Paper IV

Averaged solvent embedding potential parameters for multiscale modeling of molecular properties

M. T. P. Beerepoot, A. H. Steindal, N. H. List,
J. Kongsted and J. M. H. Olsen
J. Chem. Theory Comp. **12** (2016), 1684–1695.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Averaged Solvent Embedding Potential Parameters for Multiscale Modeling of Molecular Properties

Maarten T. P. Beerepoot,*[,†] Arnfinn Hykkerud Steindal,[†] Nanna Holmgaard List,[‡,§] Jacob Kongsted,[‡] and Jógvan Magnus Haugaard Olsen[‡,¶]

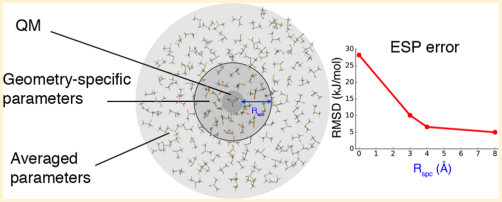
[†]Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø—The Arctic University of Norway, N-9037 Tromsø, Norway

[‡]Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark

[¶]Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: We derive and validate averaged solvent parameters for embedding potentials to be used in polarizable embedding quantum mechanics/molecular mechanics (QM/MM) molecular property calculations of solutes in organic solvents. The parameters are solvent-specific atom-centered partial charges and isotropic polarizabilities averaged over a large number of geometries of solvent molecules. The use of averaged parameters reduces the computational cost to obtain the embedding potential, which can otherwise be a rate-limiting step in calculations involving large environments. The parameters are evaluated by analyzing the quality of the resulting molecular electrostatic potentials with respect to full QM potentials. We show that a combination of geometry-specific parameters for solvent molecules close to the QM region and averaged parameters for solvent molecules further away allows for efficient polarizable embedding multiscale modeling without compromising the accuracy. The results are promising for the development of general embedding parameters for biomolecules, where the reduction in computational cost can be considerable.



1. INTRODUCTION

The calculation of molecular properties of large and complex chemical systems has long been a challenge for computational chemists. Over the last decades, the hybrid quantum mechanics/molecular mechanics (QM/MM) method has emerged as a popular tool for this purpose, because it combines the flexibility and accuracy of QM methods with the efficiency of MM methods.[1,2] MM methods lack the description of the electronic structure that is needed for the calculation of electronic properties, whereas accurate QM methods are too expensive to describe a large molecular system. An efficient way of including a homogeneous environment is through continuum models.[3] However, an atomistic description of the environment is needed to accurately describe heterogeneous surroundings and specific interactions between the quantum and classical parts of the molecular system. The combination of a QM method for a central subsystem with an atomistic classical description of the environment has the potential to overcome these difficulties. The coupling between the QM and MM methods can be described in different ways.[2] In mechanical embedding, the QM−MM interaction is treated classically. In electrostatic embedding, the permanent electrostatic potential (ESP) of the MM region is included as a one-electron operator in the QM Hamiltonian, thereby allowing the MM region to influence the electron density of the QM region but not the other way around. In polarizable embedding (PE), the environment can also be polarized by the QM region. In this work, we use the PE method by Olsen et al.[4,5] Polarization is described by induced dipoles that are determined self-consistently with respect to the electron density of the central subsystem, allowing for mutual polarization in both the electronic ground and excited states. All molecules in the environment of the central subsystem are defined by partial charges or multipoles—to describe the permanent charge distribution—and by isotropic or anisotropic dipole−dipole polarizabilities—to allow for the calculation of the induced dipoles representing polarization of the surroundings. The collection of parameters in the surroundings required by the PE calculations is referred to as the embedding potential.

An accurate approach to obtain the parameters in the embedding potential is to derive them from QM calculations on fragments making up the environment. For a solvent environment this is a straightforward task that requires one QM calculation for each solvent molecule. The drawback of such QM embedding potentials is the large amount of computational

Received: January 19, 2016
Published: March 3, 2016

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

resources used to generate the embedding parameters. Indeed, in some cases the generation of the embedding parameters can take more time than the calculation of the molecular property of the solute itself and becomes the rate-limiting step in the computational protocol. This is especially the case for calculations with a relatively small QM region and a large embedding region consisting of large molecular fragments such as acetone in a hexane solvent.[6] Another case where the computational effort of generating embedding potentials becomes very large is when molecular properties are averaged over snapshots from a molecular dynamics simulations, where a geometry-specific embedding potential is used for each snapshot.

A much simpler alternative to obtain the embedding parameters is to take them from a force field, such as AMBER,[7] OPLS,[8] or CHARMM.[9] The PE method requires at least polarizabilities to describe the effects of polarization both in the ground state and upon electronic excitation, so force fields based on fixed partial charges only cannot be used. Many different polarizable force fields exist,[10−13] but none of these has been developed specifically for the present purpose, i.e., molecular property calculations of solutes in solvents with solvent-specific isotropic parameters for an embedding method based on induced dipoles.

The motivation of this work comes from previous works in the literature.[6,14−19] Söderhjelm et al. have investigated several ways of constructing an embedding potential for the calculation of excitation energies in proteins.[14] In particular, they found that the use of anisotropic polarizabilities and charges, dipoles and quadrupoles—as opposed to isotropic polarizabilities and partial charges only—is important mainly for the part of the embedding region that is closest to the QM region. They suggested that the outer region of the embedding potential can be described in a less accurate way, for example by general parameters from a polarizable force field.[14] Söderhjelm, Aquilante, and Ryde have also shown that different approximations in the outer region of the embedding potential can still give accurate protein−ligand interaction energies.[15] Schwabe et al. have shown the importance of an accurate embedding potential close to the QM region[17] and how this reduces the need for a QM description of the closest surroundings of the embedded molecule.[17,18] Several works have shown that the deviation of an ESP calculated with embedding parameters from a QM ESP quickly decreases with the distance at which it is evaluated,[17,19] also indicating that the accuracy of the embedding potential is less important for the parts of the embedding region that are further away from the QM region. Söderhjelm, Kongsted, and Ryde have shown that averaged isotropic polarizabilities can be used as transferable parameters, as long as they are assigned for specific atoms rather than according to element or atom type.[16] We have previously shown that the computational effort of obtaining embedding potentials can be reduced by dividing the embedding region into different parts and discussed several ways of doing this.[6]

The goal of this work is to present and validate an approach for the generation of an embedding potential for solvent molecules that gives an accurate ESP for molecular property calculations while at the same time being efficient to compute. We present and validate the use of embedding potentials that contain atom-centered partial charges and isotropic polarizabilities that are obtained by averaging over parameters of 1000 different solvent geometries. Different schemes exist for deriving such charges and polarizabilities. Charges can for example be fitted to the ESP[20−22] or based on the atoms-in-molecule approach.[23] In this paper charges will be based on the former approach using the restrained electrostatic potential (RESP) method.[22] Polarizabilities are derived using the LoProp procedure.[24] Even though we use the PE method by Olsen et al.,[4,5] the derived averaged parameters can also be used in other charge- and isotropic polarizability-based embedding methods such as the MMpol method by Curutchet et al.[25] or the discrete reaction field method by Jensen, van Duijnen, and Snijders.[26] We focus in this work on embedding parameters for several commonly used solvents (see Figure 1), but the presented approach can also be applied to other solvents and (bio)-molecules.

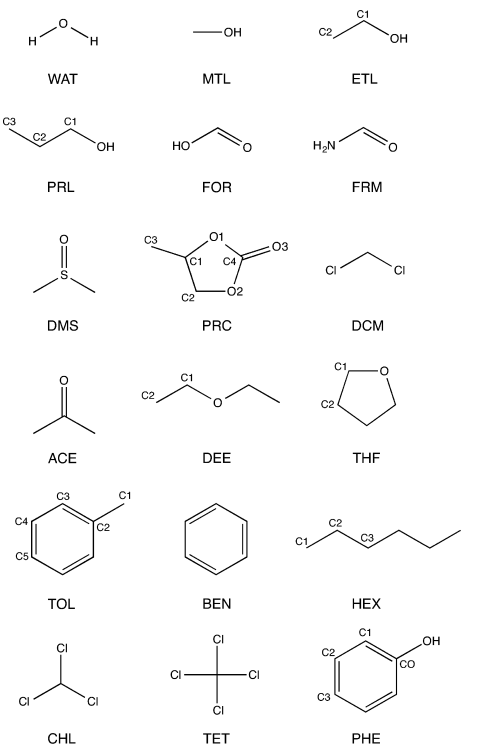


Figure 1. Chemical structures, atom names, and abbreviations for the solvent molecules used in this work: water (WAT), methanol (MTL), ethanol (ETL), 1-propanol (PRL), formic acid (FOR), formamide (FRM), dimethyl sulfoxide (DMS), propylene carbonate (PRC), dichloromethane (DCM), acetone (ACE), diethyl ether (DEE), tetrahydrofuran (THF), toluene (TOL), benzene (BEN), hexane (HEX), chloroform (CHL), tetrachloromethane (TET), and phenol (PHE).

The remainder of this work is structured as follows. In Section 2 we describe in detail the way in which we obtain our solvent-specific parameters and the way in which we validate them by calculating ESPs. A validation of our method is given in Section 3: the variation of the averaged parameters with structure and basis set is shown in Sections 3.1 and 3.2,

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

respectively, and the accuracy of the ESPs generated by the averaged parameters is evaluated in Section 3.3. In Section 3.4, we will show how a combination of geometry-specific embedding parameters for solvent molecules close to the QM region and our averaged parameters for the outer region gives the desired embedding potentials that are accurate yet efficient to generate. In Section 4 we summarize the main findings and discuss the applicability and limitations of the presented embedding parameters.

2. COMPUTATIONAL DETAILS

2.1. Generation of the Averaged Embedding Parameters. The partial charges and distributed polarizabilities for the solvent molecules shown in Figure 1 were calculated as averages over 1000 geometries from one molecular dynamics snapshot of a solvent box. Details of the generation of the solvent boxes are given in Section 7.1 in the Supporting Information.

Atom-centered RESP charges[22] and LoProp[24] isotropic polarizabilities were calculated with density-functional theory (DFT) for each of the 1000 solvent molecules in the solvent boxes. ESP charges were calculated using the Antechamber[27] module from AMBER[28] on an ESP obtained with Gaussian 09.[29] The molecular ESP was obtained using grid points from the Merz−Singh−Kollman scheme[20,21] with ten molecular surfaces defined by 1.4 to 2.7 times the vdW radii of the solvent atoms and a density of 17 grid points per square Ångström.

The RESP procedure introduces the constraint that chemically equivalent atoms get the same charge within the ESP-fitting procedure. Moreover, the RESP approach yields charges with a smaller dependence on the structure than other ESP-fitting schemes—especially for buried atoms—by restraining the charges toward zero.[22] This makes RESP charges especially useful in the development of transferable parameters, usually at the cost of a slightly worse fit with respect to the QM-derived ESP.[22] This is illustrated for ethanol in Table SI-II in the Supporting Information. In fact, the error of a geometry-specific RESP ESP with respect to a QM ESP (B3LYP/aug-cc-pVTZ) is large compared to other standard ESP fitting schemes for ethanol. However, the error from using averaged charges is lowest for the RESP scheme. Overall, the averaged RESP charges are more accurate than averaged charges from other ESP-fitting schemes when compared to the QM reference (Table SI-II).

Isotropic polarizabilities were calculated in Molcas[30,31] using the LoProp[24] approach with the basis set recontracted to an atomic natural orbital type basis. In all cases the charges and isotropic polarizabilities are calculated with the B3LYP exchange−correlation functional[32−35] and the same basis set, which is aug-cc-pVTZ[36] unless otherwise specified.

The embedding potentials consisting of RESP charges (Q) and isotropic polarizabilities (P1) will be referred to as QP1. The distinction is made between embedding potentials that were calculated for specific solvent geometries (geometry-specific QP1) and those using parameters that were averaged over 1000 solvent geometries (averaged QP1). Embedding potentials consisting of multipoles up to quadrupoles (M2) and anisotropic dipole−dipole polarizabilities (P2) were also calculated using the LoProp approach[24] in Molcas[30,31] for comparison to the QP1 embedding potentials. These embedding potentials were in all cases calculated for specific solvent geometries and are referred to as geometry-specific M2P2.

2.2. Molecular Electrostatic Potentials. Molecular ESPs were calculated as described in ref 19 to assess the accuracy of the embedding potentials from averaged parameters. Classical ESPs were calculated using the PE library[37] on a cubic grid with 5 grid points per bohr up to a distance of 12 bohr. Quantum-mechanical ESPs were calculated with B3LYP[32−35]/aug-cc-pVTZ[36] using the Dalton program[38,39] on the same grid. The quality of different ESPs was evaluated by calculating the root-mean-square deviation (RMSD) of the ESPs compared to a reference ESP on a molecular surface containing all points within (2.00 ± 0.01) times the van der Waals (vdW) radius for all atoms (unless otherwise stated) using the cubic grid specified above. The reference ESP was chosen as either the QM ESP—to obtain the error of using a classical description of the ESP—or the ESP generated by geometry-specific RESP charges—to obtain the error of the averaging procedure. Induced ESPs were calculated by applying an electric field of equal magnitude along the *x*-, *y*-, and *z*-axes defining the periodic box. Thus, the total electric field resulting from a 0.01 au field in the *x*-, *y*-, and *z*-directions is 0.017 au.

2.3. Property Calculations. The main goal of this work is to obtain accurate embedding potentials for molecular property calculations at a reduced computational effort. The quality of embedding potentials that are based on the averaged embedding parameters was tested by calculating the ESP and molecular properties of organic molecules in solvent environments. The ESP (here evaluated at the vdW surface of the solute molecule) is fundamental to obtain accurate molecular properties in a molecular environment. The evaluated molecular properties are dipole moments (as a measure for the ground-state electron distribution of the solvated molecule), oscillator strengths, vertical excitation energies, and two-photon transition strengths (as a measure for the change of the electron distribution upon excitation). Molecular properties were calculated for 50 snapshots of acetone in dimethyl sulfoxide and water from ref 6 and for one snapshot of *para*-nitroaniline (PNA) in water, dimethyl sulfoxide, propylene carbonate, and tetrachloromethane. Computational details of the structure generation are provided in Section 7 in the Supporting Information.

All molecular property calculations were performed with DFT using the CAM-B3LYP exchange−correlation functional[40] and the aug-cc-pVDZ basis set.[36] Using a long-range corrected exchange−correlation functional such as CAM-B3LYP (rather than B3LYP) for the calculation of the molecular properties is important to describe charge-transfer excitations such as the one in PNA that is considered in this work.[41] In addition, the ESP generated by the solvent embedding potential was evaluated on the vdW surface of the solute—which was described by CAM-B3LYP/aug-cc-pVDZ—to directly assess the influence of different embedding potentials on the ESP experienced by the solute. The PE QM/MM implementation[5] in Dalton2013[38,39] was used as provided through the PE library,[37] using Gen1Int[42] for the one-electron integrals. We emphasize that our interest in this study is not the molecular property itself but in how the different embedding potentials affect the molecular property.

3. RESULTS AND DISCUSSION

Our approach relies on using averaged atom-centered embedding parameters that are different for all chemically nonidentical atoms in a solvent molecule. At least two criteria should be satisfied for our approach to be successful: 1) the

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

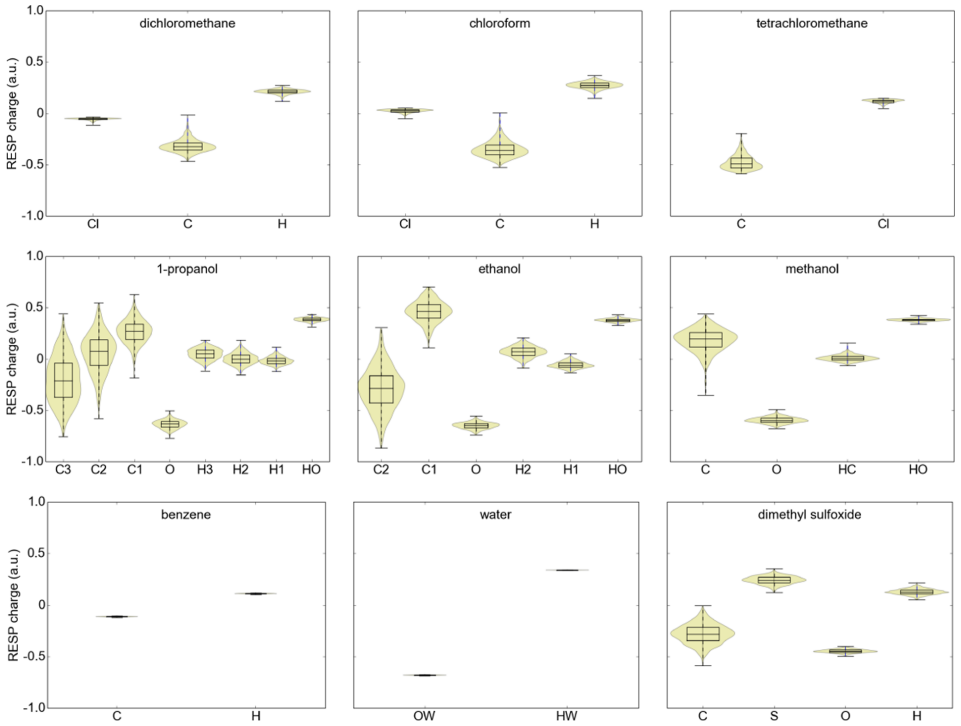


Figure 2. Variation in calculated RESP charges (B3LYP/aug-cc-pVTZ) for different solvent molecules. The diagrams show the variation in geometry-specific charges for 1000 different solvent geometries (see the Supporting Information for more solvents and tabulated data).

ESP of a solvent molecule should be sufficiently accurate when using the averaged embedding parameters and 2) molecular properties of solvated molecules should be accurate when using averaged parameters on all or part of the solvent molecules. We will start by examining the variation of the parameters with geometry (Section 3.1) and basis set (Section 3.2) and analyze the solvent ESP in Section 3.3. In Section 3.4, we will show how accurate embedding potentials can be made at reduced computational cost by combining the averaged parameters with more accurate geometry-specific parameters for the innermost molecules.

3.1. Structural Variation of the Parameters. The distribution of ESP-fitted partial charges and isotropic polarizabilities among 1000 geometries of different solvents is shown in Figures 2 and 3, respectively. The figures contain the maximum, minimum, lower quartile, median and upper quartile as horizontal bars and the distribution of all parameters as the shaded area. The averaged parameters for all solvents are tabulated in the Supporting Information (Table SI-I) together with figures showing the variation of the parameters for the remaining solvents (Figures SI-1 and SI-2).

The variation of the RESP charges differs considerably between atom types and solvents. Large variations are observed for buried carbon atoms, which is a known issue for ESP-fitting methods.[22] The atom type having the largest variation in RESP charges is the terminal carbon atom (C3) in 1-propanol, which varies by 1.21 au from −0.76 au to 0.45 au. In general, the alcohols (1-propanol, ethanol, methanol, phenol) show the largest dependence of the RESP charges on the molecular geometry. The atom types with the smallest variation in RESP charges—apart from the atoms in water, whose geometry was constrained in the simulation—are the carbon and hydrogen in benzene, varying by only 0.016 au in the 1000 geometries examined. The isotropic polarizabilities vary less with the structure (percentage-wise) with the largest variation being 2.6 au for the carbon in tetrachloromethane. Moreover, the variations between different atom types of the same element are small. For instance, the two different carbon atom types in ethanol (C1 and C2) have very similar polarizabilities, whereas the fitted charges are considerably different. Also, the polarizabilities of the terminal methyl groups have similar values in ethanol (C2) and 1-propanol (C3). This is in agreement with the conclusion of Gagliardi et al. that parameters calculated with the LoProp approach have a high degree of transferability between functional groups.[24]

Using averaged parameters for atoms that have a high structural variation (e.g., the fitted charges of the carbon atoms in the alcohols) might lead to less accurate ESPs for some of the solvent geometries (see Section 3.3). Unfortunately, our attempts to correlate structural parameters (e.g., O−C1−C2 angle in ethanol and O−C1−C2−C3 dihedral in 1-propanol) to the RESP charges using simple relationships have not been successful.

3.2. Variation with Basis Set. In order to investigate the basis set dependence of the averaged parameters, the parameters have been calculated with Dunning's aug-cc-

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

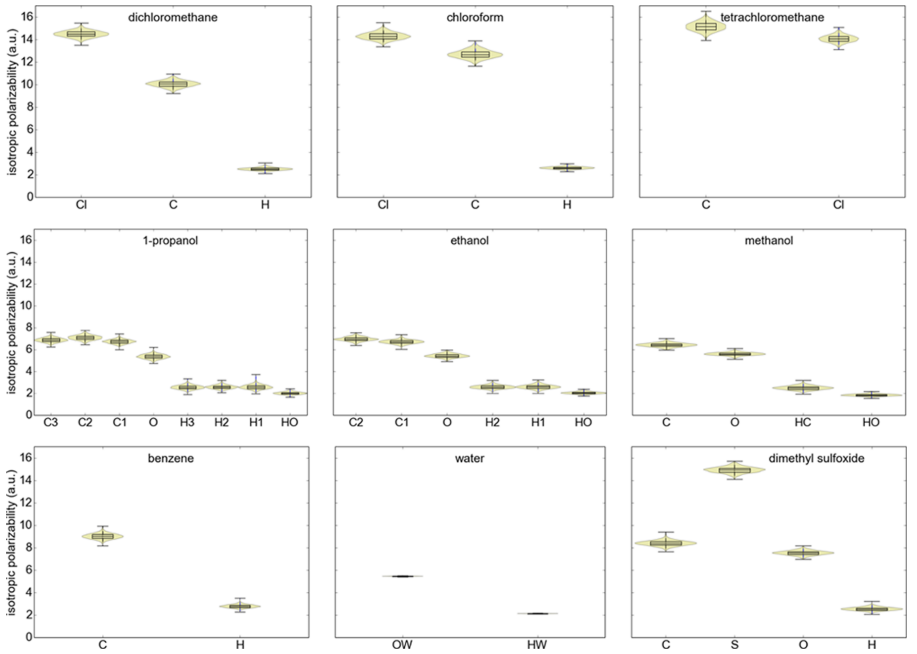


Figure 3. Variation in calculated isotropic polarizabilities (B3LYP/aug-cc-pVTZ) for different molecules. The diagrams show the variation in geometry-specific polarizabilities for 1000 different solvent geometries (see the Supporting Information for more solvents and tabulated data).

pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets[36] for 10 different solvent geometries. The deviations of the averaged values for the double-ζ and triple-ζ basis set results compared to the averaged quadruple-ζ basis set values are shown in Figure SI-3 in the Supporting Information. The averaged RESP charges computed with the different basis sets are rather similar to the largest deviations observed for the chloromethanes. Indeed, the difference between the aug-cc-pVTZ and aug-cc-pVQZ RESP charges (red in Figure SI-3) does not exceed 0.004 au with the largest deviation found for the carbon atom in tetrachloromethane. The largest deviation of the aug-cc-pVDZ basis set (black in Figure SI-3) is 0.059 au (C in tetrachloromethane) among the chloromethanes and only 0.013 au (C in dimethyl sulfoxide) among the other solvent molecules. We note that the variation of the averaged charges with basis set is an order of magnitude smaller than the variation in the parameters that results from different geometries (see Figure 2).

The difference in isotropic polarizabilities with basis set, in contrast, is of the same order of magnitude as the variation resulting from different geometries (Figure 3). The largest deviations are observed for the central carbon atoms in the chloromethanes (1.9 au for aug-cc-pVDZ and 1.1 au for aug-cc-pVTZ in tetrachloromethane, corresponding to relative errors of 15% and 7.5%, respectively) with *negative* deviations from the quadruple-ζ basis set, i.e., the more diffuse the basis, the larger the polarizability as expected. The deviations do not exceed 0.6 au in other solvent molecules. As for the RESP charges, the deviations from the quadruple-ζ basis set are larger for the double-ζ than for the triple-ζ basis set. Basis set convergence of the isotropic polarizabilities is reached satisfactorily at the B3LYP/aug-cc-pVTZ level, which was also concluded by Söderhjelm, Kongsted, and Ryde, who analyzed amino acids.[16]

It is here relevant to consider the error of using the much smaller aug-cc-pVDZ basis set, since this basis set also allows for calculations on much larger fragments, which are needed when constructing embedding potentials for biomolecules. In Table 1 we show the error—separately for RESP charges and LoProp isotropic polarizabilities—made when using aug-cc-

Table 1. Accuracy of the ESP from Averaged aug-cc-pVDZ Parameters Relative to the ESP from Averaged aug-cc-pVTZ Parameters[a]

	RMSD (kJ/mol)		
		α_{iso}	
parameter	q_{RESP}	0.0017[b]	0.017[b]
propanol	0.069	0.0064	0.064
ethanol	0.085	0.0082	0.082
methanol	0.091	0.013	0.13
chloroform	0.25	0.063	0.63
tetrachloromethane	0.27	0.064	0.64
dichloromethane	0.31	0.059	0.59
dimethyl sulfoxide	0.33	0.0086	0.086

[a]The ESPs have been calculated either with only averaged RESP charges (q_{RESP}) or with only averaged LoProp isotropic polarizabilities (α_{iso}) in the presence of an applied field of 0.001 or 0.01 au in each of the three Cartesian directions, resulting in a total electric field of 0.0017 and 0.017 au, respectively. The numbers are averages over 10 different geometries for each solvent. [b]Applied field (au).

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

pVDZ parameters for the solvent molecules having the largest variation between aug-cc-pVDZ and aug-cc-pVTZ parameters (Figure SI-3).

The errors made by using the smaller aug-cc-pVDZ basis set are small in all cases. The magnitude of the differences in the parameters (Figure SI-3) is reflected in the RMSD of the ESP (Table 1) with the largest changes for the chloromethanes and dimethyl sulfoxide. This indicates that the errors will be even smaller for the other solvent molecules. As for the quality of the polarizabilities, we emphasize that the absolute magnitude and consequently also relative error of the ESP increases with the applied field. Tests on the induced QM ESP of methanol (Figure SI-4) show that the response is approximately linear up to around 0.04 au. Similar tests have shown that nonlinearities may already arise at a total field strength of 0.01 au in some amino acids.[43] The classical response to an electric field in our current PE model is linear for all field strengths. We have previously calculated the average solvent electric field at the center of mass of the carbonyl group in solvated acetophenone.[44] The calculated fields were 0.013 au in water, 0.0053 au in dimethyl sulfoxide, 0.0024 au in chloroform, and much lower in nonpolar solvents. Since the magnitude of these fields are around or below 0.01 au the linear response approximation employed is reasonable. Comparing to the results in Table 1, we find that the error of using the aug-cc-pVDZ basis set instead of the aug-cc-pVTZ basis set is comparable for RESP charges and LoProp isotropic polarizabilities. Since the averaged parameters have to be calculated only once on relatively small molecules (up to 20 atoms), the aug-cc-pVTZ basis set rather than the smaller aug-cc-pVDZ basis set will be used in this work.

3.3. Molecular Electrostatic Potentials of the Solvent Molecules. Having obtained the averaged parameters and analyzed their variation with structure and basis set, we can now return to the first of the two criteria needed for our approach to be successful: the ESP of the solvent molecule should be accurately represented by the averaged parameters. The evaluation of molecular ESPs is a stringent test for the quality of embedding potentials of a molecule, because an accurate representation of the solvent ESP ensures correct incorporation of the electrostatic effects of the solvent in molecular property calculations. We note that the variation of the parameters with solvent geometry (as discussed in Section 3.1) is not a problem when the resulting ESP is reasonably reproduced. In Figure 4 we show ESP plots of randomly chosen geometries of four solvents. The plots show the difference in ESP calculated with geometry-specific RESP charges and averaged RESP charges on a molecular surface defined by spheres of twice the vdW radius on all atoms.

We see that the deviation from geometry-specific parameters is larger for 1-propanol and ethanol—which are flexible molecules—than for acetone and benzene—which are molecules with less structural variations. In order to obtain statistically meaningful results, we have calculated the RMSD of the ESP difference between embedding potentials from geometry-specific and averaged RESP charges for 1000 different geometries of the different solvents. The averaged numbers and standard deviations for each of the solvents are shown in Figure 5.

The RMSD is below 5 kJ/mol on average for all solvents. As expected, the solvents with the smallest structural variation (water, benzene, tetrachloromethane, see Figure 2) also show the lowest error compared to the ESPs calculated with geometry-specific parameters.

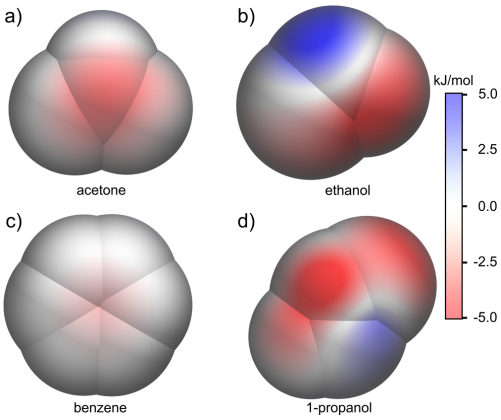


Figure 4. Molecular ESP plots for a geometry of a) acetone, b) ethanol, c) benzene, and d) 1-propanol. The ESPs are shown on a molecular surface defined by spheres of twice the vdW radius on all atoms. The figures show the difference between geometry-specific and averaged RESP ESPs. The ESPs are computed as interaction energies with a unit point charge. The darkest color in the ESP means that the error is the maximum/minimum on the scale or larger/smaller. The plots were created using VMD.[45]

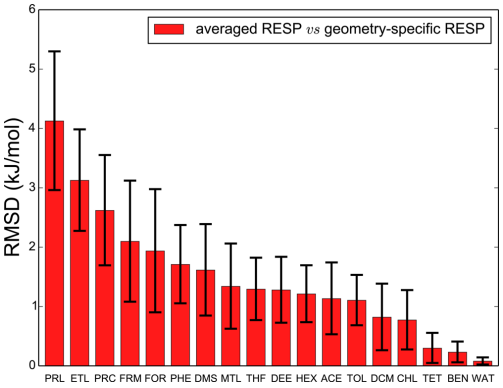


Figure 5. RMSD (in kJ/mol) of the ESP difference between the averaged and geometry-specific RESP ESPs on a molecular surface defined by spheres of twice the vdW radius on the atoms. See Figure 1 for the chemical structures and three-letter abbreviations of the solvent molecules. The numbers are averages over 1000 different geometries for each solvent with the standard deviation shown as error bars.

In Figure 6 the averaged (red) and geometry-specific (blue) ESPs based on RESP charges are compared to the full QM (B3LYP/aug-cc-pVTZ) ESP together with an embedding potential based on multipoles up to quadrupoles (M2; black).

We see that the magnitude of the error for the averaged RESP charges (red) is comparable to the error of the geometry-specific RESP charges (blue) with the QM potential as a reference, albeit being larger by up to 50% (for 1-propanol). In other words, the largest part of the error of the averaged solvent potentials does not come from the approximation of using

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

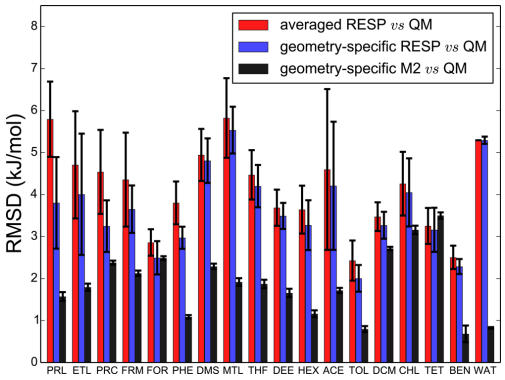


Figure 6. RMSD (in kJ/mol) of the ESP difference between the embedding potential obtained with averaged RESP parameters (red), with geometry-specific RESP parameters (blue), and with geometry-specific multipoles up to quadrupoles (M2) parameters (black) in comparison to the full QM potential (B3LYP/aug-cc-pVTZ). The ESPs are calculated on a molecular surface defined by spheres of twice the vdW radius on all atoms. See Figure 1 for the chemical structures and three-letter abbreviations of the solvent molecules. The numbers are averages over 10 different geometries for each solvent with the standard deviation shown as error bars.

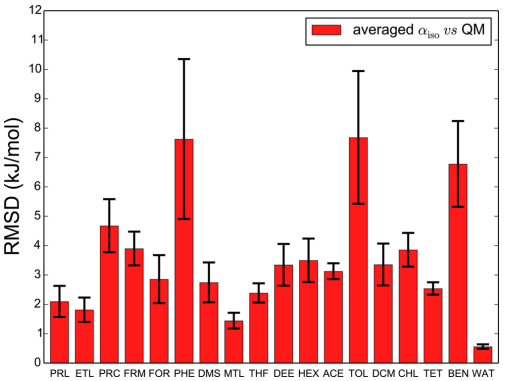


Figure 7. RMSD (in kJ/mol) of the induced ESP difference between the embedding potential obtained with averaged LoProp isotropic polarizabilities in comparison to the induced QM ESP (B3LYP/aug-cc-pVTZ) at an applied field of 0.01 au in all Cartesian directions (corresponding to a total field of 0.017 au). The induced ESPs are calculated on a molecular surface defined by spheres of twice the vdW radius on all atoms. See Figure 1 for the chemical structures and three-letter abbreviations of the solvent molecules. The numbers are averages over 10 different geometries for each solvent with the standard deviation shown as error bars.

averaged parameters, but from using RESP charges rather than more accurate potentials. We illustrate in Table SI-II in the Supporting Information for ethanol that this does not hold for other ESP-fitting schemes, which have a larger error for the averaging than for the fitting. We also note that the solvents with a low error due to the averaging are not necessarily the ones with a low error with respect to the QM reference. For instance, formic acid has one of the lowest errors with respect to the QM potential (Figure 6), but one of the largest as a result of the averaging (Figure 5). It is known that embedding potentials based on an expansion of multipole moments (M2) have a lower error compared to a QM reference than ESP-fitted charges and this error is well quantified in previous work.[5,17,19] The same is observed here (in black) for all solvents except for tetrachloromethane, where the dipoles and quadrupoles do not contribute due to the molecules's isotropy. Tetrachloromethane also has the highest error of M2 with respect to the QM ESP (3.5 kJ/mol). Averaging multipole moments, however, is more challenging than averaging isotropic parameters and is usually done only with fixed molecular geometries such as in refs 17 and 46.

In Figure 7 the quality of the isotropic polarizabilities is tested by comparing the induced ESP to a QM reference at an applied field of 0.01 au in all Cartesian directions, corresponding to a total electric field of 0.017 au. Comparison with a lower field strength (which is not shown here) reveals that this is in the linear region for all solvents considered, even though nonlinearities have been observed close to this field strength for larger molecules.[43] We emphasize again that the induced ESP and its error increase with the strength of the (applied) field.

The errors of the induced ESP are below 5 kJ/mol except for the aromatic solvents phenol, toluene, and benzene. A comparison of the error of the induced ESP (0.001 au in all Cartesian directions) of ten phenol geometries for the averaged isotropic polarizabilities (0.768 kJ/mol) with that of the geometry-specific isotropic polarizabilities (0.759 kJ/mol) and that of the geometry-specific anisotropic polarizabilities (0.125 kJ/mol) indicates that the error is not a result of the averaging procedure but of the use of isotropic parameters. Indeed, the isotropic polarizabilities of phenol, toluene, and benzene show only a very small variation with structure (Figures 3 and SI-2). It is likely that the anisotropy of the polarizability of the aromatic rings causes the larger error for these three solvents. We note that also the analysis of the electrostatic parameters (Figure 6) indicates that isotropic parameters (averaged or not) for these molecules introduce a much higher error than the anisotropic parameters compared to the QM reference. Combination of the error of the induced ESP in Figure 7 with the estimated fields in solvents mentioned in Section 3.2 reveals that it is unlikely that the error of the averaged isotropic polarizabilities exceeds the error of the averaged ESP-fitted charges in Figure 6.

The results in Figures 5−7 are based on ESPs on a molecular surface defined by spheres of twice the vdW radius on all atoms. This corresponds to a typical distance between a solute and a solvent molecule in the first solvation shell. Most solvent molecules, however, will be located further away from the central solute molecule. The variation of the RMSD of the ESP (defined here as the difference between averaged and geometry-specific RESP charges) is shown in Figure 8 as a function of the distance from the molecular surface at which it is evaluated. The numbers are averages over 10 geometries of 1-propanol, ethanol, dimethyl sulfoxide, and tetrachloromethane.

Figure 8 clearly illustrates that the error of the embedding potential with averaged parameters compared to the geometry-specific potential decreases sharply with the distance from the solvent molecule where the ESP is evaluated. The RMSD can be fitted to ax^{-n} with x the number of times the vdW distance and n equal to 1.9, 1.8, 2.0, and 3.3 for 1-propanol, ethanol, dimethyl sulfoxide, and tetrachloromethane ($R^2 \geq 0.997$),

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

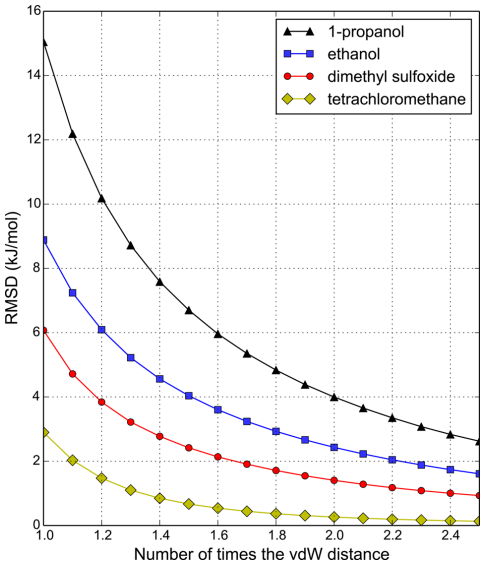


Figure 8. RMSD (in kJ/mol) of the ESP at different distances for the embedding potential with averaged RESP charges relative to the potential with geometry-specific RESP charges. The RMSDs (*y*-axis) are calculated at a molecular surface defined by spheres of the vdW atomic radius times a factor (*x*-axis). The numbers are averages over 10 different geometries for each solvent.

respectively. The decrease in the RMSD is partially due to the decrease of the absolute value of the ESP, which can be fitted to the same function with n equal to 1.6, 1.7, 1.5, and 3.2 ($R^2 \geq 0.996$), respectively, for the ESP resulting from the embedding potential with averaged RESP charges. Thus, the error in the ESP decreases slightly faster than the absolute value of the ESP when increasing the distance from the molecule.

A direct consequence of the results shown in Figure 8 is that the difference in quality of the embedding potential obtained with geometry-specific or averaged parameters is especially small for solvent molecules at a larger distance from the solute. Thus, for the solvent molecules closest to the central region of interest it is more important to use embedding parameters that are more accurate than the averaged parameters. It has previously been demonstrated that the RMSD from QM ESPs decreases almost as sharply for ESP-fitted charges as for a charge distribution with multipoles up to quadrupoles or octopoles.[5,17,19] The observation that the error in the ESP decreases sharply with the distance at which it is evaluated will be taken advantage of in the next section.

3.4. Molecular Properties from Polarizable Embedding Calculations. The second criterion determining the success of our approach is the accurate calculation of molecular properties using the averaged parameters on all or some solvent molecules. Since this involves more than one molecular fragment, polarization between the fragments is here also included through polarizabilities in the embedding potential. In section 3.3 we have shown that the ESP generated by a solvent molecule is more accurate with the multipole expansion up to quadrupoles (M2) than with the RESP charges (Figure 6), but that the error in the ESP decreases quickly with the distance where the ESP is evaluated (Figure 8). This implies that accurate embedding potentials can be generated at a reasonable cost by combining geometry-specific anisotropic parameters for the solvent molecules closest to the QM region (determined by a threshold R_{spc}) with averaged isotropic parameters for the rest of the solvent molecules up to the system size determined by a threshold R_{tot}. This approach is illustrated in Figure 9 and will be used in this section.

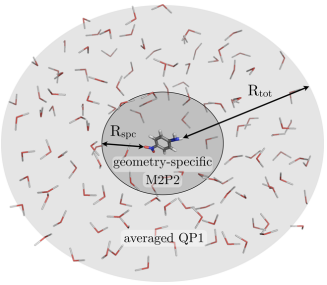


Figure 9. Schematic representation of the way the embedding potentials are built up in Section 3.4. R_{tot} is the system size threshold, R_{spc} is the threshold for geometry-specific parameters: if one of the atoms of a solvent molecule is within R_{spc} of one of the atoms of the solute, the embedding parameters are multipoles up to quadrupoles and anisotropic polarizabilities (M2P2) calculated specifically on each geometry. Beyond this threshold the embedding parameters are averaged RESP charges and isotropic polarizabilities (QP1).

We show in Figure 10 the effect of choosing different values for R_{spc} on the ESP on the molecular surface of acetone

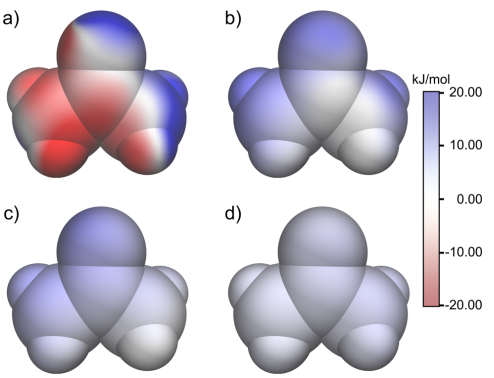


Figure 10. ESPs generated by the dimethyl sulfoxide solvent embedding potential on the vdW surface of the acetone solute. a) Difference between a geometry-specific M2P2 embedding potential ($R_{spc} = R_{tot} = 15$ Å) and the potential with averaged parameters ($R_{spc} = 0$ Å). b)-d) The geometry-specific M2P2 embedding potential is compared to potentials in which part of the solvent molecules are described by averaged parameters (see Figure 9): the solvent molecules beyond b) $R_{spc} = 3$ Å, c) $R_{spc} = 4$ Å, and d) $R_{spc} = 8$ Å. The ESPs are computed as interaction energies with a unit point charge. The darkest color on the scale means that the error is the maximum/minimum on the scale or larger/smaller. The plots were created using VMD.[45]

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

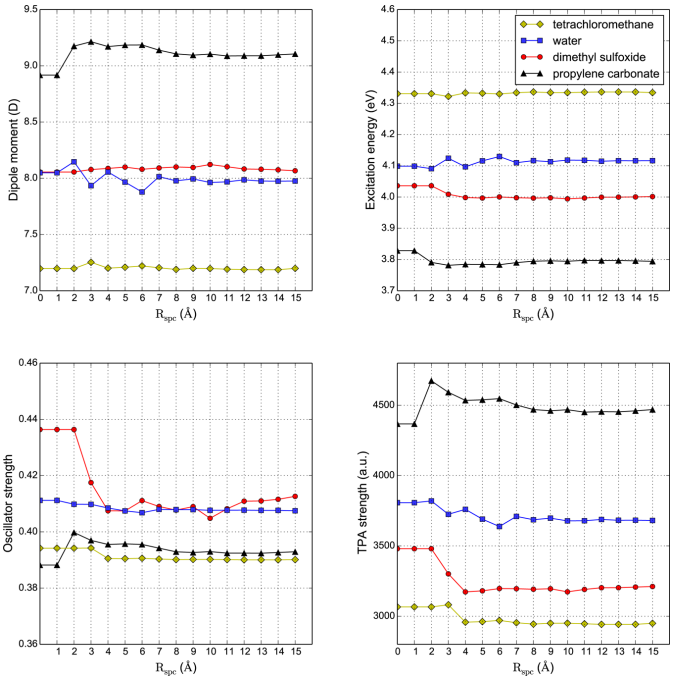


Figure 11. Dipole moment (top left) and excitation energy (top right), oscillator strength (bottom left) and two-photon absorption (TPA) transition strength (bottom right) of the charge-transfer excitation of a PNA molecule in water (blue), dimethyl sulfoxide (red), tetrachloromethane (green), and propylene carbonate (black) solvents. The embedding potential is made up of geometry-specific M2P2 parameters for solvent molecules with at least one atom within R_{spc} (x-axis) and averaged solvent embedding parameters beyond that threshold up to a system size of R_{tot} = 15 Å (see Figure 9).

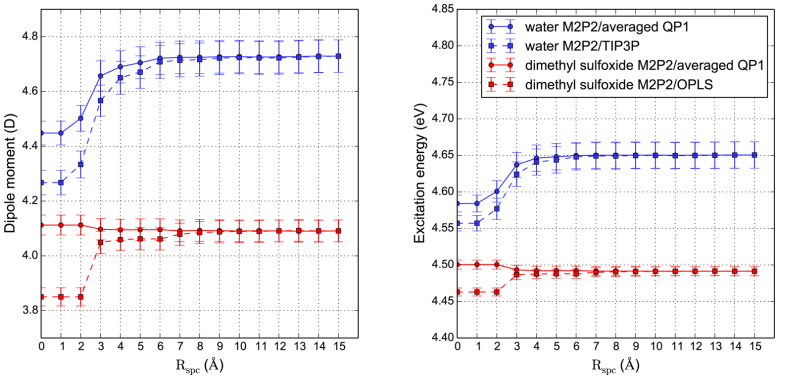


Figure 12. Dipole moment (left) and excitation energy (right) for the n → π^* excitation of acetone in water (blue) and dimethyl sulfoxide (red). The embedding potential is made up of geometry-specific M2P2 parameters for solvent molecules with at least one atom within R_{spc} (x-axis) and TIP3P/OPLS (squares) or averaged solvent embedding parameters (circles) beyond that threshold up to a system size of R_{tot} = 15 Å (see Figure 9). The numbers are averages over 50 snapshots from a molecular dynamics simulation with standard errors shown as error bars.

generated by a dimethyl sulfoxide solvent (R_{tot} = 15 Å). Note that the figure shows the ESP generated by all solvent molecules evaluated at the vdW surface of the QM region (in this case acetone), in contrast to the ESPs generated by one solvent molecule evaluated on its own surface in Section 3.3.

Figure 10 illustrates that the geometry-specific M2P2 ESP on the vdW surface of the QM region is reproduced well when only the innermost solvent molecules are described by geometry-specific M2P2 parameters and all other solvent molecules with averaged QP1 parameters. Indeed, the RMSD

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

of the ESP (evaluated on 2213 points at the vdW surface of acetone) decreases from 28.2 kJ/mol (R_{spc} = 0 Å) to 10.1 kJ/mol (R_{spc} = 3 Å), 6.6 kJ/mol (R_{spc} = 4 Å) and 5.0 kJ/mol (R_{spc} = 8 Å) when more solvent molecules are assigned geometry-specific M2P2 parameters.

The accuracy of the ESP generated by the solvent is crucial in determining the solvent effect on molecular properties of solute molecules. The convergence of molecular properties with the threshold R_{spc} is investigated systematically in Figure 11 for the charge-transfer excitation of PNA in different solvents. The molecular system consists of PNA and all solvent molecules with at least one atom within R_{tot} = 15 Å from the solute (see Figure SI-5 for the convergence of the molecular properties with R_{tot}).

The convergence of the molecular properties with R_{spc} is fastest for the least polar solvent, tetrachloromethane. In fact, using the averaged QP1 parameters on all solvent molecules (R_{spc} = 0 Å) gives almost converged properties for this snapshot. Note that tetrachloromethane is the only solvent where the solvent ESP is more accurate with averaged RESP charges than with M2 parameters (see Figure 6). For the three polar solvents, largest changes occur for the innermost molecules, i.e. when increasing R_{spc} from 0 to 4 Å. We emphasize that choosing R_{spc} = 4 Å and R_{tot} = 15 Å means that only a very small fraction of the solvent molecules need QM calculations: 15 out of 234 (6%) for dimethyl sulfoxide, 32 out of 765 (4%) for water, and 15 out of 215 (7%) for propylene carbonate. Thus, the approach yields accurate molecular properties at a greatly reduced computational cost.

Even though the best choice of R_{spc} will depend on the property investigated, the specific molecular system, and the desired accuracy, our data indicate that choosing R_{spc} to minimum 7 Å results in converged properties. *Averaging* molecular properties over snapshots can also affect the convergence of the properties. This is investigated in Figure 12 for 50 snapshots of acetone in water and dimethyl sulfoxide. In addition, the difference in convergence is investigated between embedding potentials combining geometry-specific M2P2 parameters with averaged QP1 and TIP3P/OPLS parameters, respectively.

Especially the convergence with system size R_{tot} is fast when averaging over more snapshots (Figure SI-6). Indeed, hardly any changes in the dipole moment and excitation energy are observed when the system size is increased beyond R_{tot} = 7 Å. Up to R_{spc} = 7 Å, the convergence of the properties is faster when the outer part is described by averaged QP1 parameters rather than by parameters (i.e., charges only) from TIP3P (water) or OPLS (dimethyl sulfoxide). Moreover, having *only* averaged parameters (thus avoiding any QM calculations to generate an embedding potential) leads to results closer to the geometry-specific M2P2 results for averaged QP1 parameters than for TIP3P/OPLS parameters.

4. CONCLUDING REMARKS

We have described how averaged solvent embedding parameters can be obtained and used to increase the computational efficiency of obtaining accurate embedding potentials for molecular property calculations. The solvent embedding parameters presented here are based on B3LYP/aug-cc-pVTZ calculations on 1000 different geometries. The error introduced by using averaged parameters is especially low for solvents such as benzene and tetrachloromethane, where the embedding parameters show only a small variation with structure. We have shown (Figure 6) that the largest error when using averaged RESP charges to reproduce the ESP of a solvent molecule does *not* come from the averaging procedure but from using fitted (isotropic) charges rather than an (anisotropic) multipole expansion. The use of isotropic polarizabilities is less successful on a few aromatic solvent molecules with high anisotropy (benzene, phenol, toluene). We estimate that the error of averaging isotropic polarizabilities is smaller than the error from averaging RESP charges (Section 3.3), in line with the smaller structural variation in the polarizabilities (Figure 3).

The deviation of the molecular ESP from the geometry-specific embedding potentials decreases sharply with the distance from the solvent molecules at which it is evaluated. Thus, for accurate molecular property calculations we advise the use of embedding potentials containing geometry-specific anisotropic parameters for the innermost molecules combined with averaged isotropic solvent-specific parameters for all other molecules. For calculations in which some of the accuracy can be sacrified to an easy or computationally efficient workflow, the averaged parameters can be used for all solvent molecules. Assigning averaged parameters to some or all parts of the embedding region leads to a considerable speedup in the calculation of the embedding potentials with the gain increasing with the number and size of molecular fragments in the classical region. This dramatic decrease in the computational effort needed to generate embedding potentials removes one of the disadvantages of accurate QM/MM models with respect to implicit solvent models, namely the computational resources spent on obtaining accurate embedding potentials.

The averaged embedding parameters obtained here depend on the geometry of the solvent molecules, which in turn depends on the way these geometries were obtained. However, we have shown that the quality of the ESPs is good enough also for geometries deviating from the average. In fact, as noted before, errors in the ESP resulting from using averaged parameters (Figure 5) are lower than those resulting from the use of RESP charges in comparison with QM ESPs for most of the solvents taken into consideration (Figure 6). Thus, the averaged parameters derived in this work can safely be used on solvent geometries that are obtained with another force field.

The approach to obtain averaged embedding parameters that has been presented and verified in this work can easily be extended to other solvents and even small biomolecules such as lipids. The computational effort to be gained from using averaged potential parameters is especially high for proteins and other biomolecules. Indeed, the molecular fragments used in obtaining the embedding parameters for proteins, i.e., capped amino acids, are of the same size or greater than the largest solvent molecules used in the present work. In such cases, the use of averaged parameters can lead to a dramatic decrease of computational costs in the calculation of embedding potentials for proteins. We have shown that the aug-cc-pVDZ basis set leads to only a small error in the permanent and induced ESP compared to the total error from using a classical potential and can thus safely be used to calculate ESP-fitted charges and LoProp polarizabilities. However, we recommend the use of the aug-cc-pVTZ basis set (as we have also done in this work) when possible to achieve converged ESP-fitted charges and LoProp isotropic polarizabilities. Even though our main goal is to calculate accurate embedding potentials in an efficient way to be used in molecular property calculations, the parameters can in principle also be used in molecular dynamics simulations as

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

long as compatibility with the nonelectrostatic interatomic interactions (i.e., Lennard−Jones parameters) is ensured for accurate overall nonbonded interactions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b01000.

Averaged embedding parameters for all solvents. Structural variation of calculated embedding parameters (cf. Figures 2 and 3). Comparison of averaged parameters calculated with different basis sets. RMSD of the ESP difference between embedding potentials of ethanol for different ESP-fitting schemes. RMSD of the induced ESP of methanol at different field strength. System size threshold test for PNA in different solvents and for acetone in water and dimethyl sulfoxide. Computational details for the generation of the molecular structures of the solvent boxes and of PNA and acetone in solvents (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 47 77623103. E-mail: maarten.beerepoot@uit.no.

Present Address

[§]Department of Physics, Chemistry and Biology, Linköping University, SE-58183 Linköping, Sweden.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.T.P.B. and A.H.S. acknowledge support from the Research Council of Norway through a Centre of Excellence Grant (Grant No. 179568/V30), from the European Research Council through a Starting Grant (Grant No. 279619), and from the Norwegian Supercomputer Program (Grant No. NN4654K). M.T.P.B. gratefully acknowledges financial support from BioStruct, the Norwegian national graduate school in structural biology. J.M.H.O. and J.K. acknowledge financial support from the Danish Council for Independent Research (DFF) through the Sapere Aude research career program. N.H.L. and J.K. acknowledge the Lundbeck Foundation for financial support. J.K. thanks the Villum foundation for financial support.

■ REFERENCES

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(2) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(3) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.
(4) Olsen, J. M. H.; Kongsted, J. *Adv. Quantum Chem.* **2011**, *61*, 107−143.
(5) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721−3734.
(6) Beerepoot, M. T. P.; Steindal, A. H.; Ruud, K.; Olsen, J. M. H.; Kongsted, J. *Comput. Theor. Chem.* **2014**, *1040−1041*, 304−311.
(7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(10) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.
(11) Wang, X.; Yan, T.; Ma, J. *Int. J. Quantum Chem.* **2015**, *115*, 545−549.
(12) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(13) Xu, P.; Wang, J.; Xu, Y.; Chu, H.; Liu, J.; Zhao, M.; Zhang, D.; Mao, Y.; Li, B.; Ding, Y.; Li, G. In *Advance in Structural Bioinformatics*; Wei, D. et al., Eds.; Springer: Dordrecht, 2015; Chapter 3, pp 19−32.
(14) Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. *J. Chem. Theory Comput.* **2009**, *5*, 649−658.
(15) Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085−11094.
(16) Söderhjelm, P.; Kongsted, J.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 1404−1414.
(17) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2011**, *7*, 2209−2217.
(18) Schwabe, T.; Beerepoot, M. T. P.; Olsen, J. M. H.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2582−2588.
(19) Olsen, J. M. H.; List, N. H.; Kristensen, K.; Kongsted, J. *J. Chem. Theory Comput.* **2015**, *11*, 1832−1842.
(20) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.
(21) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.
(22) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(23) Bader, R. W. F. *Atoms in Molecules − A Quantum Theory*; Oxford University Press: Oxford, 1990.
(24) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494−4500.
(25) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838−1848.
(26) Jensen, L.; van Duijnen, P. Th.; Snijders, J. G. *J. Chem. Phys.* **2003**, *118*, 514−521.
(27) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(28) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, 2010.
(29) 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 D.01*; Gaussian Inc.: Wallingford, CT, 2009.
(30) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222−239.

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

(31) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224−247.
(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(33) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(34) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(35) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(36) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(37) Olsen, J. M. H. *PElib: the Polarizable Embedding library, version 1.0.8*; 2014.
(38) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. A.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269−284.
(39) Dalton, a molecular electronic structure program, Release DALTON2013.0 (2013). See http://daltonprogram.org/ (accessed March 1, 2016).
(40) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(41) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. *J. Chem. Phys.* **2008**, *128*, 044118.
(42) Gao, B.; Thorvaldsen, A. J.; Ruud, K. *Int. J. Quantum Chem.* **2011**, *111*, 858−872.
(43) Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 1083−1090.
(44) List, N. H.; Beerepoot, M. T. P.; Olsen, J. M. H.; Gao, B.; Ruud, K.; Jensen, H. J. Aa.; Kongsted, J. *J. Chem. Phys.* **2015**, *142*, 034119.
(45) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(46) Schwabe, T. *J. Phys. Chem. B* **2015**, *119*, 10693−10700.

DOI: 10.1021/acs.jctc.5b01000
J. Chem. Theory Comput. 2016, 12, 1684−1695

Supporting Information to 'Averaged Solvent Embedding Potential Parameters for Multiscale Modeling of Molecular Properties'

Maarten T.P. Beerepoot,*,† Arnfinn Hykkerud Steindal,† Nanna Holmgaard List,‡,§ Jacob Kongsted,‡ and Jógvan Magnus Haugaard Olsen‡,¶

Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø—The Arctic University of Norway, N-9037 Tromsø, Norway, Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark, and Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

E-mail: maarten.beerepoot@uit.no

*To whom correspondence should be addressed
†University of Tromsø
‡University of Southern Denmark
¶École Polytechnique Fédérale de Lausanne
§Current address: Department of Physics, Chemistry and Biology, Linköping University, SE-58183, Linköping, Sweden

1 Averaged solvent embedding parameters

Table SI-I: Averaged solvent embedding parameters. The list gives the solvent name, atom name (see Figure 1 in main text for chemical structures, atom names and abbreviations), averaged RESP charge and averaged isotropic polarizability, both in atomic units. The numbers are averages over 1000 solvent configurations and have been calculated with B3LYP/aug-cc-pVTZ. See Section 2.1 in the main text for details.

```
WAT, OW, -0.67444, 5.73935
WAT, HW, 0.33722, 2.30839
MTL, C, 0.18382, 6.42738
MTL, O, -0.59543, 5.60297
MTL, HC, 0.01031, 2.50587
MTL, HO, 0.38068, 1.85100
ETL, C2, -0.29149, 6.96636
ETL, C1, 0.46155, 6.71482
ETL, O, -0.64779, 5.43599
ETL, H2, 0.07114, 2.58899
ETL, H1, -0.05680, 2.60105
ETL, HO, 0.37791, 2.04999
PRL, C3, -0.20048, 6.87397
PRL, C2, 0.06032, 7.09144
PRL, C1, 0.26720, 6.73554
PRL, O, -0.63447, 5.36537
PRL, H3, 0.04946, 2.56215
PRL, H2, 0.00238, 2.58205
PRL, H1, -0.01521, 2.59075
PRL, HO, 0.38471, 2.01746
FOR, HC, 0.08889, 2.52195
FOR, C, 0.53375, 7.42106
FOR, O, -0.49944, 5.69005
FOR, HO, 0.44069, 1.90088
FOR, OH, -0.56389, 5.45949
FRM, O, -0.51671, 6.20124
FRM, C, 0.52831, 8.10950
FRM, N, -0.72791, 6.66352
FRM, HC, 0.01015, 2.83651
FRM, HN, 0.35308, 2.26518
DMS, C, -0.28061, 8.40024
DMS, S, 0.24198, 14.92582
DMS, O, -0.44636, 7.54139
DMS, H, 0.12760, 2.55569
PRC, O1, -0.40892, 5.09781
PRC, C1, 0.38218, 6.26783
PRC, C2, 0.19825, 5.90126
PRC, O2, -0.37004, 5.21390
PRC, C3, -0.49200, 6.76455
PRC, C4, 0.80281, 7.82828
PRC, O3, -0.56716, 6.35162
PRC, H1, 0.01277, 2.52143
PRC, H2, 0.01185, 2.41556
PRC, H3, 0.13947, 2.46080
DCM, Cl, -0.05512, 14.48529
DCM, C, -0.31546, 10.06117
DCM, H, 0.21285, 2.51584
ACE, CO, 0.71719, 7.92284
ACE, CH, -0.45789, 7.19331
ACE, O, -0.53671, 5.83421
ACE, H, 0.12255, 2.50309
DEE, C2, -0.29619, 7.10478
DEE, H2, 0.07524, 2.58403
DEE, C1, 0.36805, 6.98709
DEE, H1, -0.04045, 2.62941
DEE, O, -0.43336, 5.82582
THF, C1, 0.22575, 6.42111
THF, H1, -0.00177, 2.63818
THF, C2, -0.04339, 6.73355
THF, H2, 0.01731, 2.56256
THF, O, -0.42688, 5.41568
TOL, C1, -0.48679, 7.31076
TOL, H1, 0.12857, 2.57127
TOL, C2, 0.34481, 9.41904
TOL, C3, -0.26155, 9.19246
TOL, H3, 0.13992, 2.71511
TOL, C4, -0.10270, 9.25031
TOL, H4, 0.12153, 2.83538
TOL, C5, -0.16049, 9.20304
TOL, H5, 0.12236, 2.88794
BEN, C, -0.11186, 9.02146
BEN, H, 0.11186, 2.79156
HEX, C1, -0.22011, 7.05133
HEX, C2, 0.14687, 7.17271
HEX, C3, -0.04616, 7.20426
HEX, H1, 0.04816, 2.63085
HEX, H2, -0.02291, 2.55841
HEX, H3, 0.01037, 2.52112
CHL, Cl, 0.02469, 14.29891
CHL, C, -0.34805, 12.68519
CHL, H, 0.27398, 2.61349
TET, C, -0.47432, 15.15766
TET, Cl, 0.11858, 14.07009
PHE, C1, -0.25722, 9.19780
PHE, H1, 0.15054, 2.75387
PHE, C2, -0.10722, 9.13020
PHE, H2, 0.12766, 2.77871
PHE, C3, -0.16229, 9.24398
PHE, H3, 0.12068, 2.88251
PHE, CO, 0.36980, 9.26636
PHE, O, -0.49794, 5.63527
PHE, HO, 0.34223, 1.93145
```

2 Structural variation of the parameters

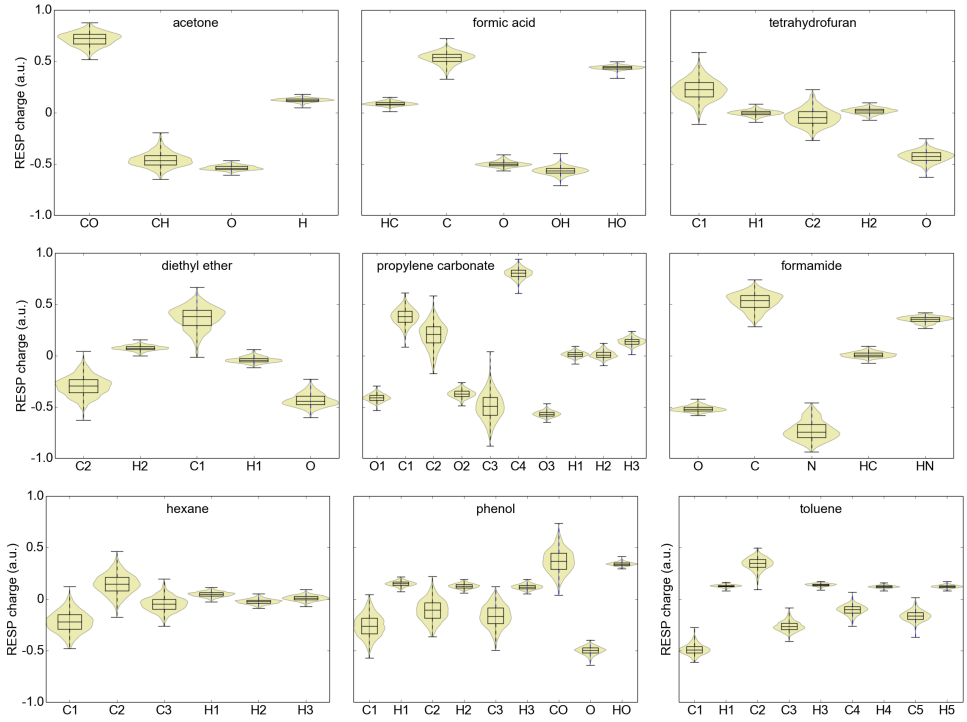


Figure SI-1: Variation in calculated RESP charges (B3LYP/aug-cc-pVTZ) for different solvent molecules. The diagrams show the variation in geometry-specific charges for 1000 different solvent geometries. Figure 2 in the main text contains the same diagrams for additional solvents.

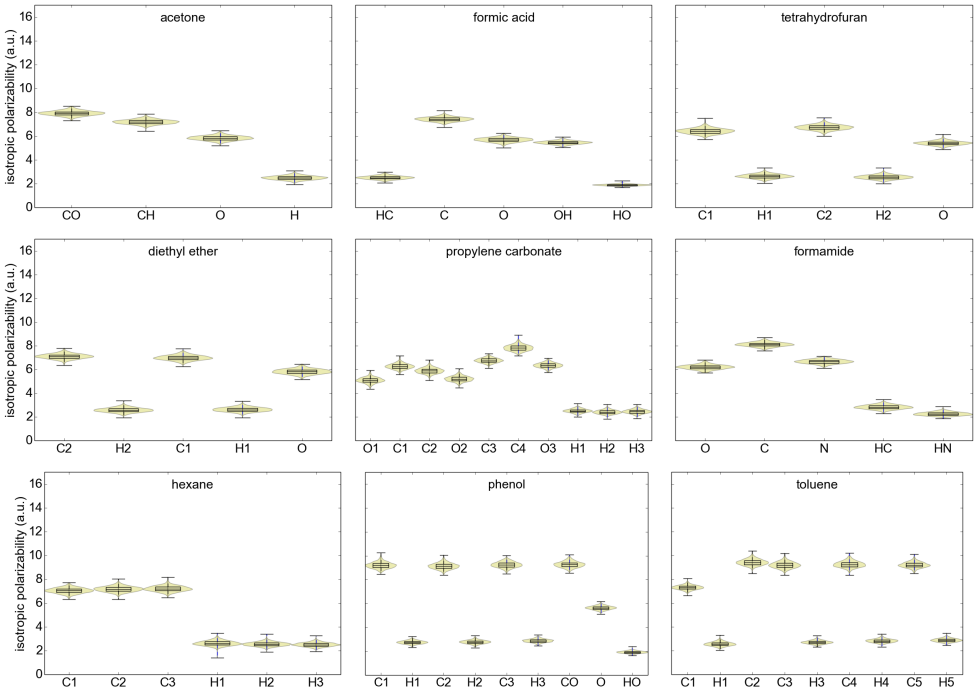


Figure SI-2: Variation in calculated LoProp isotropic polarizabilities (B3LYP/aug-cc-pVTZ) for different solvent molecules. The diagrams show the variation in geometry-specific polarizabilities for 1000 different solvent geometries. Figure 3 in the main text contains the same diagrams for additional solvents.

3 Variation with basis set

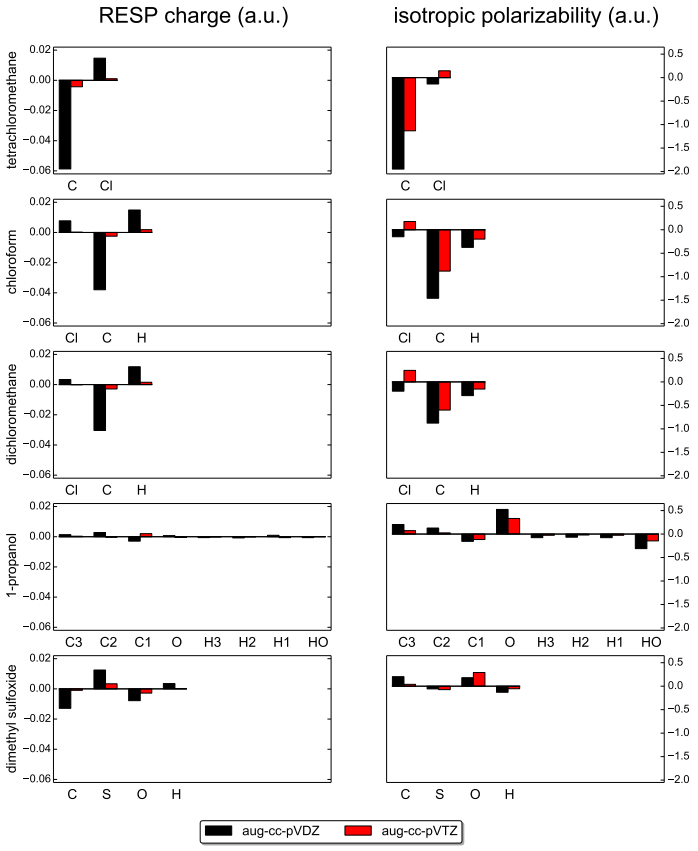


Figure SI-3: a) Basis set dependence of the RESP charges (left) and LoProp isotropic polarizabilities (right) for different solvent molecules. The figures show the deviation of the aug-cc-pVDZ (black) and aug-cc-pVTZ (red) basis sets relative to the aug-cc-pVQZ basis set. The numbers are averages over 10 different solvent geometries and are all calculated with the B3LYP functional.

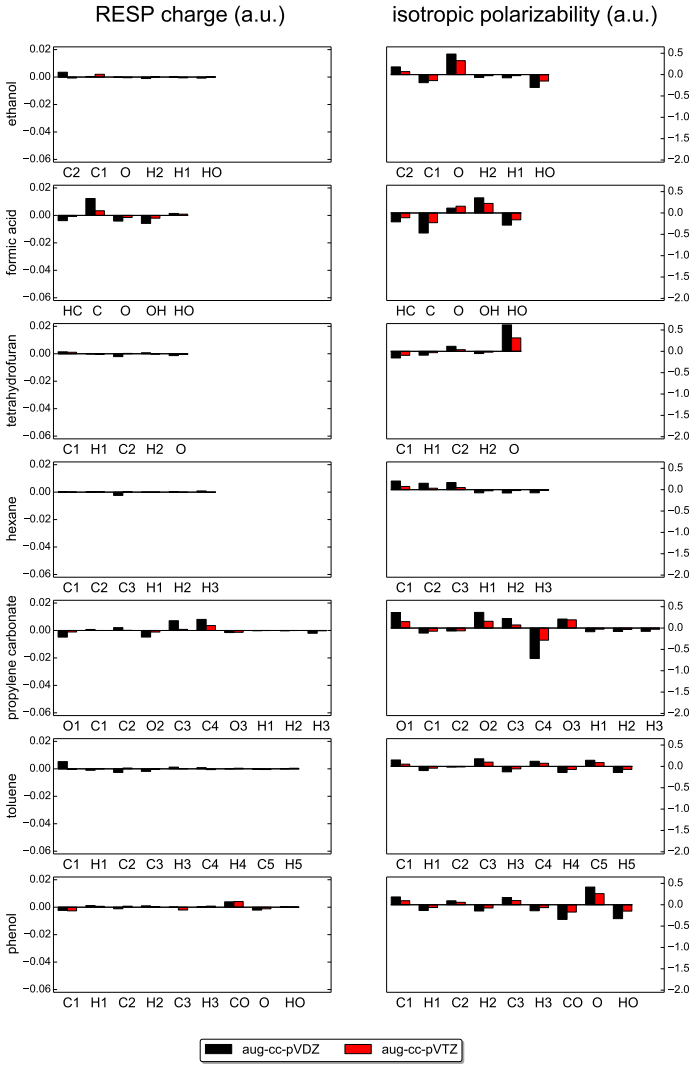


Figure SI-3: b) Basis set dependence of the RESP charges (left) and LoProp isotropic polarizabilities (right) for different solvent molecules. The figures show the deviation of the aug-cc-pVDZ (black) and aug-cc-pVTZ (red) basis sets relative to the aug-cc-pVQZ basis set. The numbers are averages over 10 different solvent geometries and are all calculated with the B3LYP functional.

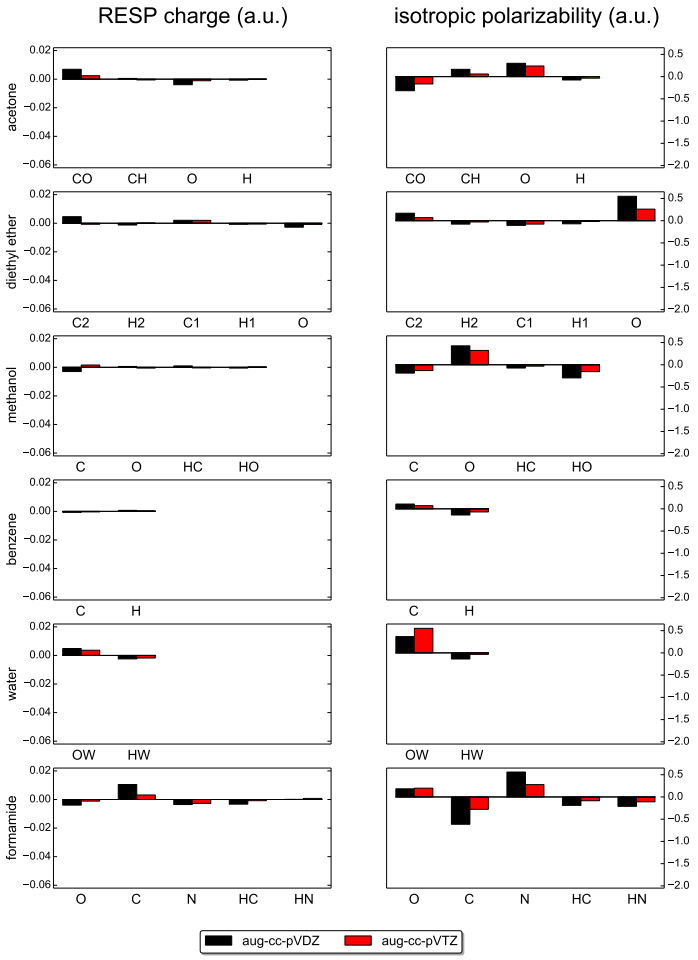


Figure SI-3: c) Basis set dependence of the RESP charges (left) and LoProp isotropic polarizabilities (right) for different solvent molecules. The figures show the deviation of the aug-cc-pVDZ (black) and aug-cc-pVTZ (red) basis sets relative to the aug-cc-pVQZ basis set. The numbers are averages over 10 different solvent geometries and are all calculated with the B3LYP functional.

4 Accuracy of different ESP fitting schemes

Table SI-II: RMSD (in kJ/mol) of the ESP between different embedding potentials for ethanol. The ESP is calculated with DFT (QM), with ESP-fitted charges calculated on the specific geometries (spc Q) and using averaged ESP-fitted charges (avg Q), where the ESP is in all cases calculated with B3LYP/aug-cc-pVTZ. The different ESP-fitting methods are RESP,[1] MK,[2,3] HLY[4] and CHelpG.[5] The ESP is calculated as an interaction energy with a unit point charge and evaluated on a molecular surface defined by spheres of twice the vdW radius on all atoms. The averaged ESP-fitted charges are obtained by averaging over all chemically equivalent atoms in 1000 geometries of ethanol. The RMSD is calculated as an average over 10 different geometries.

method	spc Q *vs* QM	avg Q *vs* QM	avg Q *vs* spc Q
RESP	4.01	4.71	2.43
MK	2.72	5.23	4.32
HLY	2.71	5.35	4.47
CHelpG	2.79	5.01	6.27

5 Linearity of response to applied electric field

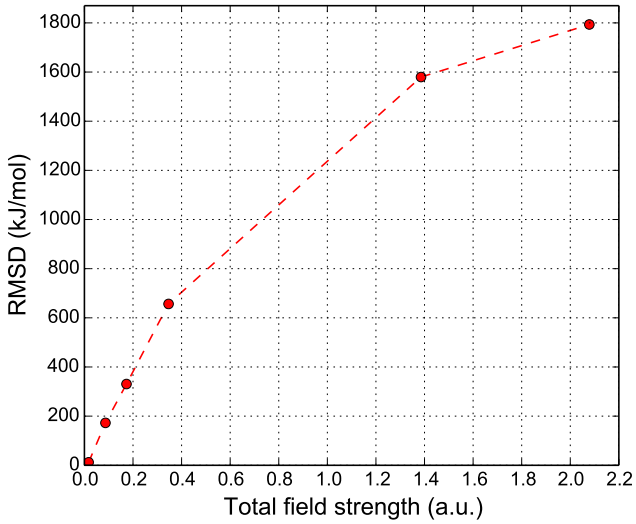


Figure SI-4: RMSD (in kJ/mol) of the induced QM ESP of methanol at different total field strengths with respect to the QM ESP without an applied electric field. The ESPs are calculated with B3LYP/aug-cc-pVTZ as an interaction energy with a unit point charge and evaluated on a molecular surface defined by spheres of twice the vdW radius on all atoms.

6 System size threshold tests

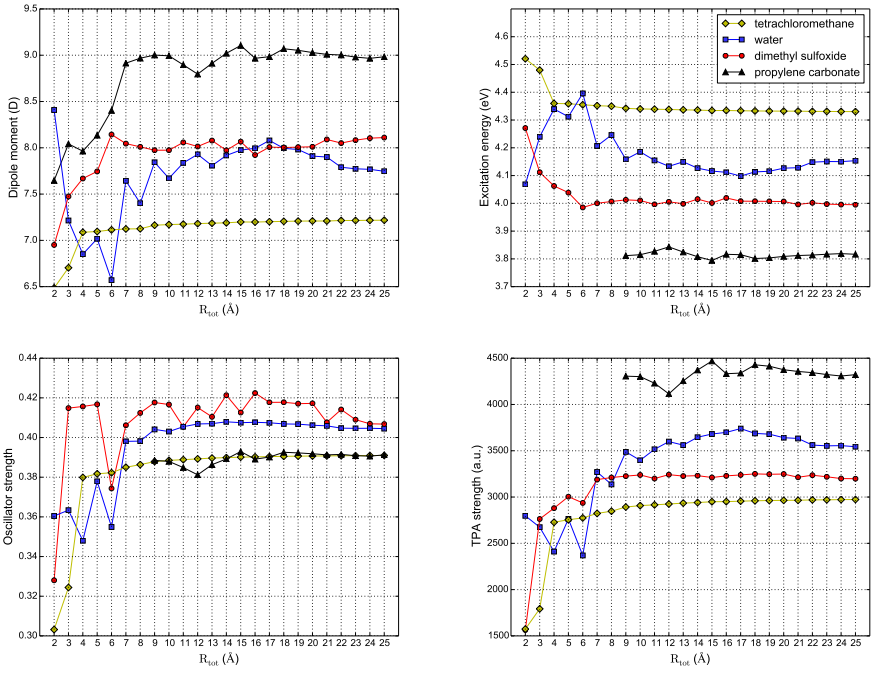


Figure SI-5: The dipole moment (top left), excitation energy (top right), oscillator strength (bottom left) and two-photon transition strength (bottom right) of the charge-transfer excitation of a *para*-nitroaniline (PNA) molecule in water (blue), dimethyl sulfoxide (red), tetrachloromethane (green) and propylene carbonate (black) solvents. The embedding potential is made up of an M2P2 potential for solvent molecules with at least one atom within the system size threshold R_{tot} (x-axis) and nothing beyond that threshold. The excitation energy, oscillator strength and two-photon transition strength of the charge-transfer excitation of PNA in propylene carbonate are not shown for system thresholds below 9 Å because the character of the excitations is not the same in the calculations on those systems.

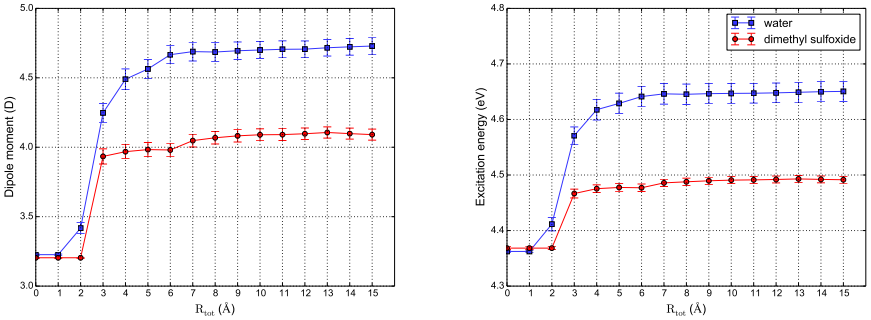


Figure SI-6: The dipole moment (left) and excitation energy (right) of the n→π^* excitation of acetone in water (blue) and dimethyl sulfoxide (red) solvents. The embedding potential is made up of an M2P2 potential for solvent molecules with at least one atom within the system size threshold R_{tot} (x-axis) and nothing beyond that threshold. The numbers are averages over 50 snapshots from a molecular dynamics simulation with standard errors shown as error bars.

7 Generation of the molecular structures

7.1 Solvent boxes

The atom-centered charges and isotropic polarizabilities for the solvent molecules shown in Figure 1 in the main article were calculated as averages over 1000 solvent geometries extracted from a single snapshot of a solvent box. The solvent boxes (containing exactly 1000 solvent molecules) were minimized and equilibrated in GROMACS[6] using three-dimensional periodic boundary conditions.

The optimized potential for liquid simulations (OPLS) force field[7] was used for all solvent molecules except for water. The TIP3P model[8] was used for water, which was constrained with the SETTLE algorithm[9] as is usually done. Topologies for solvent molecules were taken from the GROMACS molecule & liquid database[10,11] (www.virtualchemistry.org) where available. Nonbonded interactions were treated with a cutoff radius of 15 Å. Electrostatic interactions beyond this threshold were treated with the smooth particle-mesh Ewald method[12] with a tolerance of 10^{-5}.

The minimization of the solvent boxes consisted of 20 steps of steepest descent followed by 1000 steps of conjugate gradient. Initial velocities were obtained from a Maxwell distribution at 298 K. An NPT equilibration of 500 ps was run with the Berendsen[13] thermostat (298 K) and barostat (1 bar) to optimize the density of the solvent boxes. The relaxation constant for the Berendsen temperature and pressure coupling was set to 0.5 ps. Subsequently, the systems were equilibrated for 2 ns in the NVT ensemble using the Berendsen thermostat at 298 K. The time step used was 1 fs in both equilibration steps.

7.2 Snapshots of *para*-nitroaniline in water and dimethyl sulfoxide

Structures of *para*-nitroaniline (PNA) in a (60 Å)^3 solvent box (water, dimethyl sulfoxide, propylene carbonate and tetrachloromethane) were optimized and equilibrated using the same procedure as used for the solvent boxes, which is described in Section SI-7.1. For PNA the OPLS[7] topology of nitrobenzene from the Gromacs molecule & liquid database[10,11]

(www.virtualchemistry.org) was used with additional parameters for the amine group taken from the aniline topology from the same database. The force field used was deemed good enough to proceed without geometry optimization of the MD structures, which ensures that all temperature effects from the MD simulation are preserved. All solvent molecules with one or more atoms within R_{tot}=15 Å from one of the atoms of PNA were included in the embedding region for the QM/MM calculations based on convergence tests shown in Figure SI-7.

7.3 Snapshots of acetone in different solvents

The 50 molecular solute–solvent structures of acetone in various solvents (also using a threshold of R_{tot}=15 Å) are taken from a molecular dynamics simulation and were subsequently geometry optimized within the frozen solvent environment. Full details of the preparation of these structures are found in ref 14. See Figure SI-8 for the effect of the system size (R_{tot}) on the dipole moment and n→π^* excitation energy of acetone.

References

(1) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.

(3) Besler, B. H.; Merz Jr., K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.

(4) Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007**, *3*, 1004–1013.

(5) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.

(6) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(7) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(9) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.

(10) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.

(11) van der Spoel, D.; van Maaren, P. J.; Caleman, C. *Bioinformatics* **2012**, *28*, 752–753.

(12) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(13) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(14) Beerepoot, M. T. P.; Steindal, A. H.; Ruud, K.; Olsen, J. M. H.; Kongsted, J. *Comput. Theor. Chem.* **2014**, *1040–1041*, 304–311.