculations, except where otherwise noted.

Figure 3 also compares the results obtained with (red line) and without (blue line) optimization of the hydrogen atoms of the central water molecule. It is clear that relaxation of the positions of the hydrogen atoms leads to an increase in the degree of polarization, manifested in a systematically higher dipole moment. The difference for the largest cluster is about 0.07 D, which is not insignificant. Although the classical models are designed to capture, in a mean field sense, the average polarization of water in the liquid state, they are not able to adapt to local fluctuations in the electric field, which distort the molecule and, quite often, give rise to asymmetric geometries. This is particularly the case when we are dealing with hydrogen-bonding molecules like water. We believe it makes physical sense to allow the geometry of the QM molecule to relax in response to the local electric field, and

that calculations that enforce a rigid geometry are underestimating the true consequences of polarization. We note that a recent study has also found that using configurations from classical MD followed by a constrained QM/MM optimization provided the best compromise between accuracy and computational speed [112]. At this point, we note that the vast majority of previous QM/MM calculations on water have considered fixed geometries for both the QM and MM molecules throughout the calculation.

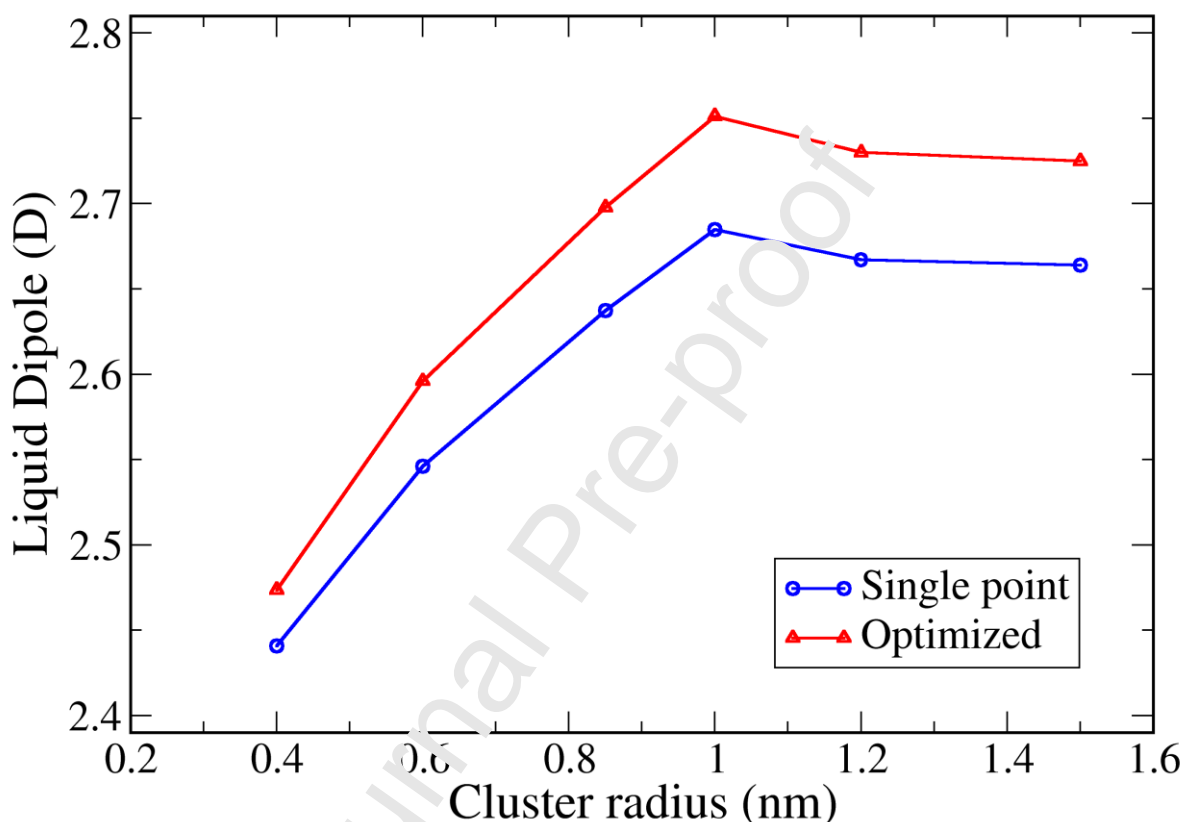


Figure 3 – SCEE dipole obtained from the same configuration, but with increasing number of surrounding MM molecules, determined by the value of R_{cut} . Calculations were done on a single MD configuration for the SPC/E model using the B3LYP/PTZ++ level of theory. The red line shows results with optimization of the water hydrogens, while the blue line corresponds to results without geometry optimization.

The next variable we analyze is the level of theory of the QM calculation. In Table 1, we report results obtained using several functionals and basis sets for the dipole moment in vacuum (μ_G), the liquid dipole moment (μ_L) for a single SPC/E configuration, and the associated dipole moment shift (the difference between the liquid and the gas phase values, $\Delta\mu$). When comparing results obtained with different levels of theory, $\Delta\mu$ is a better figure of merit for our purposes than μ_L , as it takes into account eventual inaccuracies in predicting the gas phase dipole moment and quantifies only the ability of a particular method to polarize the central molecule. We will thus refer primarily to $\Delta\mu$ when comparing our results to those of

previous studies. We note also that the dipole moment shift as calculated here is identical to the induced dipole moment if the latter is collinear with the permanent dipole moment. In practice, only small differences in orientation between the permanent and induced dipole moments have been observed for water [73] and hence the dipole shift is an appropriate measure of the extent of polarization.

Our first observation following from the data in Table 1 is that, for the B3LYP functional, all of Dunning’s augmented basis sets yield gas-phase dipole moments in very close agreement with the experimental value of 1.855 D [16]. In contrast, the PTZ++ and PTZ+ basis sets consistently overestimate μ_G , albeit by a rather small amount. Crucially, however, $\Delta\mu$ is seen to be relatively insensitive to the choice of basis set, except perhaps for DDZ. In contrast, the PTZ basis set, which is the only one that does not include diffuse functions, significantly underestimates μ_G and leads to a much lower value of $\Delta\mu$. This confirms previous conclusions that diffuse functions are essential to accurately describe dipole moments [100, 101]. Although some studies have shown that using large basis sets with diffuse functions in QM/MM calculations can lead to the so-called “electron spill-over” effect [113], the consistent values of the dipole moment shift obtained with basis sets of different size in this work (Table 1) suggest that such effects are likely to be minor here. On the whole, and based also on previous studies [100, 101], we decided to use the DTZ basis set in subsequent calculations, as it strikes a good balance between accuracy and computational speed.

Table 1 – Results for the gas phase dipole moment (μ_G), the liquid phase dipole moment (μ_L), and the dipole moment shift ($\Delta\mu$) obtained using different functionals and basis sets. Results are for a single configuration collected from a simulation with the SPC/E model.

Functional/Theory	Basis Set	μ_G (D)	μ_L (D)	$\Delta\mu$ (D)
B3LYP	PTZ++	1.890	2.725	0.835
B3LYP	PTZ+	1.892	2.723	0.831
B3LYP	PTZ	1.738	2.400	0.662
B3LYP	DDZ	1.854	2.664	0.810
B3LYP	DTZ	1.847	2.689	0.842
B3LYP	DQZ	1.846	2.709	0.863
M06-2X	PTZ++	1.913	2.720	0.807
M06-2X	DTZ	1.896	2.717	0.821

M05	DTZ	1.838	2.666	0.828
PBE0	DTZ	1.854	2.675	0.821
ω B97X-D	DTZ	1.859	2.694	0.835
CAM-B3LYP	DTZ	1.879	2.722	0.843
BMK	DTZ	1.901	2.730	0.829
B2PLYP	DTZ	1.850	2.682	0.832
MP2	DTZ	1.859	2.690	0.831
HF	DTZ	1.939	2.680	0.741

It has been shown that the choice of functional can also play an important role in the accuracy of electronic property calculations [100, 101]. We have tested some of the functionals that performed best for non-spin-polarized molecules in the extensive benchmark study of Hait and Head-Gordon [101]. Our choices were also limited to functionals that were available in Gaussian 09. We also calculated the dipoles with MP2 and HF approaches for comparison. It can be seen from Table 1 that all functionals perform rather well at predicting the gas-phase dipole moment of water. Furthermore, the results for $\Delta\mu$ are not very sensitive to the choice of functional, with values for the DTZ basis set varying between 0.82 and 0.84 D. The estimate obtained with MP2 falls squarely within this range, but the result of HF calculations is significantly lower. This agrees with previous observations that HF performed systematically worse than higher-level wave function methods [60, 62, 69, 75]. We decided to use B3LYP in our calculations, again based on a balance between accuracy and computational speed, and also in the interest of generality, since this is a functional that is widely available in computational chemistry codes.

3.2 Configurational sampling

All the tests presented in section 3.1 were carried out for a single MD configuration, but a pre-requisite for obtaining statistically meaningful results is to average over a large number of uncorrelated configurations [59]. In Figure 4, we show typical results for the convergence of μ_L with the number of MD configurations sampled. We selected those two examples because they represent the best (Figure 4a) and worst (Figure 4b) case scenarios – similar plots for other systems are provided in Supporting Information (Figure S2). For the SPC/E model (Figure 4a), the calculation appears to converge after about 60 configurations. However, such a conclusion is premature, as for some systems significant fluctuations were

still observed after more than 100 configurations (see Figure 4b). It is important to note that these differences are not related to the underlying molecular model but are simply a consequence of sampling different local arrangements around the central water molecule, which arise from thermal fluctuations of the liquid state. Overall, we found that, even in the most challenging cases, 150 configurations were sufficient to obtain converged results. This is significantly larger than the estimate of 35 uncorrelated configurations suggested by Coutinho et al. [62], although in their study they considered a maximum of 50 configurations as a benchmark. At the other end of the spectrum, Osted et al. [77] suggested that 500 configurations are necessary to obtain converged values for non-linear optical properties of water, although they mentioned that the dipole moment converged faster. Given that they used a maximum spacing of 3 ps between configurations, which may imply some degree of correlation between successive sampling points [59], our result of 150 configurations with a spacing of 20 ps appears consistent with theirs. Nevertheless, to obtain improved statistics, we used all 200 configurations in subsequent calculations.

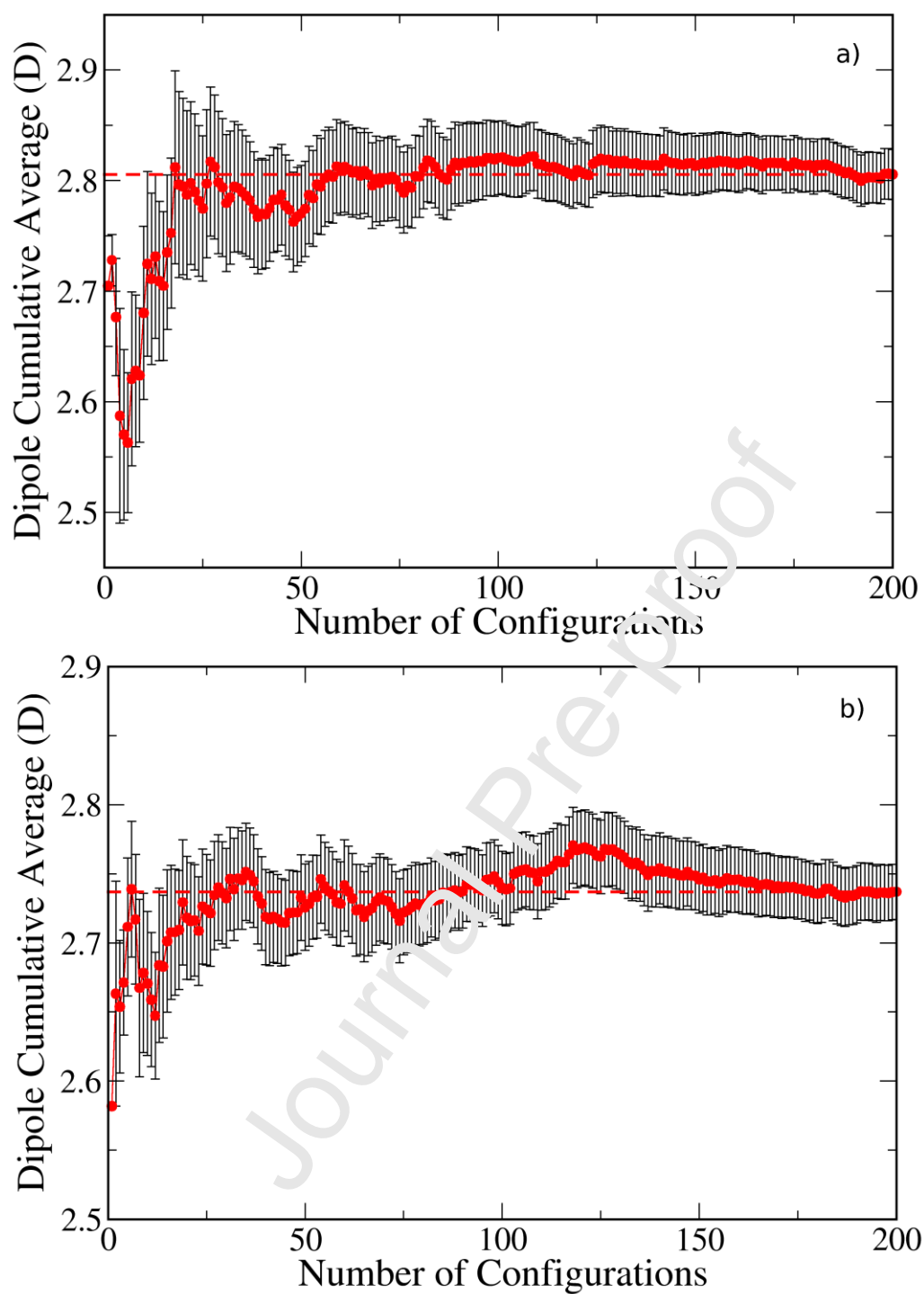


Figure 4 – Cumulative average of the liquid phase dipole moment calculated at the B3LYP/DTZ level as a function of the number of MD configurations sampled. The error bars represent the standard error of the mean, while the dashed red line shows the converged value. Panel (a) is for the SPC/E model and (b) is for the TIP3P model in the NVT ensemble (see text for details).

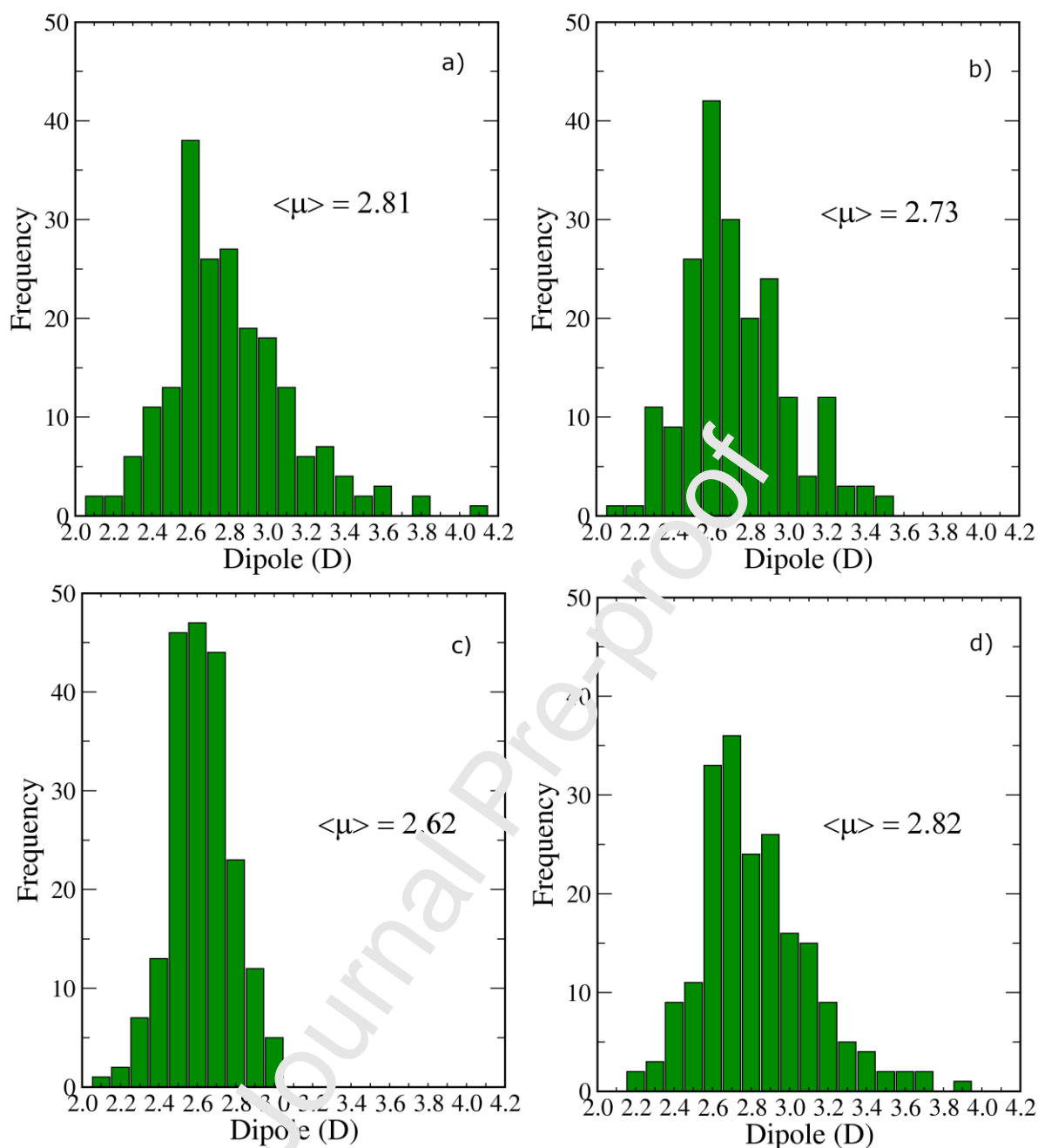


Figure 5 – Liquid dipole moment distributions over the same 200 MD configurations obtained with the SPC/E model for different calculation methods: (a) B3LYP/DTZ with SCEE and optimization of water hydrogens; (b) B3LYP/DTZ with SCEE but without geometry optimization; (c) B3LYP/DTZ with standard EE and optimization of water hydrogens; (d) M06-2X/PTZ++ with SCEE and optimization of water hydrogens. The corresponding average dipole moment is included in each panel.

In Figure 5, we show the distributions of liquid dipole moments over the same 200 configurations, obtained with the SPC/E model, using four different calculation methods. The average results are reported in Table 2. Figure 5a corresponds to the selected B3LYP/DTZ level of theory using the standard SCEE approach, i.e. the positions of the hydrogen atoms of the central molecule are optimized and μ_L is determined self-consistently for each individual

configuration, as described above. In Figure 5b, we compare the results obtained without geometry optimization. Confirming our previous observations for a single configuration (Figure 3), the effect of hydrogen optimization is to increase the degree of polarization. The average dipole is about 0.08 D lower when the central QM molecule is not allowed to relax during the calculation. Figure 5d shows results obtained using the M06-2X/PTZ++ level of theory. This was chosen because it showed the largest μ_G among all combinations studied here (except HF, see Table 1). As we can see, the distribution is very similar to the one obtained with B3LYP/DTZ. The average liquid dipole moment is somewhat larger but, because of the overestimation of μ_G , $\Delta\mu$ is actually lower by 0.05 D (Table 2).

Table 2 – Dipole moments obtained after averaging over the same 200 MD configurations obtained with the SPC/E model using different approaches.

Level of theory	Geometry ¹	Procedure ²	μ_G (D)	μ_L (D)	$\Delta\mu$ (D)
B3LYP/DTZ	opt	SCEE	1.847	2.81 ± 0.02	0.96 ± 0.02
B3LYP/DTZ	no-opt	SCEE	1.847	2.73 ± 0.02	0.88 ± 0.02
B3LYP/DTZ	opt	EE	1.847	2.62 ± 0.02	0.78 ± 0.02
M06-2X/PTZ++	opt	SCEE	1.913	2.82 ± 0.02	0.91 ± 0.02

¹ - opt = optimization of the QM molecule hydrogen atoms; no-opt = QM molecule held fixed; ² - SCEE = self-consistent calculation of the dipole moment for each separate configuration; EE = the MM charges are kept fixed at the SPC/E values

It is worth reiterating that all the functional/basis-set combinations considered in this work have been shown to perform well in predicting gas-phase dipole moments in previous benchmark studies [100-101]. Therefore, our observed difference of ~0.05 D between B3LYP/DTZ and M06-2X/PTZ++ calculations gives us an idea of the uncertainty associated with a particular choice of “good” QM method. This does not mean, however, that we should expect *any* QM method to yield results within this margin of error. Indeed, most early QM/MM studies of solvation made use of semi-empirical approaches to circumvent limitations in computer power, and they obtained very small values of $\Delta\mu$, between 0.25 and 0.45 D [52, 66, 67, 78]. This is due to the well-known limitations of semi-empirical approaches to describe highly polar and hydrogen-bonding systems [78]. Later studies that compared different wave-function methods showed that HF gives rise to significantly underestimated values of $\Delta\mu$ when compared with higher-level theories like MP2 or coupled cluster (CC) approaches [60, 62, 69, 75]. The importance of employing a sufficiently large basis set that includes diffuse functions was also recognized previously [53, 60, 62, 64, 68, 72, 73, 76]. In summary, results of previous studies using semi-empirical or HF approaches,

as well as those that employ basis sets that are too small or do not contain diffuse functions, should be regarded with care.

It is important to compare the result of our SCEE approach, where the dipoles of the MM molecules are adjusted self-consistently until they match the QM dipole of the central molecule, to the result of a standard EE calculation, in which the dipole of the MM molecules is kept constant at the value of the classical fixed-charge model (Figure 5c). In this case, for the SPC/E model, the MM dipole is 2.35 D and this corresponds to an average QM dipole of 2.623 D (averaged over all 200 configurations, at the B3LYP/DTZ level with hydrogen atom optimization). This value is lower than the corresponding SCEE result by nearly 0.2 D. In this context, it is very interesting to notice that previous EE calculations of liquid water using adequate levels of theory yielded values of $\Delta\mu$ between 0.55 D and 0.74 D [53-55, 58, 61, 62], which are indeed about 0.2 D lower than the values obtained with our SCEE approach (see Table 2). This confirms the need to allow the dipole moments of the surrounding MM molecules to adapt self-consistently during the calculation.

We also repeated the above EE calculation but fixing the dipole moment of the MM molecules at the gas-phase value (1.855 D) instead of at the value of the SPC/E model. In this case, the resulting average QM dipole was only 2.457 D, or about 0.35 D lower than the SCEE value. This difference can be interpreted as the effect caused by polarizing the MM molecules in the calculation. Previous QM/MM calculations using the polarizable embedding approach also obtained differences of similar magnitude [64, 73-76], which shows that our SCEE method is correctly capturing the effects of liquid-phase polarization.

As a final test of the technical aspects of the SCEE approach, we examined the effect of adding a dielectric continuum to the QM/MM cluster, as this was an approach used in some previous studies to represent the long-range effects of the solvent on solute polarization [56, 57, 68, 69]. We first note that when a single QM water molecule is optimized in an IEFPCM continuum (i.e., without any surrounding MM charges) using the Gaussian 09 default SCRF parameters, the dipole moment is only 2.165 D, which is much lower than any QM/MM, AIMD or experimental estimates. This has also been observed previously [56, 72, 75, 76] and reflects the inability of dielectric continuum approaches to represent the local effects of polarization such as the contribution due to hydrogen bonding. In solvents of low polarity, where such interactions are absent, continuum methods are likely to more accurately describe the degree of induced polarization.

When our QM/MM clusters were surrounded by a continuum, we noticed that the computational time dramatically increased with R_{cut} (see Figure S3). This is likely because the PCM approach works by representing the effects of the surrounding solvent by a set of effective charges on the cavity surface, and thus the computational effort will increase with the size of the surface. This issue became so extreme that all calculations with $R_{\text{cut}} > 0.85$ nm produced a runtime error and did not complete. Therefore, we used $R_{\text{cut}} = 0.85$ nm as the cluster size in SCEE/SCRF calculations. Averaging over all 200 configurations, we obtained a dipole moment of 2.816 D (see Table 3). Comparing with a standard SCEE calculation with the same value of $R_{\text{cut}} = 0.85$ nm, this yields an increase of only 0.011 D caused by the presence of the dielectric continuum. In contrast, increasing R_{cut} from 0.85 to 1.5 nm (without a continuum) increases the dipole moment by 0.024 D, which is a larger effect than adding the dielectric continuum. This confirms our previous assertion that the SCEE calculations on the larger clusters are converged and capture all of the long-range effects of polarization.

Table 3 – Dipole moments obtained after averaging over the same 200 MD configurations obtained with the TIP4P2005 model using the B3LYP functional and the PTZ++ basis set with and without a surrounding dielectric continuum. The row labelled “SCRF” represents a single IEFPCM calculation with default parameters, since there are no surrounding MM molecules.

Approach	R_{cut} (nm)	Procedure	μ_{G} (D)	μ_{L} (D)	$\Delta\mu$ (D)
SCEE	1.5	SCEE	1.890	2.83 ± 0.02	0.94 ± 0.02
SCEE	0.85	SCEE	1.890	2.81 ± 0.02	0.92 ± 0.02
SCEE/SCRF	0.85	EE	1.890	2.82 ± 0.02	0.93 ± 0.02
SCRF	default	n/a	1.890	2.17	0.28

3.3 Effect of the classical molecular model

One of the key choices in any QM/MM approach is the molecular model to represent the MM part of the system. Very little attention has been paid to this issue, with the majority of previous QM/MM studies on water using the rather flawed TIP3P fixed-charge model [52, 54-59, 68-71]. In Table 4 we compare the results obtained from SCEE calculations with four fixed-charge models. The corresponding dipole moment distributions are shown in Figure 6 (see also Figure 5a for SPC/E). The first important observation is that TIP3P leads to a consistently lower liquid dipole moment when compared with the other fixed-charge models. This is most likely due to the shortcomings of this model in describing the structural and thermodynamic properties of water [42, 43]. In particular, it yields a bulk liquid density at

298 K (980 kg/m³ [43]) that significantly underestimates the experimental value (997.04 kg/m³), while all other models considered here are able to predict that property quite accurately. Given that an incorrect density is likely to affect the local packing and hence the molecular dipole moment, we have carried out an additional calculation with the TIP3P model at a fixed density (i.e. in the *NVT* ensemble), equal to the experimental value at 298 K. Perhaps surprisingly, the resulting liquid dipole moment is only slightly larger than for the original *NpT* calculation; in fact, the two values are statistically equivalent. This indicates that the main limitation of the TIP3P model is related to its incorrect representation of the local structure of the liquid [114]. Indeed, it has been shown that TIP3P leads to a consistently poorer description of the radial and angular distribution functions in liquid water when compared with the other non-polarizable models used here [1–5]. The other three fixed-charge models yield results that are all within statistical error of each other, reflecting their significantly more accurate representation of the local water structure and thermodynamic properties of the liquid.

Table 4 – Dipole moments and polarization energies obtained from SCEE calculations with four different fixed-charge models. In all cases, the results are averages over 200 MD configurations with the B3LYP/DTZ level of theory and hydrogen optimization of the central molecule. For the TIP3P model, we also report results from an *NVT* calculation at the experimental density of water (termed TIP3P (*NVT*)).

Model	Method	μ_L (D)	$\Delta\mu$ (D)	E_{Elec} (kJ/mol)	E_{Dist} (kJ/mol)	E_{Tot} (kJ/mol)
SPC/E	SCEE	2.81 ± 0.02	0.96 ± 0.02	-32.7 ± 0.5	32.0 ± 1.4	-0.7 ± 0.9
TIP3P	SCEE	2.72 ± 0.02	0.87 ± 0.02	-30.2 ± 0.4	24.2 ± 1.1	-6.0 ± 0.7
TIP3P (<i>NVT</i>)	SCEE	2.74 ± 0.02	0.89 ± 0.02	-30.5 ± 0.4	24.8 ± 1.0	-5.7 ± 0.6
TIP4P2005	SCEE	2.31 ± 0.02	0.97 ± 0.02	-32.8 ± 0.4	32.1 ± 1.2	-0.7 ± 0.7
TIP4P-FB	SCEE	2.80 ± 0.02	0.96 ± 0.02	-32.4 ± 0.4	30.5 ± 1.1	-1.9 ± 0.7

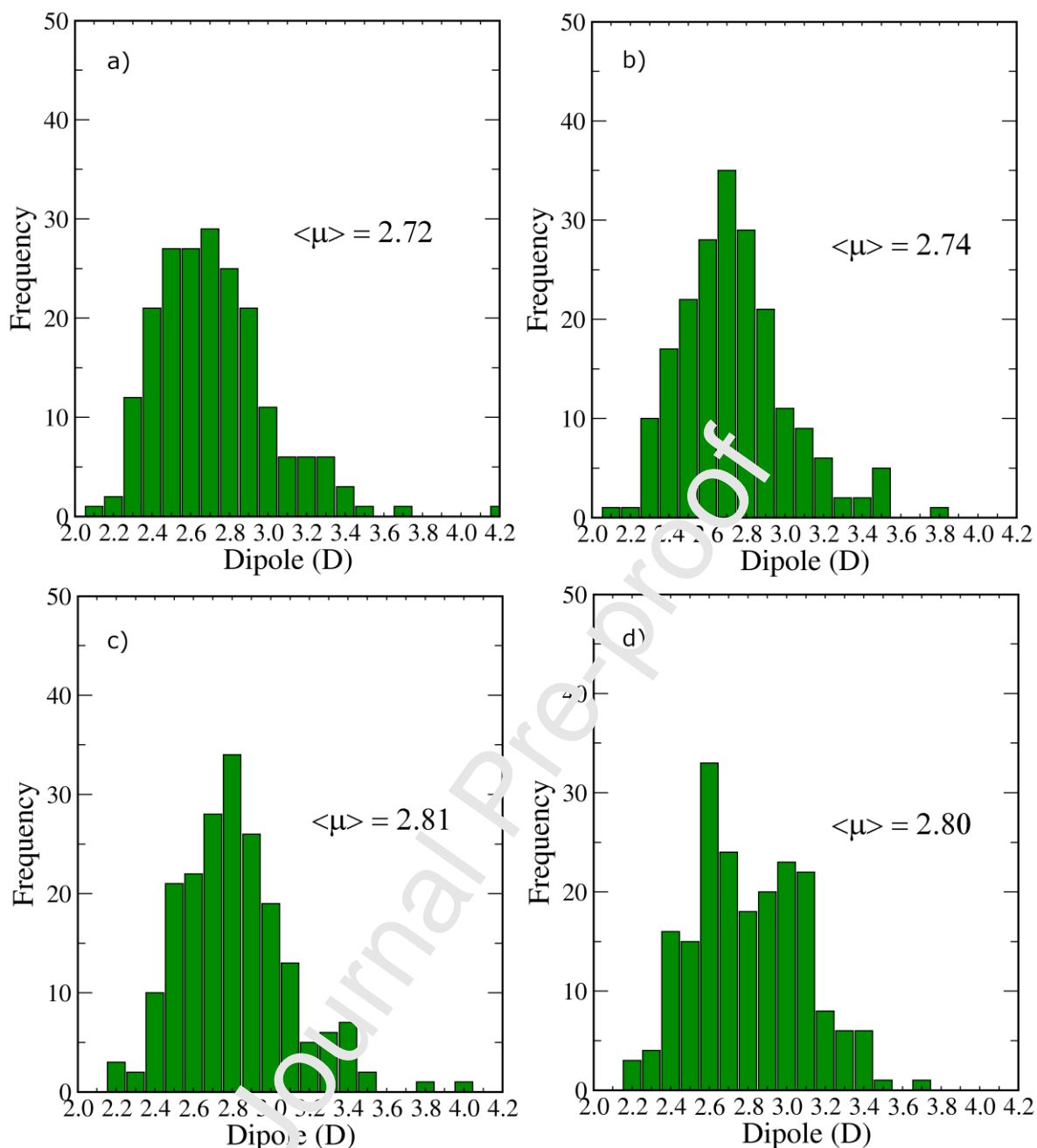


Figure 6 – Liquid dipole moment distributions over 200 MD configurations obtained at the B3LYP/DTZ QM level with hydrogen atom optimization and using different classical models: (a) TIP3P; (b) TIP3P in the NVT ensemble at the experimental water density; (c) TIP4P2005; (d) TIP4P-FB (see also Figure 5a for the SPC/E model at the same conditions). The corresponding average dipole moment is included in each panel.

A major assumption of our SCEE method when using fixed-charge models in the MM part of the system is that the dipole moment of all surrounding molecules takes on the same value (as implied by the fixed-charge nature of those models). In reality, the dipole moment of each surrounding molecule will adopt a different value depending on the specific environment felt by that molecule. To test the validity of this assumption, we carried out SCEE calculations on 200 configurations obtained from MD simulations of a classical

polarizable model, SWM4-DP [48]. Because this model is based on the Drude oscillator approach, it allows us to vary the value of the average dipole moment of the surrounding MM molecules by simply applying a constant scaling factor to all the point charges (including those on the Drude oscillator) without changing the geometry. We then apply the SCEE approach described in Figure 1b using different values of the scaling factor, and find the self-consistent value of the liquid dipole moment. Using this approach, we obtained a liquid dipole of 2.82 ± 0.05 D (i.e. $\Delta\mu$ of 0.98 D), in perfect agreement with SCEE results obtained on configurations from all fixed-charge models except TIP3P. The dipole moment distribution (Figure S4) is also very similar to those obtained with the above-mentioned fixed-charge models. In other words, the assumption that all surrounding MM molecules have the same dipole moment does not adversely affect the outcome of the calculations.

With the information obtained from the methodological tests described above, we are in a good position to systematically compare our SCEE estimates of $\Delta\mu$ against those of previous studies. As mentioned previously, EE calculations lead to significantly underestimated dipole moments compared to approaches like PE and SCEE, which take into account the polarization of the MM region. The precise magnitude of this underestimation depends on the details of the fixed-charge model used in the MM part; for example, the difference $\Delta\Delta\mu$ between our EE and SCEE calculations is -0.18 D for SPC/E and -0.14 D for TIP3P. Furthermore, we observed that using an MM model that does not provide a realistic representation of the local water structure also leads to a systematic deviation; in this case, using TIP3P instead of the more accurate SPC/E model underestimates $\Delta\mu$ by -0.09 D (or -0.07 D if the TIP3P simulations are run at constant experimental density). It was also clear from our studies that when the structure of the QM molecule was kept fixed, as in the majority of previous studies, the dipole shift was underestimated by -0.08 D compared to calculations where the hydrogen atoms were allowed to relax.

We found that all previous studies suffered from at least one of the shortcomings described above, even when the level of theory employed in the QM calculations was sufficiently accurate. Using the values from the above paragraph as correction terms, we can bring several previous results into close agreement with our SCEE calculations. Table S2 presents a comprehensive analysis of previous QM/MM results after applying these corrections. Here, we concentrate on a few particular examples representing the “best estimates” obtained with different QM/MM approaches. For example, the best estimate of

Tuñón et al. was $\Delta\mu = 0.68$ D obtained from an EE calculation using TIP3P in the *NVT* ensemble without geometry optimization [54], which translates into $\Delta\mu = 0.97$ D when we take into account the corrections for MM polarization (0.14 D), the choice of MM model (0.07 D) and the QM geometry relaxation (0.08 D). Tu and Laaksonen [58] used a similar methodology (EE with TIP3P in the *NVT* ensemble, no optimization) and obtained $\Delta\mu = 0.65$ D, which yields $\Delta\mu = 0.94$ D after applying the same corrections. The results of Coutinho et al. [62] are more difficult to compare because they used the TIP5P water model, which was not considered here. Nevertheless, if we assume that it leads to a reasonable representation of liquid water, hence no correction for the MM model is needed, we still need to correct their best value of $\Delta\mu = 0.74$ D for polarization and geometry relaxation, which yields $\Delta\mu = 0.96$ D.

Previous PE studies on liquid water report systematically higher dipole shifts than EE calculations, as expected, but they still tend to be somewhat lower than our SCEE estimates. However, the central QM water molecules were held rigid in all of those calculations. Particularly relevant are several systematic PE calculations on water from a series of studies by Michelsen and co-workers [72-77] using CC methods. Their best estimate of the dipole moment shift, using CCSD with the d-aug-cc-VTZ basis set, large water clusters and sampling over 500 configurations, is $\Delta\mu = 0.86$ D [77]. After correcting for lack of geometry relaxation of the central QM molecule, this becomes $\Delta\mu = 0.94$ D, again in very close agreement with our SCEE calculations. The values reported by Sánchez Mendoza et al. [68, 69] need to be corrected for both geometry relaxation and the choice of MM model, since they sampled the liquid configurations using TIP3P in the *NPT* ensemble. Their best estimate of $\Delta\mu = 0.81$ D [69] becomes $\Delta\mu = 0.98$ D after correcting for both those shortcomings. Finally, the value of $\Delta\mu = 0.77$ D reported by Tu and Laaksonen [71] becomes $\Delta\mu = 0.95$ D after correcting for lack of geometry relaxation (0.08 D), for using the TIP3P model at constant density (0.07 D), and for self-consistently iterating the MM point charges instead of the dipole moments (we estimate this deviation as 0.03 D based on our calculations reported in section 3.1).

3.4 Polarization energies

In this section, we analyze the different contributions to the polarization energy that arise from our SCEE calculations. Because the liquid phase multipole moments and the energy contributions calculated with each of the SPC/E, TIP4P2005, and TIP4P-FB models

are all statistically indistinguishable (see Table S1), we have decided to compile together all the data obtained with those three models, at the B3LYP/DTZ level of theory. We note that this was only for the purpose of improving the statistical sampling, since we are now able to analyze a total of 600 independent configurations. We start by analyzing collected statistics for the main multipole moments of liquid water and their shifts with respect to the vapor phase. The dipole moment vector can be reduced to a single scalar value that is independent of the origin of the coordinate system due to the symmetry of the water molecule. This symmetry also means that the non-diagonal elements of the quadrupole tensor should be zero (or at least negligible compared to the other components). One of the non-zero diagonal components can be related to the other two by rendering the quadrupole tensor traceless, leaving us with two independent quadrupole components for water – we use the nomenclature Q_0 for the minor quadrupole term and Q_T for the major quadrupole term. Contrary to the dipole moment, the quadrupole terms are dependent on the choice of origin. In this work, they were calculated after rotating each of the central water molecules to the same Cartesian reference frame, centered on the oxygen atom. Further details regarding these calculations are reported in our previous publication [9].

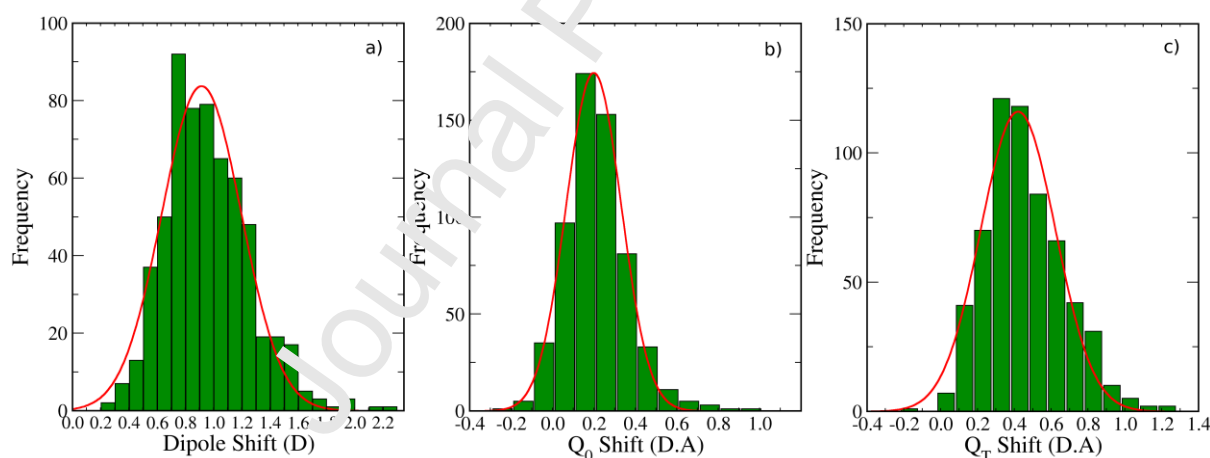


Figure 7 – Distributions of the shifts (i.e. difference between liquid and gas) in multipole moments of water: (a) dipole moment, (b) minor quadrupole moment, Q_0 , and (c) major quadrupole moment, Q_T . Results were sampled over 600 configurations obtained with the SPC/E, TIP4P2005 and TIP4P-FB models, using SCEE at the B3LYP/DTZ QM level of theory with hydrogen atom optimization. The red lines show the best fit of each distribution to a Gaussian expression.

Figure 7 shows the overall distribution (over all 600 configurations) for the shifts in dipole moment as well as for the two quadrupole components. The corresponding average values are reported in Table 5. For all multipole components, it can be seen that the distributions are not exactly Gaussian-shaped, but are somewhat skewed, showing a small tail extending to larger values. This slight asymmetry has not been observed in previous AIMD

calculations of the water dipole moment [23-26, 31-34, 116, 117], and is possibly due to the fact that the geometry of the central molecule is optimized in the SCEE method for each configuration. In terms of the average values, all of the multipole moments are significantly enhanced with respect to the vapor. Although the absolute values of the quadrupole moments in the vapor are underestimated with respect to experimental values ($Q_0 = 0.11 \text{ D}\cdot\text{\AA}$; $Q_T = 2.57 \text{ D}\cdot\text{\AA}$ [16]), it is nevertheless relevant to compare the shifts to previous studies in the literature. The average shifts in the dipole and minor quadrupole moments are very close to our recent analysis [9] based on high quality AIMD trajectories [34], which yielded $\Delta\mu = 1.02 \text{ D}$ and $\Delta Q_0 = 0.19 \text{ D}\cdot\text{\AA}$. However, the shift in the major quadrupole moment is much lower in our SCEE calculations ($\Delta Q_T = 0.46 \text{ D}\cdot\text{\AA}$) than in the AIMD data ($\Delta Q_T = 0.75 \text{ D}\cdot\text{\AA}$). It is, nonetheless, of the same order as values reported by Niu et al. [118] from other QM/MM calculations. The reason for the discrepancy in the major quadrupole shift between QM/MM and AIMD is unclear at this point, although it could be related to any ambiguities in partitioning the electron density of each molecule when analyzing data from AIMD.

Table 5 – Main multipole moments for the water molecule in the liquid phase, in the gas phase, and the corresponding shift. Liquid results are averaged over 600 SCEE calculations from configurations obtained with the SPC/E, TIP4P2005 and TIP4P-FB classical models using the B3LYP/DTZ level of theory with hydrogen atom optimization, while vapor results were obtained from a single full optimization at the same level of theory.

	Vapor	Liquid	Shift
$\mu \text{ (D)}$	1.85	2.81 ± 0.2	0.96 ± 0.2
$Q_0 \text{ (D}\cdot\text{\AA)}$	0.04	0.25 ± 0.1	0.21 ± 0.1
$Q_T \text{ (D}\cdot\text{\AA)}$	1.67	2.13 ± 0.2	0.46 ± 0.2

We have also carried out a statistical analysis of the geometry of the central water molecule after being polarized by the surrounding environment. Using our SCEE method, we obtained an average bond length of 0.98 \AA and an average angle of 105° (both distributions are shown in Figure S5). As expected, due to the formation of hydrogen bonds with surrounding molecules (here represented by point charges), the O-H bond length is elongated with respect to the experimental gas-phase value of 0.957 \AA [119], while the H-O-H angle is slightly widened compared to the experimental gas-phase angle of 104.5° [119]. In this context, it is interesting to note that McGrath et al. [117] have demonstrated, through AIMD calculations, the existence of a correlation between the dipole moment enhancement in the

liquid phase and the degree of tetrahedral order of the molecule, as well as the number of hydrogen bonds formed with adjacent neighbors.

In Table 6, we show average results for several energy components, calculated as described in section 2. We also report results for the distortion, electronic and polarization correction energies obtained from approximate analytical formulas. To obtain these approximate estimates, we followed the same procedure as in our previous work [9], where full equations and computational details are presented. In short, the electronic contribution was calculated from a Kirkwood-Onsager [2, 3] solvation model in a surrounding continuum with a dielectric constant equal to the high-frequency dielectric constant of water (1.776), using only the dipole moment (labelled “Dipoles”) or both dipole and quadrupole moments (labelled “Dipoles + Quadrupoles”). For these calculations, we used a fixed cavity radius of 1.554 Å, corresponding to the experimental molar volume of the liquid. Similarly, the distortion contribution was estimated from the Swopce expressions [105], either by taking into account only the dipole moment difference between vapor and liquid (labelled “Dipoles”), or taking into account the distortion of both dipoles and quadrupoles (labelled “Dipoles + Quadrupoles”). In all calculations, we made use of experimental or high-level QM data for the dipole and quadrupole polarizability matrices, the gas-phase multipole moments, and the static and high-frequency dielectric constants of water.

Table 6 – Liquid phase energy contributions for water, in kJ/mol. Results are averages over 600 SCEE calculations from configurations obtained with the SPC/E, TIP4P2005 and TIP4P-FB classical models using the B3LYP/DTZ level of theory with hydrogen atom optimization. For some of the energy components, we also report results obtained using analytical approximations (see text and [9] for details).

	SCEE	Dipoles	Dipoles + Quadrupoles
E_{Dist}	31.6 ± 1.4	21.6 ± 1.1	27.0 ± 1.3
E_{Elec}	-32.6 ± 0.5	-6.9 ± 0.1	-31.2 ± 0.5
E_{Corr}	-1.0 ± 1.5	14.6 ± 1.1	-4.2 ± 1.8
E_{Int}	-89.8 ± 2.6	--	--
E_{Stab}	-62.2 ± 2.7	--	--
E_{Pol}	-30.6 ± 3.0	--	--

The first observation that can be made from Table 6 is that the magnitude of the stabilization energy is approximately twice that of the distortion energy. In other words, the favorable interaction energy gained by polarizing the central water molecule to a liquid-state

wave function is about twice the unfavorable energy necessary to distort the wave function from a vapor to a liquid state. This can be confirmed in Figure S6, which plots individual values of E_{Stab} against $-2 \times E_{\text{Dist}}$. The values fall very close to the diagonal, with only a slight deviation observed for configurations of very low energy. This relationship has been predicted from linear response theory and observed in previous studies of polarization energies [107, 108]. However, it is important to note that the magnitude of the individual contributions was significantly lower in those studies, since they applied SCRF methods or QM/MM with semi-empirical approaches, which significantly underestimate the extent of liquid phase polarization as discussed above. As a result of this relationship, the total polarization energy (which is just the sum of the stabilization and distortion contributions) is negative and substantial, amounting to about a third of the total interaction energy per molecule.

The balance between the distortion component and the purely electronic component of polarization, i.e. the interaction between the polarized central molecule and the electronic degrees of freedom of the surrounding liquid, is of particular interest in the context of force field development [13]. This is because they account for the two missing elements in simulations that make use of classical non-polarizable models [8], and hence yield a polarization correction that can be applied to phase-change energies and free energies [9] to improve agreement with experiment and force field transferability. Our SCEE results show that those two contributions have opposite sign and very similar magnitude, so that the total polarization correction is very close to zero (in fact, it *is* zero, within statistical uncertainty). This confirms previous assertions by ourselves [9] and others [44], now making use of a more accurate procedure for estimating the two separate contributions, and confirms that predictions of vaporization enthalpies and hydration free energies using fixed-charge water models are biased if only the distortion correction is applied.

Finally, it is interesting to assess to what extent approximate analytical expressions are able to predict the distortion and electronic contributions of polarization. In Figure 8a, we plot E_{Dist} against the dipole moment shift for each individual configuration, both from SCEE and from the Swope [105] formulas. A similar plot for E_{Elec} , but using the liquid dipole moment as the x-axis, is shown in Figure 8b, while the net polarization correction is analyzed in Figure 8c. As expected, the distortion correction tends to zero as the liquid dipole becomes closer to the gas-phase dipole and shows an approximately quadratic dependence on the dipole shift. This dependence is exact for the dipole formula, but both the SCEE results and

the approximation of Swope et al. [105] with dipoles and quadrupoles show some fluctuations. This is because the latter also take into account distortions of higher order multipole moments, up to quadrupoles for the Swope approximations and up to infinity for SCEE, since it calculates the distortion energy from differences in the total molecular wave function. Both analytical approximations somewhat underestimate the magnitude of the energy, but by including the quadrupole distortion the agreement is significantly improved. This suggests that, as conjectured in our previous work [9], multipole moments beyond the quadrupole have very small contributions to the distortion energy.

Similar conclusions can be drawn from analyzing the electronic contribution to the polarization energy (Figure 8b). The dependence is now approximately quadratic with respect to the liquid dipole moment (not the shift), and the energy tends to zero as $\mu_L \sim 0$. The underestimation of the energy magnitude by the simple dipole formula is, however, much larger than observed for the distortion correction. Consequently, the total correction estimated by assuming that only dipole moments are important is much more positive than the results of our SCEE calculations (Figure 8c). The results from the Kirkwood-Onsager model including up to quadrupole moments provide an excellent approximation of the SCEE data. Once more, this suggests that higher order multipole contributions, beyond the quadrupole, are negligible for water. Taken together, the two approximations that go up to quadrupole moments are able to capture the trend of the overall polarization correction very well (Figure 8c), despite a small underestimation at high dipole moments that leads to a slightly more negative correction term.

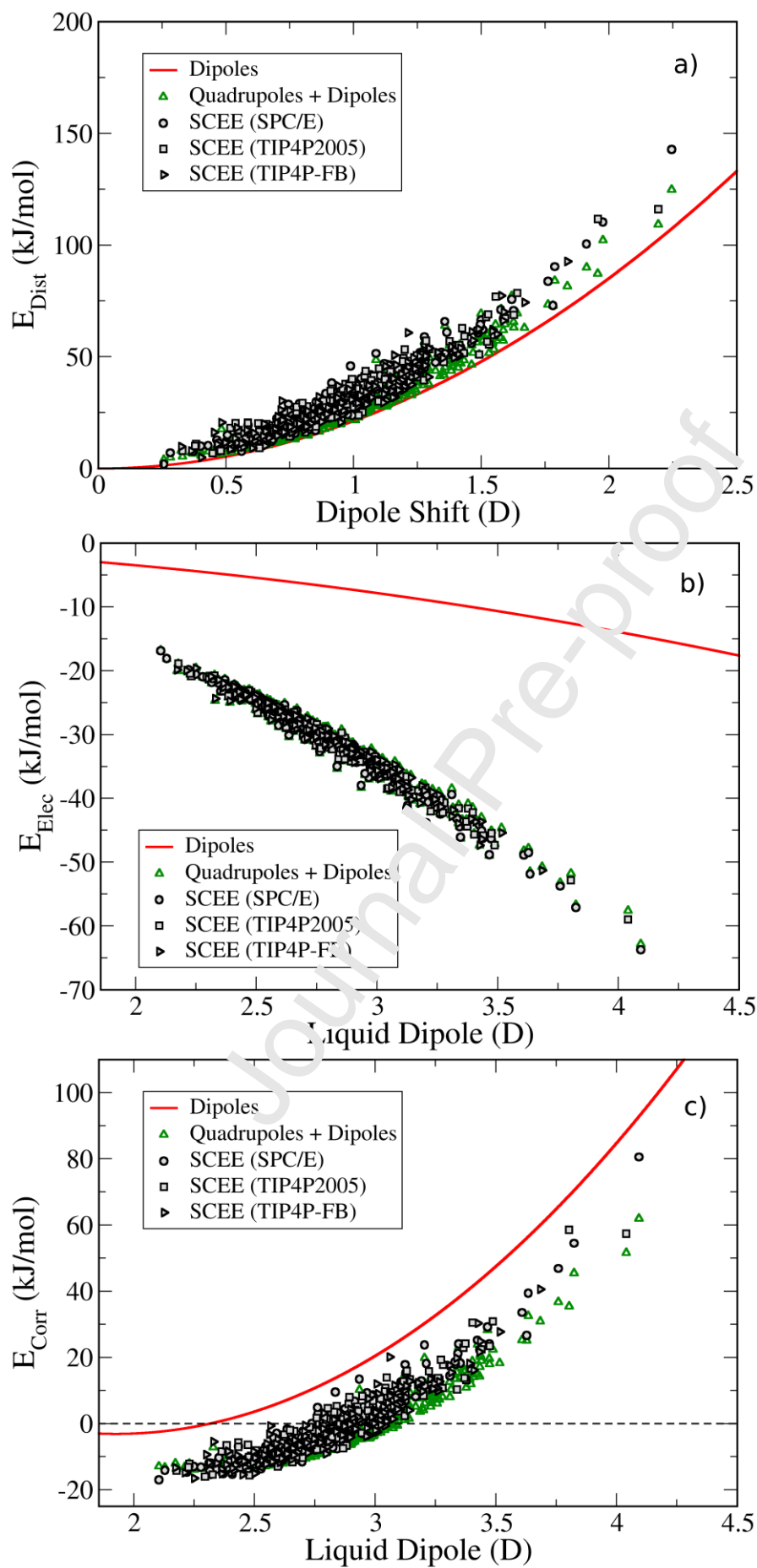


Figure 8 – Polarization energy contributions as a function of the liquid phase dipole moment for 600 configurations obtained with the SPC/E, TIP4P2005 and TIP4P-FB models (points for each separate model are denoted by different symbols): a) Distortion energy; b) Electronic polarization energy; c) Total polarization correction. SCEE results used the B3LYP/DTZ QM level of theory with hydrogen atom optimization. The analytical approximations were based on the expressions of Swope *et al.* [105] for the distortion energy and on the Kirkwood-Onsager [2, 3] solvation model for the electronic polarization energy.

3.5 Other state points

After an in-depth analysis of liquid water at ambient conditions, we now apply SCEE to two additional state points: i) liquid water at elevated temperature; ii) ice Ih. The optimal set-up determined above was applied in both cases, i.e. B3LYP/DTZ level of theory, optimization of the hydrogen atoms of the central water molecule, cluster radius of 1.5 nm, averaging over 200 configurations. For the high-temperature calculation, we selected the TIP4P2005 model for configurational sampling, set the temperature of an *NVT* ensemble simulation at a value of 500 K and kept the density identical to the equilibrium value for that model at 298 K (i.e. 997.04 kg/m³). This was done in order to isolate the effect of temperature from that of density, and to enable comparison to a previous AIMD study of Dyer and Cummings [40]. For the ice Ih calculation, we generated 200 independent proton-disordered configurations using the software GenIce of Tanaka and co-workers [120] and considering different random seeds. The resulting dipole moment distributions obtained from SCEE calculations are shown in Figure 9, while Table 7 reports the average dipole moments and polarization energy contributions for each case.

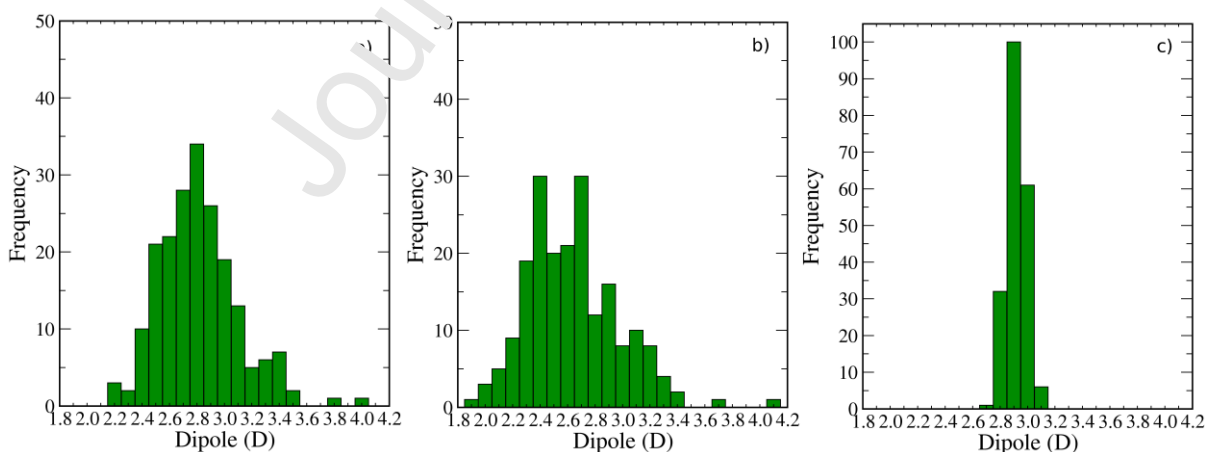


Figure 9 – Liquid dipole moment distributions over 200 configurations of water for different state points: (a) liquid water at 298 K sampled with TIP4P2005 in the NPT ensemble (average density of 997.04 kg/m³); (b) liquid water at 500 K sampled with TIP4P2005 in the NVT ensemble (constant density of 997.04 kg/m³); (c) ice Ih, with configurations sampled using GenIce [120]. The B3LYP/DTZ level of theory was employed with optimization of water hydrogens.

Table 7 – Average liquid phase dipole moments and polarization energies for water at different state points. Results are averages over 200 SCEE calculations, with configurations obtained with the TIP4P2005 model for the liquid and with GenIce [120] for ice Ih, using the B3LYP/DTZ level of theory with hydrogen atom optimization.

	TIP4P2005 (298 K)	TIP4P2005 (500 K)	Ice Ih
μ (D)	2.81 ± 0.04	2.64 ± 0.05	2.92 ± 0.01
E_{Dist} (kJ/mol)	32.1 ± 2.3	22.9 ± 2.3	70.2 ± 0.8
E_{Elec} (kJ/mol)	-32.8 ± 0.8	-29.0 ± 0.9	-45.2 ± 0.3
E_{Corr} (kJ/mol)	-0.6 ± 1.5	-6.1 ± 1.4	24.9 ± 0.5

Our results show that increasing the temperature by ~ 200 K causes a statistically significant decrease in the liquid dipole moment. We obtained a difference of 0.17 D, which is consistent with the difference of ~ 0.2 D observed by Dyer and Cummings in their AIMD calculations [40]. We note also that other theoretical studies have also shown a decrease in the dipole moment of water with temperature, although in most cases the density was also changing [34, 116, 117, 121]. We can also see that the distribution of dipole moments is somewhat wider for 500 K (*cf* Figures 9a and 9b), which reflects the higher degree of thermal fluctuations in the high-temperature liquid. An interesting observation can be seen in the polarization energies – while both the distortion and the electronic contributions decrease in magnitude, the effect of temperature is more significant for the distortion energy. As a consequence, the overall polarization correction is now significantly more negative (and statistically distinct from zero). This suggests that predictions of phase-change energies (e.g. enthalpy of vaporization) from simulations using fixed-charge force fields may require polarization corrections that are state-dependent. We aim to carry out a more in-depth analysis of this effect, and will report on our results in due course.

As for ice Ih, our SCEE calculations yield a dipole moment of 2.92 D, which is 0.11 D higher than in the liquid state at room temperature, in qualitative agreement with previous theoretical studies [37, 122]. This shows that the three-dimensional ordered structure of ice leads to a statistically significant increase in the molecular dipole moment. This observation is also consistent with discrete changes in the dielectric constant of water when changing from liquid to solid state and suggests, as argued by Vega and co-workers [10, 114, 123], that one should not expect a simple fixed-charge model to be able to describe dielectric properties of both ice and liquid water. As expected, the conformational fluctuations in the ice structure are much less pronounced than in the liquid, leading to a dipole distribution (Figure 9c) that

is much narrowed than those obtained for the liquid state and to a much lower statistical error (Table 7). Regarding the polarization energies, our calculations indicate that the distortion energy increases dramatically relative to the liquid, i.e. the water molecule wave function is significantly more disrupted when transferring from the gas to the solid than when transferring from the gas to the liquid state. Although the electronic contribution also increases in magnitude, the polarization correction is now dominated by the distortion energy, and is significantly positive. This is relevant information for those attempting to predict sublimation enthalpies and free energies from fixed-charge molecular models of water.

4. Conclusions

In this paper, we presented a new QM/MM approach for representing polarization in a liquid phase that satisfies the five key criteria described in the introduction: 1) the central molecule is polarized by the electric field of the surrounding molecules; 2) the surrounding molecules are polarized by the presence of their neighbors, including the central molecule; 3) the structure of the liquid, in terms of molecular arrangements, is correctly described; 4) an adequate statistical sample of liquid-phase fluctuations at the relevant temperature is carried out; 5) the electronic properties of the central molecule are decoupled from those of adjacent molecules. Our Self-Consistent Electrostatic Embedding (SCEE) approach is based on a relatively small number of computationally inexpensive electrostatic embedding calculations with different values of surrounding charges, so that the dipole of the central molecule can be determined self-consistently. We carried out a systematic test of methodological parameters and a thorough comparison with previous studies in the literature, leading to the following conclusions:

- The level of theory of the QM calculation on the central molecule must be able to accurately predict the gas-phase dipole moment and the polarizabilities, otherwise Criterion 1 is not satisfied. Most hybrid, double-hybrid and meta-GGA DFT functionals yield good results [100, 101], as do higher-level wave-function methods like MP2 and CC. However, lower-level methods like HF, LDA or GGA DFT functionals, as well as semi-empirical approaches, are much less reliable and underestimate the dipole moment shift from vapor to liquid. Similarly, the basis set needs to be large enough and include diffuse functions so that dipole moments are accurately predicted. However, within the subset of “good enough” functional/basis set combinations, the results for the dipole moment shift and polarization energies are

not very sensitive to that particular choice – for example, we estimate a maximum uncertainty of 0.05 D on the liquid dipole moment arising from the choice of QM method. Overall, we recommend the use of B3LYP/aug-cc-VTZ as a good compromise between accuracy, computational speed and general availability.

- A sufficiently large cluster of surrounding MM molecules must be considered to satisfy Criteria 1 and 3. We observed that our results converged beyond a cut-off radius of ~1.2 nm, in broad agreement with previous observations. This radius may need to be extended for larger molecules than water, and we intend to test this in the future. Adding a dielectric continuum beyond the MM cut-off sphere was found to yield only a minor increase in the liquid dipole moment of small clusters, at a dramatic increase in computational expense. We found this to be unnecessary, as using a large enough cluster of surrounding molecules effectively accounts for long-range effects on the polarization of the central molecule.
- Allowing the hydrogen atoms of the central molecule to relax was found to lead to a statistically significant increase in the liquid dipole moment and is therefore necessary to obtain accurate results, at least for the water models considered here. This may be related to the fact that the water configurations were obtained from MD simulations of rigid classical models. We intend to assess if this conclusion still applies for other molecules described through flexible potentials. It is important to remark that the vast majority of previous QM/MM calculations on water have not allowed for geometry relaxation; this at least partly explains why they have yielded liquid dipole moments that are systematically lower than those obtained from AIMD simulations, which explicitly allow for geometry relaxation.
- A sufficient number of liquid state configurations need to be sampled to obtain converged results. We found that Criterion 4 was satisfied if at least 150 statistically uncorrelated configurations were sampled, although we recommend using more configurations to increase statistical precision.
- The classical molecular model used to obtain liquid configurations needs to provide a realistic description of the local structure and thermodynamic properties (e.g., density) of liquid water. When this is not the case (e.g., TIP3P model), Criterion 3 is not obeyed and the liquid dipole moment is underestimated. Many previous QM/MM calculations used TIP3P and hence report lower dipole moments than estimated with

more accurate models. Our SCEE calculations yielded consistent results for three different water models (SPC/E, TIP4P2005 and TIP4P-FB), all of which provide a reasonable description of the liquid state. Therefore, it appears that the results are not very sensitive to the choice of model, as long as it is sufficiently realistic. Furthermore, statistically identical results were obtained when configurations were sampled from a polarizable model, where the dipole moments of surrounding MM molecules can take on different values, supporting the use of the much less computationally demanding fixed-charge models.

- Standard electrostatic embedding calculations, where the surrounding MM charges are held fixed at the same values as in the classical non-polarizable model, significantly underestimate the degree of polarization of the central molecule. It is thus necessary to allow the MM charges to respond self-consistently to the presence of surrounding molecules (Criterion 2), as done in our SCEE approach and in PE calculations.
- We observed that previous QM/MM calculations on liquid water suffered from at least one (often more) of the methodological limitations discussed above. When these are accounted for, the results of previous studies are consistent with our SCEE calculations.

Overall, we observed that when a water molecule is transferred from the gas to the liquid phase, its dipole moment increases by ~ 0.95 to ~ 1.00 D. This increase is consistent with previous estimates from AIMD and experiment, which are between 0.95 and 1.05 D. The two quadrupole moment components are also enhanced in the liquid phase. This has several implications for classical force field development. For instance, the significant enhancement of the liquid dipole moment supports previous assertions by ourselves [11, 13, 124] and others [8, 10, 43, 123] that classical fixed-charge models should not be expected to yield accurate predictions of the static dielectric constant of water, which depends on the dipole moment surface of the system. Instead, one needs to apply a correction for the fact that the dipole moment of fixed-charge models (normally between 2.2 and 2.4 D) is actually an effective value that is scaled down with respect to the real liquid phase dipole moment (of around 2.8 to 3.0 D) [8, 10, 11, 44]. This also means that point charges obtained from SCEE (or any other QM-based approach to describe the liquid state) should not be used directly in fixed-charge force fields. Instead, the charges that best describe the liquid-state PES for non-

polarizable models are more likely to lie somewhere between gas-phase and liquid-phase charges, as proposed in recent studies [12, 13, 44-46]. This suggests, therefore, that it may be possible to obtain a set of charges from SCEE that represent the real liquid dipole moment, and then scale them by a constant factor to provide “effective” charges for a fixed-charge force field. We are currently testing this possibility for molecules with different functional groups, and will report on those results in the near future.

Our SCEE approach, like other QM/MM methods, has the significant advantage over AIMD of easily satisfying Criterion 5. Taking advantage of this fact, it becomes straightforward to estimate different energetic contributions to the polarization process by carrying out single-point energy calculations under different conditions. Our results confirm that the energetic gain obtained from polarizing the molecule in the liquid phase (i.e., the stabilization energy) is approximately twice the magnitude of the energy lost due to distorting the molecular wave function (i.e., the distortion energy). The latter, however, is almost entirely compensated by the purely electronic component of polarization, i.e. the interaction energy between the polarized molecule and the fluctuating electron clouds of the surrounding liquid. Because those two terms (distortion and electronic) are missing in non-polarizable classical potentials, they can be added *post facto* to predictions of phase change energies (e.g., vaporization enthalpy and hydration free energies) [13]. In future work, we intend to estimate these energy contributions for several other molecules in the liquid state, to assess if the almost exact compensation observed for water is a general feature, or if, instead, it varies with the degree of polarity of the liquid. Furthermore, we aim to extend the current implementation of the SCEE method to deal with solutions, where solute and solvent have different identities, by adopting an approach along the lines reported by Georg et al. for solvation of acetone in water [125].

List of Acronyms:

AIMD – *Ab Initio* Molecular Dynamics

CC – Coupled Cluster

DFT – Density Functional Theory

EE – Electrostatic Embedding

GGA – Generalized Gradient Approximation

HF – Hartree-Fock

MD – Molecular Dynamics

MLWF – Maximally Localized Wannier Functions

MM – Molecular Mechanics

MP2 – Moller-Plesset 2nd order perturbation

PE – Polarized Embedding

PES – Potential Energy Surface

QM – Quantum Mechanics

SCEE – Self-Consistent Electrostatic Embedding

SCF – Self-Consistent Field

SCRf – Self-Consistent Reaction Field

Supplemental Material:

Scripts for processing MD simulations from the output of GROMACS and QM/MM calculations in Gaussian. Additional figures to support the discussion in the main paper. Table containing a detailed analysis of liquid water dipole moments reported in previous QM/MM studies.

Acknowledgements:

This work was developed within the scope of the project CICECO-Aveiro Institute of Materials, UIDB/50011/2020 & UIDP/50011/2020, financed by national funds through the FCT/MEC and when appropriate co-financed by FEDER under the PT2020 Partnership Agreement.

References

- [1] Born, M. “Volumen und Hydratationswärme der Ionen” *Eur. Phys. J. A* **1920**, *1*, 45–48.
- [2] Kirkwood, J. G. “The Dielectric Polarization of Polar Liquids” *J. Chem. Phys.* **1939**, *7*, 911–919.
- [3] Onsager, L. “Electric Moments of Molecules in Liquids” *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.
- [4] Böttcher, C. J. F. *Theory of Electric Polarization*. **1952**, Amsterdam, Houston, Elsevier Pub. Co.
- [5] Ball, P. “Water - an enduring mystery”, *Nature* **2008**, *452*, 291–292.
- [6] Errington, J. R.; Debenedetti, P. G. “Relationship between structural order and the anomalies of liquid water”, *Nature* **2001**, *409*, 318–321.

- [7] Russo, J.; Tanaka, H. “Understanding water’s anomalies with locally favoured structures”, *Nat. Commun.* **2014**, *5*, 3556.
- [8] Leontyev, I.; Stuchebrukhov, A. “Accounting for electronic polarization in non-polarizable force fields” *Phys. Chem. Chem. Phys.* **2011**, *13*, 2613-2626.
- [9] Milne, A. W.; Jorge, M. “Polarization Corrections and the Hydration Free Energy of Water” *J. Chem. Theory Comput.* **2019**, *15*, 1065–1078.
- [10] Vega, C. “Water: one molecule, two surfaces, one mistake” *Mol. Phys.* **2015**, *113*, 1145–1163.
- [11] Jorge, M.; Lue, L. “The Dielectric Constant: Reconciling Simulation and Experiment” *J. Chem. Phys.* **2019**, *150*, 084108.
- [12] Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. “Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization” *J. Phys. Chem. B* **2013**, *117*, 2328–2338.
- [13] Barrera, M. C.; Jorge, M. “A polarization-consistent model for alcohols to predict solvation free energies” *J. Chem. Inf. Model.* **2020**, *60*, 1352-1367. 10.1021/acs.jcim.9b01005
- [14] Coulson, C. A.; Eisenberg, D. “Interactions of H₂O molecules in ice I. The dipole moment of an H₂O molecule in ice” *Proc. R. Soc. London Ser. A* **1966**, *291*, 445-453.
- [15] Badyal, Y. S.; Saboungi, M.-L.; Price, D. L.; Shastri, S. D.; Haeffner, D. R.; Soper, A. K. “Electron distribution in water” *J. Chem. Phys.* **2000**, *112*, 9206-9208.
- [16] Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L. S. “Dipole moment of water from Stark measurements of H₂O, H₂¹⁸O, and D₂O” *J. Chem. Phys.* **1973**, *59*, 2254-2259.
- [17] Gubskaya, A. V.; Kuslik, V. G. “The total molecular dipole moment for liquid water” *J. Chem. Phys.* **2002**, *117*, 5290–5302.
- [18] Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. “The Water Dipole Moment in Water Clusters” *Science* **1997**, *275*, 814-817.
- [19] Car, R.; Parrinello, M. “Unified Approach for Molecular Dynamics and Density-Functional Theory” *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- [20] Gillan, M. J.; Alfè, D.; Michaelides, A. “Perspective: How good is DFT for water?” *J. Chem. Phys.* **2016**, *144*, 130901.
- [21] Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. “Ab initio liquid water” *J. Chem. Phys.* **1993**, *99*, 9080–9089.
- [22] Delle Site, L.; Alavi, A.; Lynden-Bell, R. M. “The electrostatic properties of water molecules in condensed phases: an ab initio study” *Mol. Phys.* **1999**, *96*, 1683–1693.

- [23] Silvestrelli, P. L.; Parrinello, M. "Water Molecule Dipole in the Gas and in the Liquid Phase" *Phys. Rev. Lett.* **1999**, *82*, 3308–3311.
- [24] Silvestrelli, P. L.; Parrinello, M. "Structural, electronic, and bonding properties of liquid water from first principles" *J. Chem. Phys.* **1999**, *111*, 3572.
- [25] Kuo, I.-F. W.; Mundy, C. J. "An ab Initio Molecular Dynamics Study of the Aqueous Liquid-Vapor Interface" *Science* **2004**, *303*, 658-660.
- [26] Allesch, M.; Schwegler, E.; Gygi, F.; Galli, G. "A first principles simulation of rigid water" *J. Chem. Phys.* **2004**, *120*, 5192-5198.
- [27] Becke, A. D. "Density-functional exchange-energy approximation with correct asymptotic-behavior" *Phys. Rev. A* **1998**, *38*, 3098-3100.
- [28] 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.
- [29] Perdew, J. P.; Burke, K.; Ernzerhof, M. "Generalized gradient approximation made simple" *Phys. Rev. Lett.* **1996**, *77*, 3865-68.
- [30] Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. "Molecular Dynamics Simulation of Liquid Water: Hybrid Density Functionals" *J. Phys. Chem. B* **2006**, *110*, 3685–3691.
- [31] Zhang, C.; Wu, J.; Galli, G.; Gygi, F. "Structural and Vibrational Properties of Liquid Water from van der Waals Density Functionals" *J. Chem. Theory Comput.* **2011**, *7*, 3054–3061.
- [32] DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. "The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water" *J. Chem. Phys.* **2014**, *141*, 084502.
- [33] Bankura, A.; Karnakar, A.; Carnevale, V.; Chandra, A.; Klein, M. L. "Structure, Dynamics, and Spectral Diffusion of Water from First-Principles Molecular Dynamics." *J. Phys. Chem. C* **2014**, *118*, 29401–29411.
- [34] Schienbein, P.; Marx, D. "Liquid-Vapor Phase Diagram of RPBE-D3 Water: Electronic Properties along the Coexistence Curve and in the Supercritical Phase." *J. Phys. Chem. B* **2018**, *122*, 3318–3329.
- [35] Adamo, C.; Barone, V. "Toward reliable density functional methods without adjustable parameters: The PBE0 model" *J. Chem. Phys.* **1999**, *110*, 6158-69.
- [36] Grimme, S. "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *J. Comp. Chem.* **2006**, *27*, 1787-99.

- [37] Batista, E. R.; Xantheas, S. S.; Jonsson, H. "Multipole moments of water molecules in clusters and ice Ih from first principles calculations" *J. Chem. Phys.* **1999**, *111*, 6011-6015.
- [38] Bader, R. F. W. *Atoms in Molecules* Oxford University Press, New York, **1994**, Vol. 22.
- [39] Marzari, N.; Vanderbilt, D. "Maximally localized generalized Wannier functions for composite energy bands" *Phys. Rev. B* **1997**, *56*, 12847-12869.
- [40] Dyer, P. J.; Cummings, P. T. "Hydrogen bonding and induced dipole moments in water: Predictions from the Gaussian charge polarizable model and Car-Parrinello molecular dynamics" *J. Chem. Phys.* **2006**, *125*, 144519.
- [41] Souza, I.; Martin, R. M.; Marzari, N.; Zhao, X.; Vanderbilt, D. "Wannier-function description of the electronic polarization and infrared absorption of high-pressure hydrogen" *Phys. Rev. B* **2000**, *62*, 15505-15522.
- [42] Guillot, B. "A reappraisal of what we have learnt during three decades of computer simulations on water" *J. Mol. Liq.* **2002**, *101*, 219-260.
- [43] Vega, C.; Abascal, J. L. F. "Simulating water with rigid nonpolarizable models: a general perspective" *Phys. Chem. Chem. Phys.* **2001**, *13*, 19663-19688.
- [44] Leontyev, I. V.; Stuchebrukhov, A. "Electronic Polarizability and the Effective Pair Potentials of Water" *J. Chem. Theory Comput.* **2010**, *6*, 3153-3161.
- [45] Karamertzanis, P. G.; Raiteri, P.; Galindo, A. "The Use of Anisotropic Potentials in Modeling Water and Free Energies of Hydration" *J. Chem. Theory Comput.* **2010**, *6*, 1590-1607.
- [46] Cole, D. J.; Vilseck, S. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. "Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning" *J. Chem. Theory Comput.* **2016**, *12*, 2312-2323.
- [47] Baker, C. M.; Besocke, R. B. "Matching of Additive and Polarizable Force Fields for Multiscale Condensed Phase Simulations" *J. Chem. Theory Comput.* **2013**, *9*, 2826-2837.
- [48] Lamoureux, G.; MacKerell, A. D.; Roux, B. "A simple polarizable model of water based on classical Drude oscillators" *J. Chem. Phys.* **2003**, *119*, 5185-5197.
- [49] Jedlovszky, P.; Richardi, J. "Comparison of different water models from ambient to supercritical conditions: A Monte Carlo simulation and molecular Ornstein-Zernike study" *J. Chem. Phys.* **1999**, *110*, 8019-8031.
- [50] Rick, S. W.; Stuart, S. J.; Berne, B. J. "Dynamical fluctuating charge force fields: Application to liquid water" *J. Chem. Phys.* **1994**, *101*, 6141-6156.
- [51] Senn, H. M.; Thiel, W. "QM/MM methods for biomolecular systems" *Angew. Chem. Int. Ed.* **2009**, *48*, 1198-1229.

- [52] Gao, J.; Xia, X. “A Priori Evaluation of Aqueous Polarization Effects Through Monte Carlo QM-MM Simulations” *Science* **1992**, *258*, 631-635.
- [53] Wei, D.; Salahub, D. R. “A combined density functional and molecular dynamics simulation of a quantum water molecule in aqueous solution” *Chem. Phys. Lett.* **1994**, *224*, 291-296.
- [54] Tuñón, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. “A Hybrid Density Functional-Classical Molecular Dynamics Simulation of a Water Molecule in Liquid Water” *J. Mol. Model.* **1995**, *1*, 196-201.
- [55] Tuñón, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F.; Rivail, J. L. “A Coupled Density Functional-Molecular Mechanics Monte Carlo Simulation Method: The Water Molecule in Liquid Water” *J. Comput. Chem.* **1996**, *17*, 19-29.
- [56] Chalmet, S.; Rinaldi, D.; Ruiz-López, M. F. “A QM/MM/Continuum Model for Computations in Solution: Comparison with QM/MM Molecular Dynamics Simulations” *Int. J. Quantum Chem.* **2001**, *84*, 559-564.
- [57] Chalmet, S.; Ruiz-López, M. F. “The reaction field of a water molecule in liquid water: Comparison of different quantum/classical models” *J. Chem. Phys.* **2001**, *115*, 5220-5227
- [58] Tu, Y.; Laaksonen, A. “On the effect of Lennard-Jones parameters on the quantum mechanical and molecular mechanical coupling in a hybrid molecular dynamics simulation of liquid water” *J. Chem. Phys.* **1999**, *111*, 7519-7525
- [59] Rocha, W. R.; Coutinho, K.; Almeida, W. B.; Canuto, S. “An efficient quantum mechanical/molecular mechanics Monte Carlo simulation of liquid water” *Chem. Phys. Lett.* **2001**, *335*, 127-133
- [60] Gubskaya, A. V.; Kusalik, P. G., “The multipole polarizabilities and hyperpolarizabilities of the water molecule in liquid state: an *ab initio* study” *Mol. Phys.* **2001**, *99*, 1107-1120.
- [61] Takahashi, H.; Hori, E.; Hashimoto, H.; Nitta, T. “A Hybrid QM/MM Method Employing Real Space Grids for QM Water in the TIP4P Water Solvents” *J. Comput. Chem.* **2001**, *22*, 1252-1261.
- [62] Coutinho, K.; Guedes, R. C.; Costa Cabral, B. J.; Canuto, S. “Electronic polarization of liquid water: converged Monte Carlo quantum mechanics results for the multipole moments” *Chem. Phys. Lett.* **2003**, *369*, 345–353.
- [63] Georg, H. C.; Canuto, S. “Electronic Properties of Water in Liquid Environment. A Sequential QM/MM Study Using the Free Energy Gradient Method” *J. Phys. Chem. B* **2012**, *116*, 11247-11254.

- [64] Jansen, G.; Colonna, F.; Ángyán, J. G. "Mixed Quantum-Classical Calculations on the Water Molecule in Liquid Phase: Influence of a Polarizable Environment on Electronic Properties" *Int. J. Quantum Chem.* **1996**, *58*, 251-265.
- [65] Moriarty, N. W.; Karlström, G., "Electronic Polarization of a Water Molecule in Water. A Combined Quantum Chemical and Statistical Mechanical Treatment" *J. Phys. Chem.* **1996**, *100*, 17791-17796.
- [66] Gao, J. "A molecular-orbital derived polarization potential for liquid water" *J. Chem. Phys.* **1998**, *109*, 2346-2354.
- [67] Field, M. J. "Hybrid quantum mechanical/molecular mechanical fluctuating charge models for condensed phase simulations" *Mol. Phys.* **1997**, *91*, 825-845.
- [68] Sánchez Mendoza, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. "A mean field approach that combines quantum mechanics and molecular dynamics simulation: the water molecule in liquid water" *J. Mol. Struct. (THEOCHEM)* **1998**, *426*, 181-190.
- [69] Sánchez, M. L.; Martín, M. E. Aguilar, M. A.; Olivares del Valle, F. J. "Solvent Effects by Means of Averaged Solvent Electrostatic Potentials: Coupled Method" *J. Comput. Chem.* **2000**, *21*, 705-715.
- [70] Martín, M. E. Aguilar, M. A.; Chalme, S.; Ruiz-López, M. F. "A comparative study of two QM/MM methods testing the validity of the mean field approximation" *Chem. Phys. Lett.* **2001**, *344*, 107-112.
- [71] Tu, Y.; Laaksonen, A. "The electronic properties of water molecules in water clusters and liquid water" *Chem. Phys. Lett.* **2000**, *329*, 283-288.
- [72] Poulsen, T. D.; Ogilby, P. K.; Mikkelsen, K. V. "Linear response properties for solvated molecules described by a combined multiconfigurational self-consistent-field/molecular mechanics model" *J. Chem. Phys.* **2002**, *116*, 3730-3738.
- [73] Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. "Dipole and quadrupole moments of liquid water calculated within the coupled cluster/molecular mechanics method" *Chem. Phys. Lett.* **2002**, *364*, 379-386.
- [74] Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. "Molecular electric properties of liquid water calculated using the combined coupled cluster/molecular mechanics method" *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 207-225.
- [75] Osted, A.; Kongsted, J.; Mikkelsen, K. V.; Christiansen, O. "A CC2 dielectric continuum model and a CC2 molecular mechanics model" *Mol. Phys.* **2003**, *101*, 2055-2071.
- [76] Jensen, L.; van Duijnen, P. T.; Snijders, J. G. "A discrete solvent reaction field model within density functional theory" *J. Chem. Phys.* **2003**, *118*, 514-521.

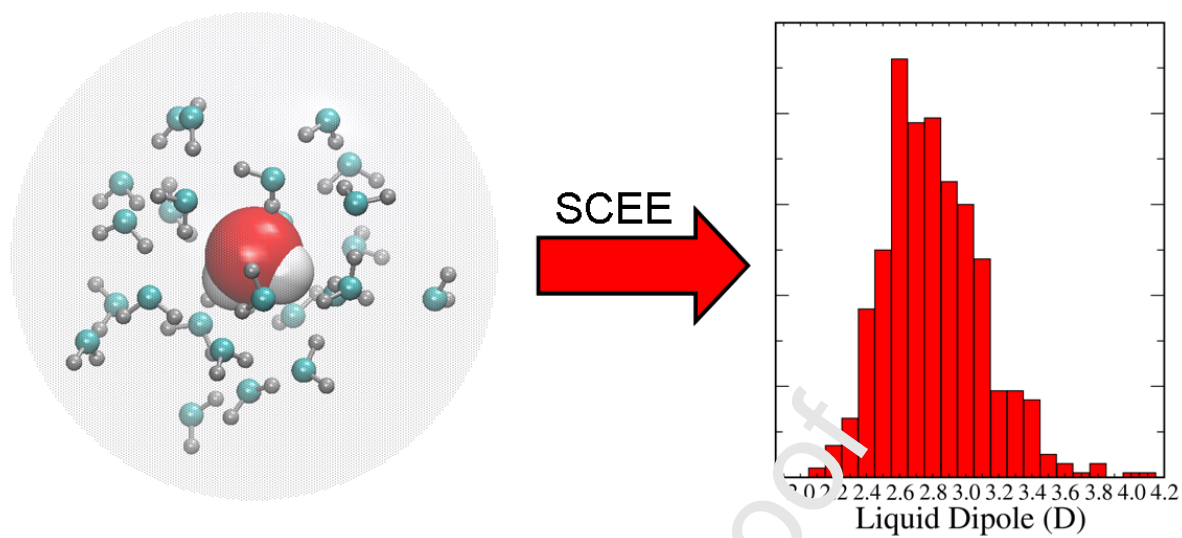
- [77] Osted, A.; Kongsted, J.; Mikkelsen, K. V.; Åstrand, P.-O.; Christiansen, O. “Statistical mechanically averaged molecular properties of liquid water calculated using the combined coupled cluster/molecular dynamics method” *J. Chem. Phys.* **2006**, *124*, 124503.
- [78] Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. “QM–MM interactions in simulations of liquid water using combined semi-empirical/classical Hamiltonians” *Phys. Chem. Chem. Phys.* **2008**, *10*, 297-302.
- [79] Millot, C.; Costa Cabral, B. J. “Electronic properties of liquid water by sequential molecular dynamics/density functional theory” *Chem. Phys. Lett.* **2008**, *460*, 466-469.
- [80] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. “Comparison of simple potential functions for simulating liquid water” *J. Chem. Phys.* **1983**, *79*, 926–935.
- [81] Berendsen, H. J.; Grigera, J. R.; Straatsma, T. P. “The missing term in effective pair potentials” *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- [82] Abascal, J. L. F.; Vega, C. “A general purpose model for the condensed phases of water: TIP4P/2005” *J. Chem. Phys.* **2005**, *123*, 234505
- [83] 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.
- [84] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. “GROMACS: A message-passing parallel molecular dynamics implementation” *Comput. Phys. Commun.* **1995**, *91*, 43-56.
- [85] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers” *SoftwareX* **2015**, *1*, 19-25.
- [86] Hockney, R.; Goel, S.; Eastwood, J. “Quiet high-resolution computer models of a plasma” *J. Comput. Phys.* **1974**, *114*, 148–158.
- [87] Nosé, S. “A unified formulation of the constant temperature molecular-dynamics methods” *J. Chem. Phys.* **1984**, *8*, 511-519.
- [88] Hoover, W. G. “Canonical dynamics: Equilibrium phase-space distributions” *Phys. Rev. A* **1985**, *31*, 1695-1697.
- [89] Parrinello, M.; Rahman, A. “Crystal Structure and Pair Potentials: A Molecular-Dynamics Study” *Phys. Rev. Lett.* **1980**, *45*, 1196–1199.
- [90] Miyamoto, S.; Kollman, P. A. Settle: “An analytical version of the SHAKE and RATTLE algorithm for rigid water models” *J. Comput. Chem.* **1992**, *13*, 952–962.

- [91] Darden, T.; York, D.; Pedersen, L. "Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems" *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [92] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. "Gaussian 09, Revision E.01" Gaussian, Inc., Wallingford CT, 2009.
- [93] Becke, A. D. "Density-functional thermochemistry. III. The role of exact exchange" *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [94] Yanai, T.; Tew, D.; Handy, N. "A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP)" *Chem. Phys. Lett.*, **2004**, *393*, 51-57.
- [95] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. "Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions" *J. Chem. Phys.* **2005**, *123*, 161103.
- [96] Zhao, Y.; Truhlar, D. G. "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals" *Theor. Chem. Acc.* **2008**, *120*, 215-241.
- [97] Chai, J.-D.; Head-Gordon, M. "Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections" *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615-6620.
- [98] Boese, A. D.; Martin, J. M. L. "Development of Density Functionals for Thermochemical Kinetics" *J. Chem. Phys.* **2004**, *121*, 3405-3416.
- [99] Grimme, S. "Semiempirical hybrid density functional with perturbative second-order correlation" *J. Chem. Phys.*, **2006**, *124*, 034108.

- [100] Hickey, A. L.; Rowley, C. N. "Benchmarking Quantum Chemical Methods for the Calculation of Molecular Dipole Moments and Polarizabilities" *J. Phys. Chem. A* **2014**, *118*, 3678–3687.
- [101] Hait, D.; Head-Gordon, M. "How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values" *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.
- [102] Krishnan, R., Binkley, J. S., Seeger, R., Pople, J. A. "Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions" *J. Chem. Phys.* **1980**, *72*, 650-654.
- [103] Dunning, T. H. "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen" *J. Chem. Phys.* **1989**, *90*, 1007-1023.
- [104] Tomasi, J.; Mennucci, B.; Cammi, R. "Quantum mechanical continuum solvation models" *Chem. Rev.* **2005**, *105*, 2999-3093.
- [105] Swope, W. C.; Horn, H. W.; Rice, J. E. "Accounting for polarization cost when using fixed charge force fields. I. Method for computing energy." *J. Phys. Chem. B* **2010**, *114*, 8621–8630.
- [106] Chipot, C. "Rational determination of charge distributions for free energy calculations." *J. Comput. Chem.* **2003**, *24*, 409–415.
- [107] Orozco, M.; Luque, F. J.; Habibollahzadeh, D.; Gao, J. "The polarization contribution to the free energy of hydration" *J. Chem. Phys.* **1995**, *102*, 6145-6152.
- [108] Luque, F. J.; Orozco, M. "Semiclassical-Continuum Approach to the Electrostatic Free Energy of Solvation" *J. Phys. Chem. B* **1997**, *101*, 5573-5582.
- [109] Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. "Solvent Effects 5. The Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation on ab initio Reaction Field Calculations" *J. Phys. Chem.*, **1996**, *100*, 16098-16104.
- [110] Besler, B. H.; Merz Jr., K. M.; Kollman, P. A. "Atomic charges derived from semiempirical methods" *J. Comp. Chem.* **1990**, *11*, 431-439.
- [111] Breneman, C. M.; Wiberg, K. B. "Determining atom-centered monopoles from molecular electrostatic potentials – the need for high sampling density in formamide conformational-analysis" *J. Comp. Chem.* **1990**, *11*, 361-373.
- [112] Kjellgren, E. R.; Olsen, J. M. H.; Kongsted, J. "Importance of Accurate Structures for Quantum Chemistry Embedding Methods: Which Strategy Is Better?" *J. Chem. Theory Comput.* **2018**, *14*, 4309–4319.

- [113] Reinholdt, P.; Kongsted, J.; Olsen, J. M. H. “Polarizable Density Embedding: A Solution to the Electron Spill-Out Problem in Multiscale Modeling” *J. Phys. Chem. Lett.* **2017**, *8*, 5949–5958.
- [114] Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragoes, J. L. “What ice can teach us about water interactions: A critical comparison of the performance of different water models” *Faraday Discuss.* **2009**, *141*, 251–276.
- [115] Lu, J.; Qiu, Y.; Baron, R.; Molinero, V. “Coarse-Graining of TIP4P/2005, TIP4P-Ew, SPC/E, and TIP3P to Monatomic Anisotropic Water Models Using Relative Entropy Minimization” *J. Chem. Theory Comput.* **2014**, *10*, 4104–4120.
- [116] McGrath, M. J.; Siepmann, J. I.; Kuo, I. W.; Mundy, C. J.; Vandevondele, J.; Hutter, J.; Mohamed, F.; Krack, M. “Simulating Fluid-Phase Equilibria of Water from First Principles” *J. Phys. Chem. A* **2006**, *110*, 640–646.
- [117] McGrath, M. J.; Siepmann, J. I.; Kuo, I. W.; Mundy, C. J. “Spatial correlation of dipole fluctuations in liquid water” *Mol. Phys.* **2007**, *105*, 1411–1417.
- [118] Niu, S.; Tan, M.-L.; Ichiye, T. “The large quadrupole of water molecules.” *J. Chem. Phys.* **2011**, *134*, 134501.
- [119] Hasted, J. B. “Liquid water: Dielectric properties”, in *Water: A comprehensive treatise*, **1972**, Vol 1, Ed. F. Franks, Plenum Press, New York, 255–309.
- [120] Matsumoto, M.; Yagasaki, T.; Tanaka, H. “GenIce: Hydrogen- Disordered Ice Generator” *J. Comput. Chem.* **2018**, *39*, 61–64.
- [121] Tu, Y.; Laaksonen, A. “Combined Hartree–Fock quantum mechanical and molecular mechanical molecular dynamics simulations of water at ambient and supercritical conditions” *J. Chem. Phys.* **2000**, *113*, 11264–11269.
- [122] Sharma, M.; Resta, R.; Car, R. “Dipolar Correlations and the Dielectric Permittivity of Water” *Phys. Rev. Lett.* **2007**, *98*, 247401.
- [123] Aragoes, J. L.; MacDowell, L. G.; Vega, C. “Dielectric Constant of Ices and Water: A Lesson about Water Interactions” *J. Phys. Chem. A* **2011**, *115*, 5745–5758.
- [124] Cardona, J.; Jorge, M.; Lue, L. “Simple corrections for the static dielectric constant of liquid mixtures from model force fields” *Phys. Chem. Chem. Phys.*, **2020**, in press.
- [125] Georg, H. C.; Coutinho, K.; Canuto, S. “Converged electronic polarization of acetone in liquid water and the role in the $n-\pi^*$ transition” *Chem. Phys. Lett.* **2006**, *429*, 119–123.

Graphical Abstract:



Author_Statement

Resources, Data Curation, Writing – Original Draft, Supervision, Project Administration; **José R. B. Gomes**: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – Review & Editing; **Andrew W. Milne**: Validation, Formal Analysis, Investigation, Writing – Review & Editing.

Journal Pre-proof

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Highlights:

- New method (SCEE) to quantify the extent of polarization in molecular liquids.
- Dipole moment of water is enhanced in the liquid phase by about 1.0 D.
- Distortion and electronic polarization contributions cancel out.
- SCEE can be used to estimate polarization corrections in non-polarizable force fields.

Journal Pre-proof