# 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

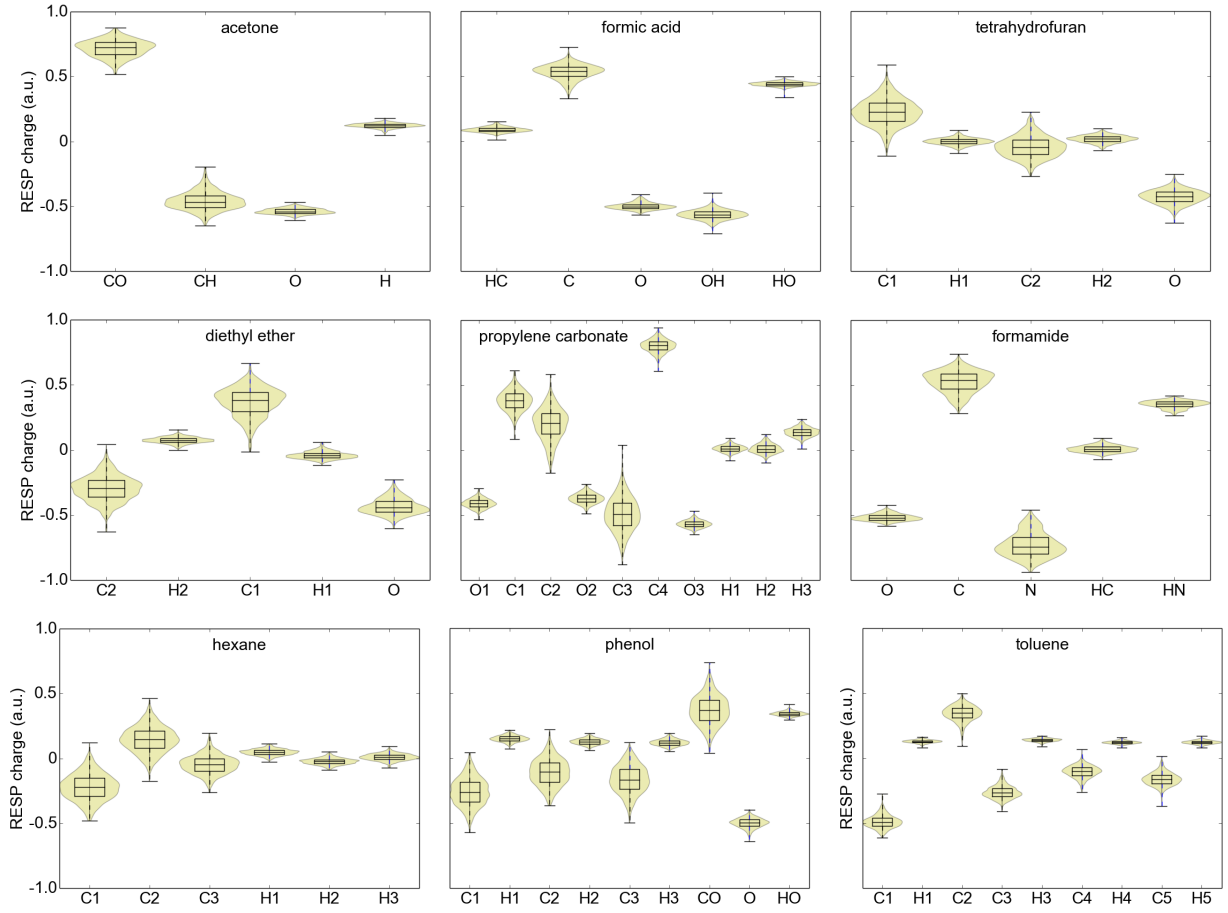


Figure SI-1: Variation in calculated RESP charges (B3LYP/aug-cc-pVTZ) for different solvent molecules. The diagrams show the variation in geometry-specific charges for 1000 different solvent geometries. Figure 2 in the main text contains the same diagrams for additional solvents.

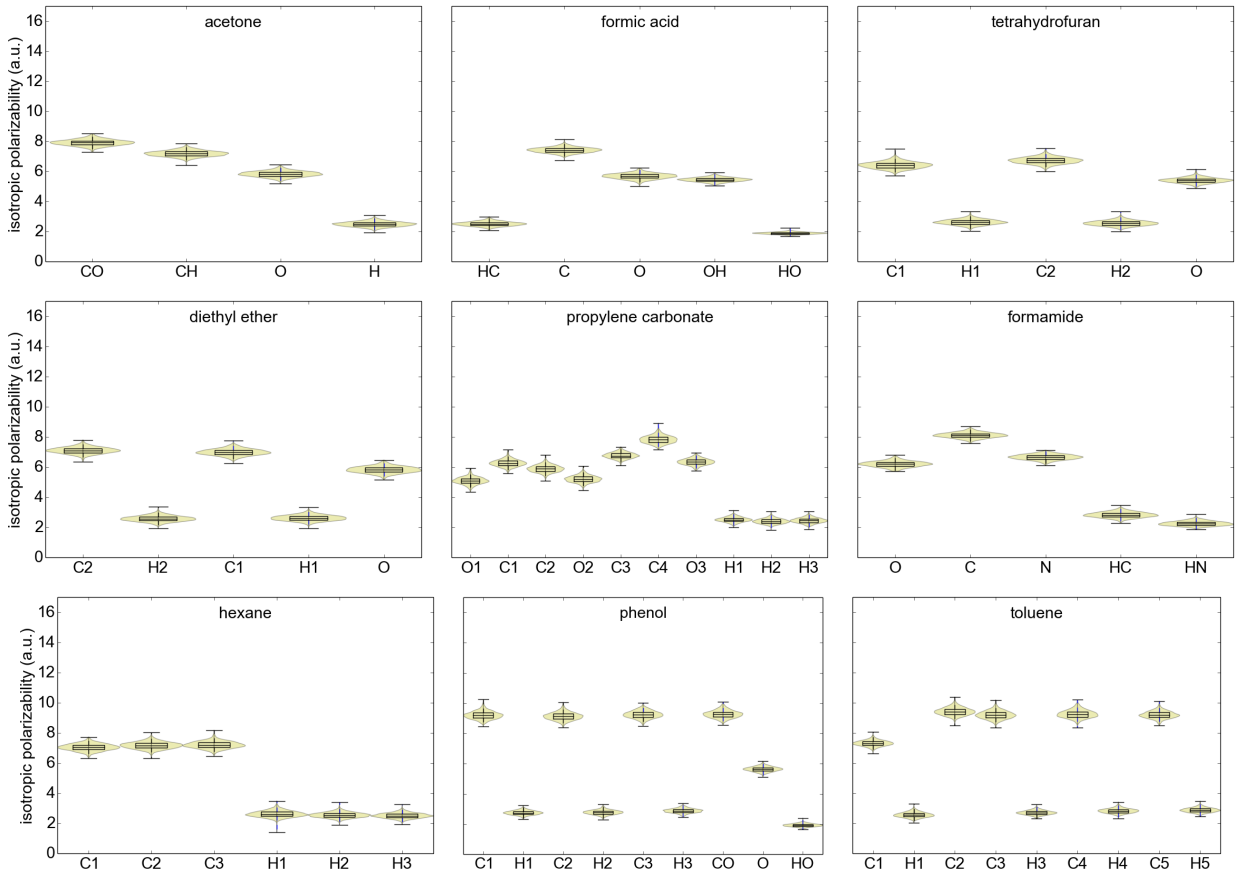


Figure SI-2: Variation in calculated LoProp isotropic polarizabilities (B3LYP/aug-cc-pVTZ) for different solvent molecules. The diagrams show the variation in geometry-specific polarizabilities for 1000 different solvent geometries. Figure 3 in the main text contains the same diagrams for additional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

# 3 Variation with basis set

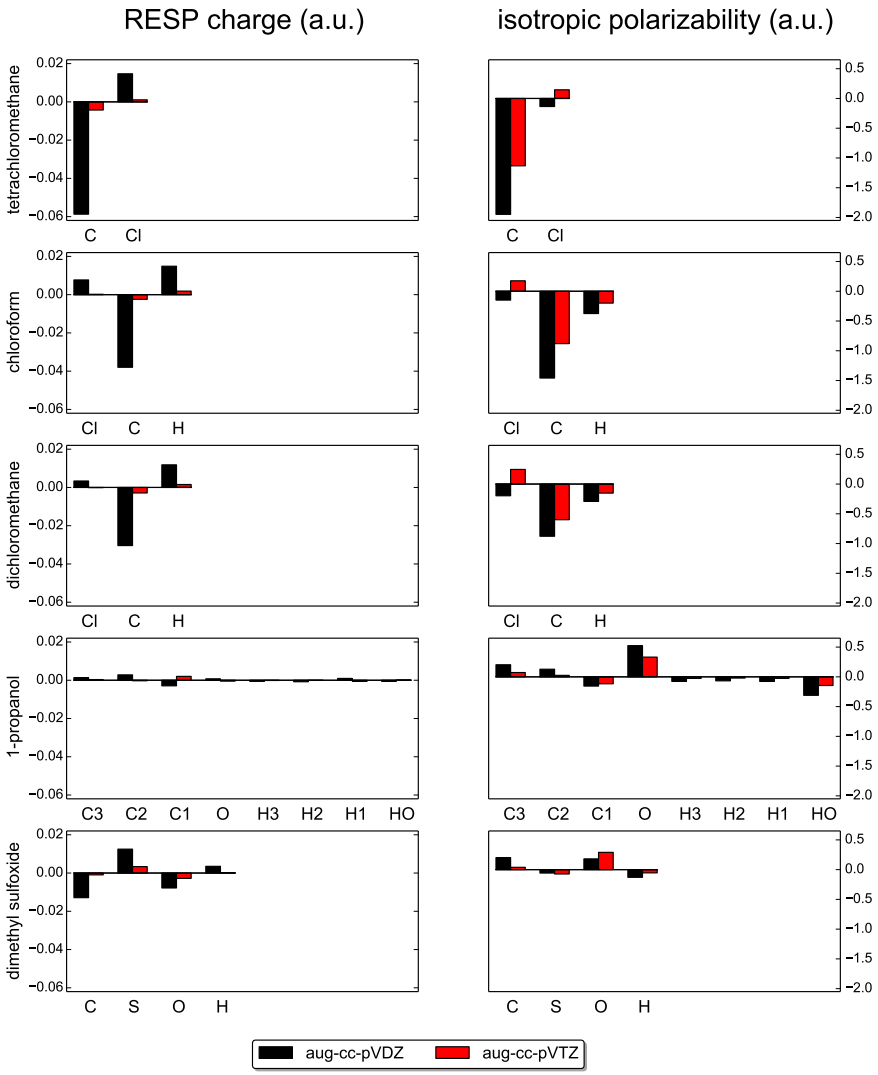


Figure SI-3: a) Basis set dependence of the RESP charges (left) and LoProp isotropic polarizabilities (right) for different solvent molecules. The figures show the deviation of the aug-cc-pVDZ (black) and aug-cc-pVTZ (red) basis sets relative to the aug-cc-pVQZ basis set. The numbers are averages over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

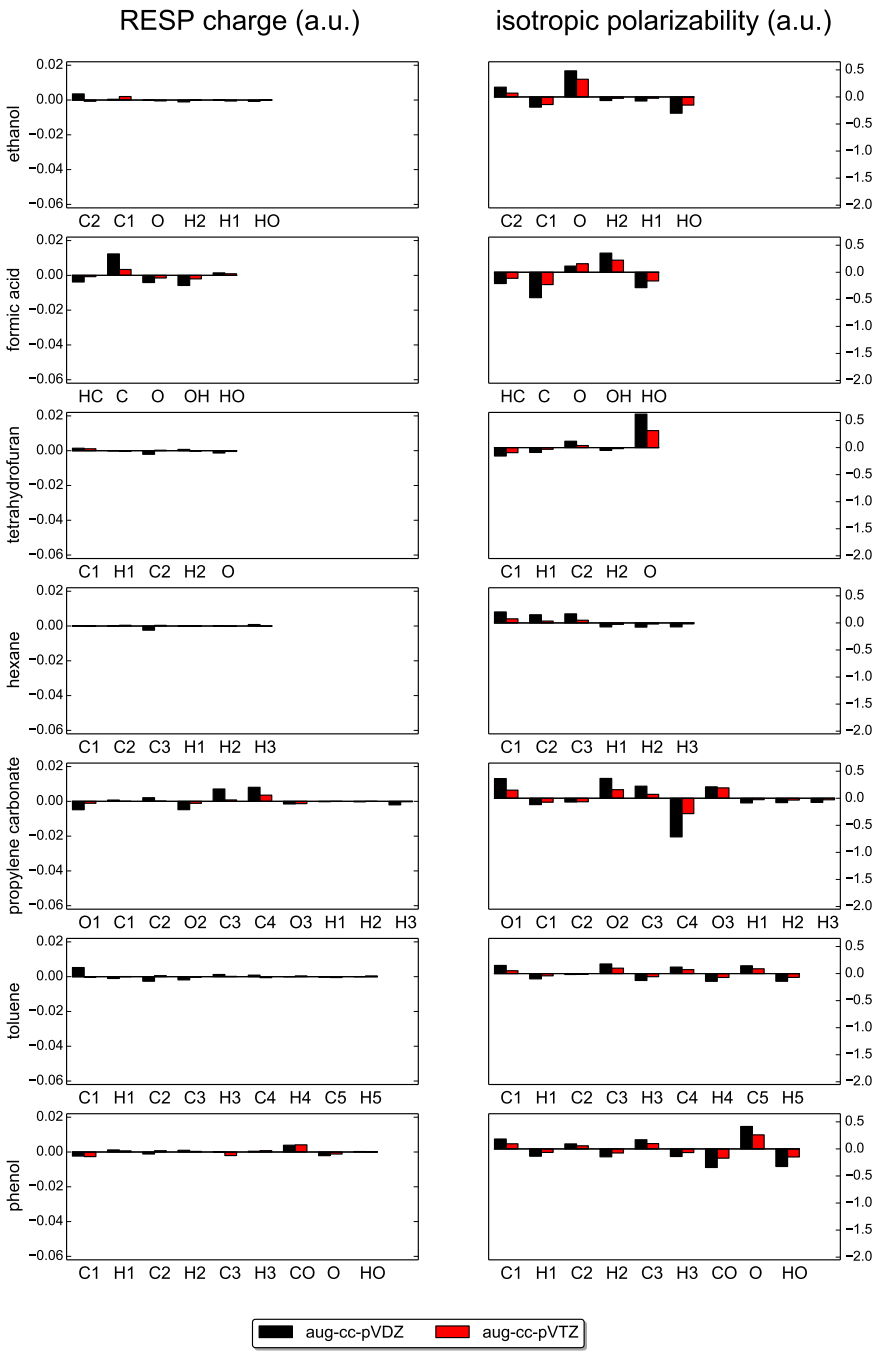


Figure SI-3: b) Basis set dependence of the RESP charges (left) and LoProp isotropic polarizabilities (right) for different solvent molecules. The figures show the deviation of the aug-cc-pVDZ (black) and aug-cc-pVTZ (red) basis sets relative to the aug-cc-pVQZ basis set. The numbers are averages over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

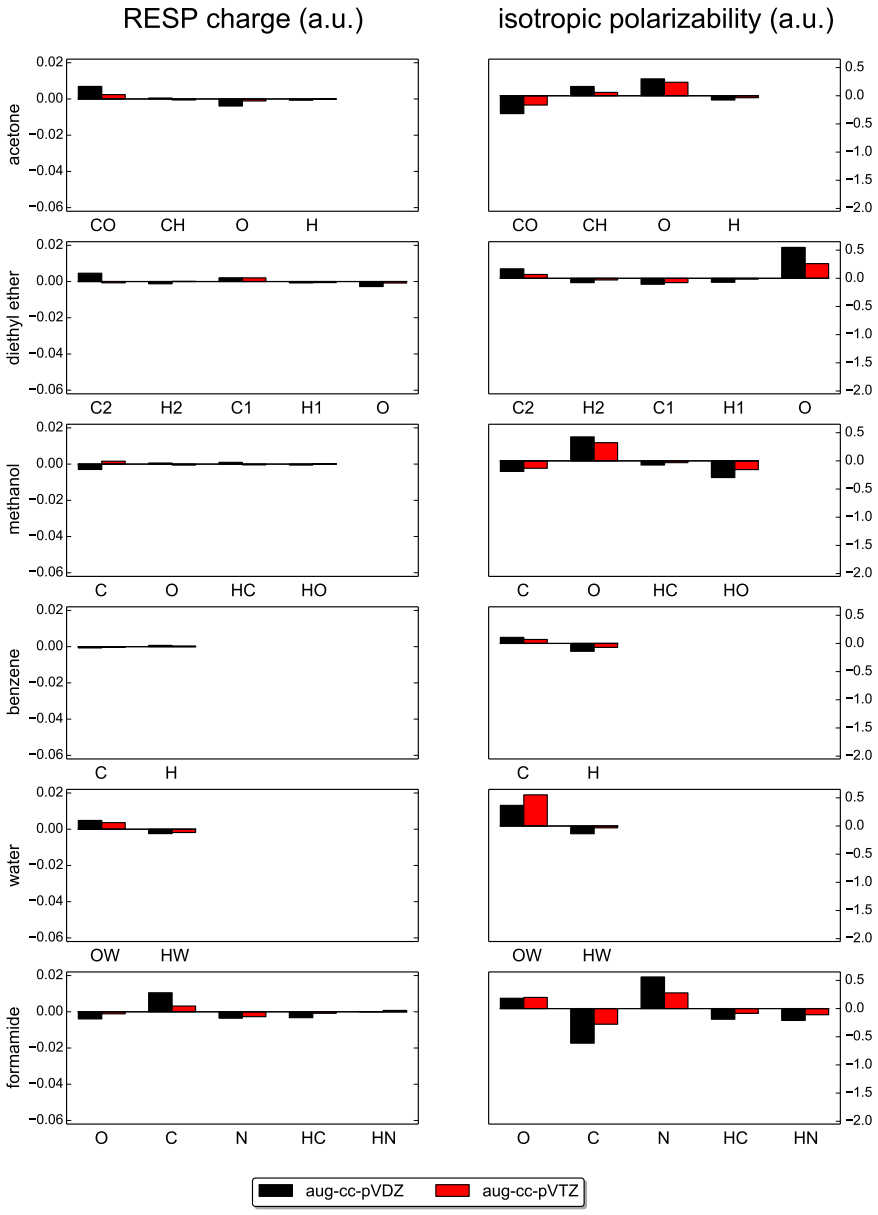


Figure SI-3: c) Basis set dependence of the RESP charges (left) and LoProp isotropic polarizabilities (right) for different solvent molecules. The figures show the deviation of the aug-cc-pVDZ (black) and aug-cc-pVTZ (red) basis sets relative to the aug-cc-pVQZ basis set. The numbers are averages over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

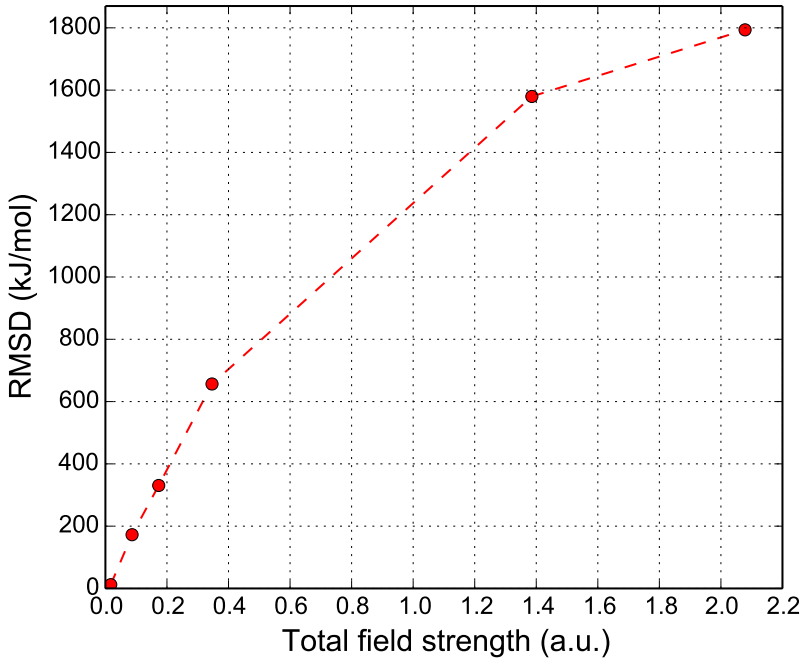


Figure SI-4: RMSD (in kJ/mol) of the induced QM ESP of methanol at different total field strengths with respect to the QM ESP without an applied electric field. The ESPs are calculated with B3LYP/aug-cc-pVTZ as an interaction energy with a unit point charge and evaluated on a molecular surface defined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

# 6 System size threshold tests

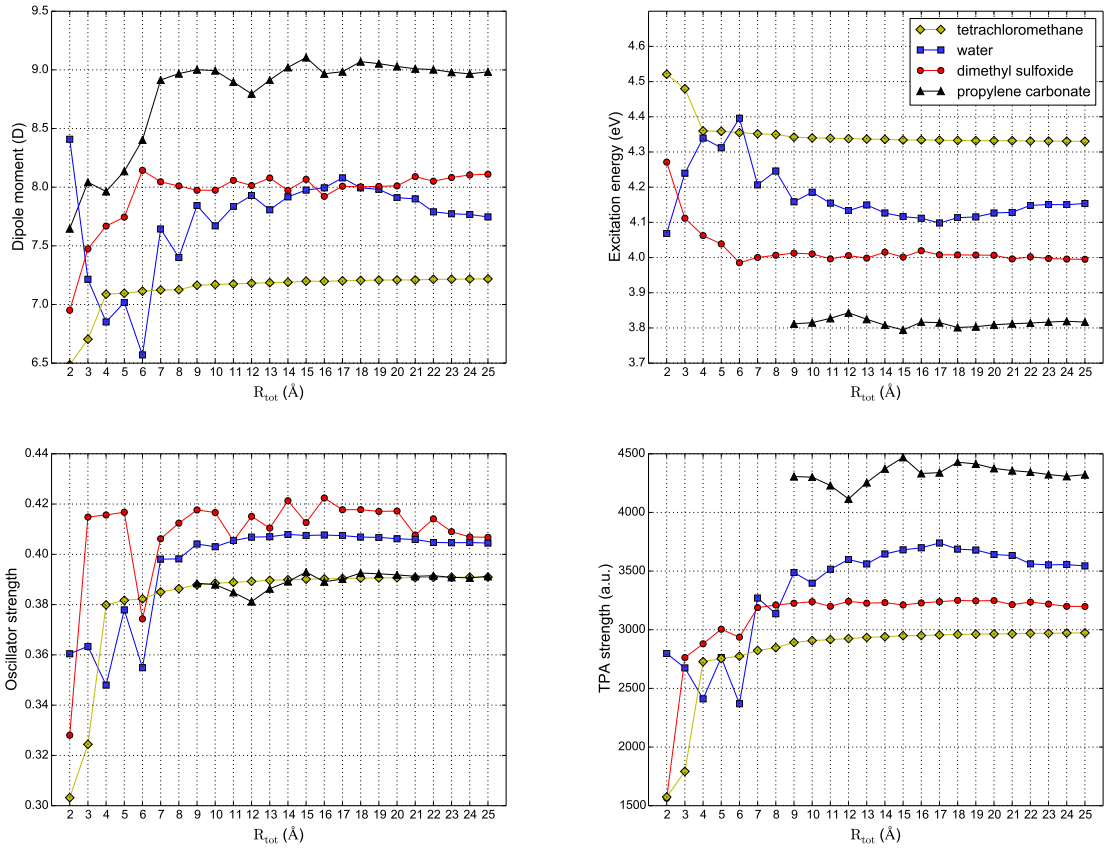


Figure SI-5: The dipole moment (top left), excitation energy (top right), oscillator strength (bottom left) and two-photon transition strength (bottom right) of the charge-transfer excitation of a *para*-nitroaniline (PNA) molecule in water (blue), dimethyl sulfoxide (red), tetrachloromethane (green) 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.

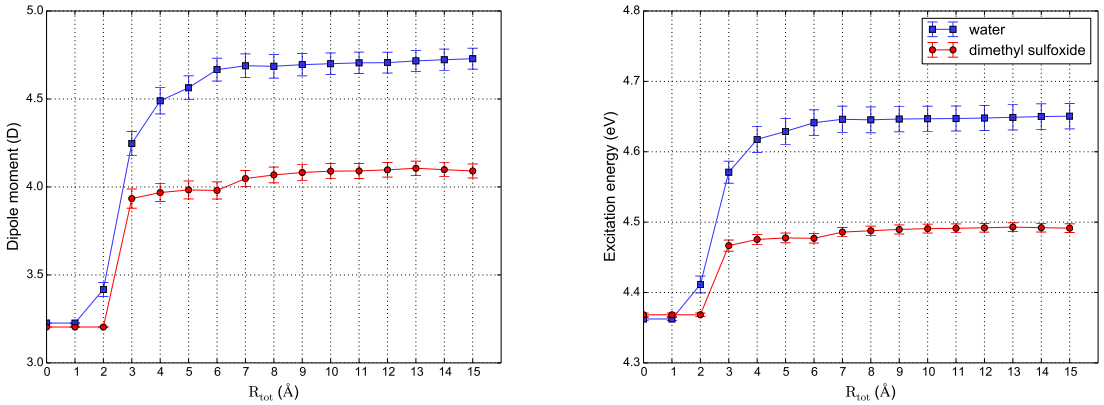


Figure SI-6: The dipole moment (left) and excitation energy (right) of the n$\rightarrow\pi^*$ excitation of acetone in water (blue) and dimethyl sulfoxide (red) solvents. The embedding potential is made up of an M2P2 potential for solvent molecules with at least one atom within the system size threshold $R_{tot}$ ($x$-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→$\pi^*$ 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.