PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2020, **22**, 8021

Prediction of aqueous free energies of solvation using coupled QM and MM explicit solvent simulations†

Daniel Sadowsky *[a,b] and J. Samuel Arey ‡[a]

A method based on molecular dynamics simulations which employ two distinct levels of theory is proposed and tested for the prediction of Gibbs free energies of solvation for non-ionic solutes in water. The method consists of two additive contributions: (i) an evaluation of the free energy of solvation predicted by a computationally efficient molecular mechanics (MM) method; and (ii) an evaluation of the free energy difference between the potential energy surface of the MM method and that of a more computationally intensive first-principles quantum-mechanical (QM) method. The latter is computed by a thermodynamic integration method based on a series of shorter molecular dynamics simulations that employ weighted averages of the QM and MM force evaluations. The combined computational approach is tested against the experimental free energies of aqueous solvation for four solutes. For solute–solvent interactions that are found to be described qualitatively well by the MM method, the QM correction makes a modest improvement in the predicted free energy of aqueous solvation. However, for solutes that are found to not be adequately described by the MM method, the QM correction does not improve agreement with experiment. These preliminary results provide valuable insights into the novel concept of implementing thermodynamic integration between two model chemistries, suggesting that it is possible to use QM methods to improve upon the MM predictions of free energies of aqueous solvation.

Received 2nd February 2020,
Accepted 26th March 2020

DOI: 10.1039/d0cp00582g

rsc.li/pccp

1 Introduction

Prediction of the free energy of aqueous solvation ($\Delta_{\text{solv}}G$), defined here as the difference in partial molar Gibbs free energy between the aqueous phase and the gas phase, is relevant to many fields in the chemical sciences. Quantitative assessments of free energies are of critical importance, for example, to atmospheric chemistry, aquatic chemistry, and biochemistry, and are often established through the experimental determination of Henry's law constants. However, there are many classes of compounds for which the feasibility of such determinations is limited by high reactivity or low aqueous solubility.[1–3] Computational alternatives to experimental determination include linear free energy relationships,[4] electronic structure calculations with continuum solvation models,[5] and molecular simulation.[6,7] Molecular simulations offer the advantage of explicitly considering the complex structural changes that both solute and solvent experience owing to interactions such as hydrogen bonding. However, this rigor comes at the cost of increased computational intensity relative to predictive models where these interactions are treated implicitly.

One possible approach to implementing molecular simulations for the prediction of $\Delta_{\text{solv}}G$ is the use of molecular mechanics (MM) methods.[6,7] This approach is based on the evaluation of computationally efficient analytical potentials that are amenable to the use of alchemical methods. When used for the prediction of $\Delta_{\text{solv}}G$, alchemical methods gradually introduce or eliminate solute–solvent interactions by way of a series of unphysical intermediate states between the coupled, or solvated, state and the uncoupled gas phase and liquid states.[8,9] The intermediate states give access to a thermodynamic path of cavity formation and structuring of the solvent, and therefore a means for the precise and efficient prediction of the hydrophobic effect,[10,11] including the free energy contributions of directed solute–solvent interactions such as hydrogen bonds. However, the approximations of the MM force field also inevitably limit the accuracy of this approach.

---

[a] *Environmental Chemistry Modeling Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), GR C2 544, Station 2, 1015 Lausanne, Vaud, Switzerland*
[b] *Division of Physical and Computational Sciences, University of Pittsburgh at Bradford, 300 Campus Drive, Bradford, Pennsylvania, 16701, USA. E-mail: sadowsky@pitt.edu*

† Electronic supplementary information (ESI) available: Equilibrium unit cell lengths, densities, and molar concentrations; a comparison of MM free energy of solvation results to previous predictions; as well as the time evolution of the difference of and the correlation between the MM and QM energy evaluations for simulations of 63 water molecules. See DOI: 10.1039/d0cp00582g

‡ Present address: ExxonMobil Biomedical Sciences, Inc., Annandale, New Jersey 08801, United States.

As an alternative to the MM approach, one could consider first principles molecular dynamics simulations based on quantum mechanical (QM) electronic structure calculations.[12] These methods potentially offer greater accuracy and applicability than the simplified potentials of MM methods, although at significantly higher computational cost. However, the alchemical transformation techniques that are commonly employed in MM simulations are not possible with current implementations of QM methods.

Here, we develop and test a methodology for the first-principles prediction of $\Delta_{\mathrm{solv}}G$ of a non-ionic solute. We propose an approach that employs two different theoretical model chemistries: a model chemistry based on a MM molecular dynamics description of the solute and solvent, and a second model chemistry based on a QM molecular dynamics description of the (identical) solute/solvent system. Unlike previous QM/MM[13–24] approaches, where each model chemistry describes a separate part of the system, the present approach uses both QM and MM model chemistries to describe the entire system, allowing for the direct comparison of these two potential energy surfaces. The prediction of $\Delta_{\mathrm{solv}}G$ is achieved through the computation of a thermodynamic path that connects the MM and QM potential energy surfaces. This method exploits the computational efficiency of the MM model chemistry, as well as the ease of implementing alchemical transformations with this model chemistry, but also exploits the accuracy and reliability of the QM model chemistry. Finally, the proposed method is modular: varying combinations of existing QM and MM models can be used, in principle, and these models can be chosen according to convenience or performance objectives.

2 Experimental data

To test the proposed approach with a preliminary experimental dataset, we select four experimental free energy of aqueous solvation data that have been carefully vetted by Warneck[25–27] and include experimental uncertainty estimates (Table 1). The test set is selected not only for low experimental uncertainty, but also for structural diversity (each organic compound contains novel functional groups), and for a diversity of capacities to engage in electrostatic (Keesom/Debye), London dispersion, hydrogen-bonding donor and hydrogen-bonding acceptor interactions.

Table 1 Experimental Gibbs free energies of aqueous solvation of the test set of 4 solutes (kJ mol^{-1})

Solute	$\Delta_{\mathrm{solv}}G^{\mathrm{exp.}\,a}$
Argon	8.38 ± 0.04 ($n = 5$)[27]
Methanol	-21.3 ± 0.2 ($n = 8$)[25]
Acetonitrile	-17.7 ± 0.1 ($n = 1$)[27]
trans-1,2-Dichloroethene	-2.36 ± 0.02 ($n = 2$)[26]

[a] Intervals refer to standard deviations of experimental determinations. n refers to number of experimental determinations used in each evaluation.

3 Methodology

3.1 Overview of approach

The free energy paths that link the MM potential energy surface to the QM potential energy surface, or free energies of model chemistry transfer ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G$) as they are referred to herein, are computed for the isolated solute (u), the pure solvent (v), and the combined, solvated system (uv) (as illustrated by the thermodynamic cycle shown in Fig. 1). We use thermodynamic integration to compute $\Delta_{\mathrm{MM}\to\mathrm{QM}}G$ for each subsystem. Thermodynamic integration[28,29] has the advantages of being conceptually simple and reliable to converge, compared to other free energy methods that could have been considered.[30] The computation of these free energy paths opens up the possibility of a QM-quality prediction of the free energy of solvation ($\Delta_{\mathrm{solv}}G^{\mathrm{QM}}$) that is computed as a correction ($\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G$) to an initial MM estimate ($\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$), in which the term $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ is computed with conventional alchemical methods:

$$\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G = \Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv}) - \Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{u}) - \Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v}) \tag{1}$$

$$\Delta_{\mathrm{solv}}G^{\mathrm{QM}} = \Delta_{\mathrm{solv}}G^{\mathrm{MM}} + \Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G \tag{2}$$

The MM model chemistry therefore functions as an auxiliary method that is used to generate an initial estimate, $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$, and not a final prediction of $\Delta_{\mathrm{solv}}G$. In principle, the QM correction, $\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G$, to the MM value could quantitatively compensate for many of the artifacts originating in the choice of MM parameters. Eqn (1) and (2) therefore represent a general, modular methodology to determine a $\Delta_{\mathrm{solv}}G^{\mathrm{QM}}$ value based on a QM potential energy surface that is sampled exhaustively with Born–Oppenheimer molecular dynamics, formulated as a QM correction to an initial $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ estimate provided by a MM method.

3.2 Thermodynamic integration

In the thermodynamic integration formalism, a coupling parameter, λ, is introduced which is varied by discrete values from 0 (defined as the system in its initial state) to 1 (defined as the

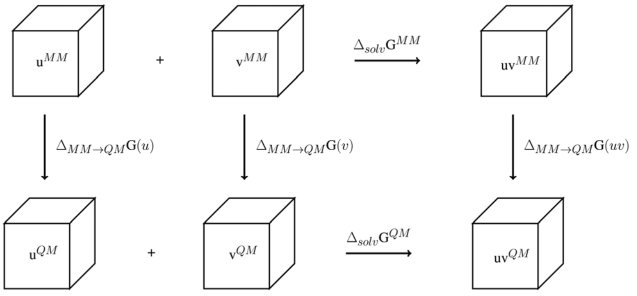


Fig. 1 Conceptual diagram showing the thermodynamic cycle that relates the free energy of aqueous solvation given by the MM method ($\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$), free energy of aqueous solvation given by the QM method ($\Delta_{\mathrm{solv}}G^{\mathrm{QM}}$), and the free energies of model chemistry transfer ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G$) of the solute (u), solvent (v), and solution (uv). Each box represents a simulation cell with the specified contents (u, v, or uv) and the model chemistry (MM or QM) applied.

system in its final state). The free energy associated with the thermodynamic path from 0 to 1 is then computed by sampling the derivative of the potential energy, U, with respect to λ at each value of λ, and integrating:

$$\Delta G = \int_0^1 \left\langle \frac{\delta U(\lambda)}{\delta \lambda} \right\rangle_\lambda \mathrm{d}\lambda \tag{3}$$

Here, all implementations of the thermodynamic integration formalism use the trapezoidal rule with a uniform grid (equally spaced values of λ) to compute the definite integral in eqn (3):

$$\Delta G = \sum_{i=1}^{n_\lambda - 1} \frac{\left\langle \frac{\delta U(\lambda)}{\delta \lambda} \right\rangle_{\lambda = \frac{i}{n_\lambda - 1}} - \left\langle \frac{\delta U(\lambda)}{\delta \lambda} \right\rangle_{\lambda = \frac{i-1}{n_\lambda - 1}}}{n_\lambda - 1} \tag{4}$$

In the present study, thermodynamic integration techniques are used to determine the individual terms in eqn (1) ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{u})$, $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v})$, and $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv})$) and also $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$, as explained in the sections that follow.

3.3 Free energies of solvation computed by MM

The $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ term is computed by a thermodynamic path that comprises three consecutive steps, as illustrated in Fig. 2. Starting with the solvent in a liquid state and the solute in a vacuum: (i) the electrostatic charges are incrementally removed from the solute, "turning off" the solute–solute electrostatic interactions; (ii) the solute–solvent Lennard-Jones interactions are incrementally "turned on", coupling the uncharged solute with the solvent for the first time; and (iii) the electrostatic charges on the solute are incrementally restored, "turning on" both the solute–solute and the solute–solvent electrostatic interactions. Parts (i) and (iii) of this path are each computed from 9 simulations where the electrostatic charges on all of the solute atoms are scaled proportionately with λ. The free energies associated with each of these processes are computed from eqn (3) with 9 equally spaced values of λ. Part (ii) of this path is computed from a series of 33 simulations where all solute–solvent interactions are described by the interatomic potential:

$$U(r) = 4\varepsilon\lambda\left(\left(\frac{1}{2}(1-\lambda) + \left(\frac{r}{\sigma}\right)^6\right)^{-2} - \left(\frac{1}{2}(1-\lambda) + \left(\frac{r}{\sigma}\right)^6\right)^{-1}\right) \tag{5}$$

Eqn (5) converges to the Lennard-Jones potential with parameters ε and σ when $\lambda = 1$, and $U(r)$ becomes 0 when $\lambda = 0$. The particular form of eqn (5) has been designed to avoid the numerical instabilities that would result from linear scaling of the Lennard-Jones potentials.[6,31–33] Because there is no thermal contact between the solute and solvent in part (ii) where $\lambda = 0$, we applied separate thermostats to the solvent and solute for all three parts of the thermodynamic path used to construct $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$.

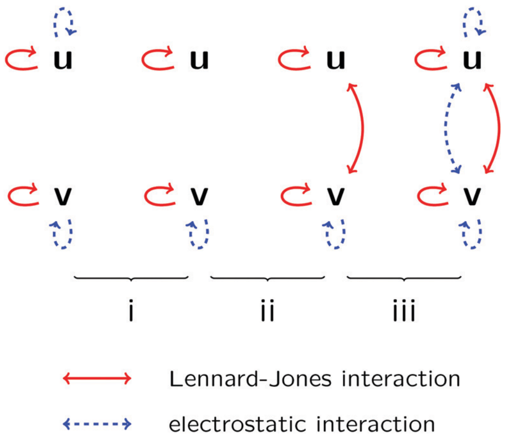


Fig. 2 Overview of thermodynamic path through which $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ is computed (see text for description).

All simulations are carried out using a trunk version (svn:15815) of CP2K version 2.7.[34,35] Each of the simulations used for the $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ thermodynamic path described above are run for 288 ps with a time step of 0.48 fs in the *NPT* ensemble. The target temperature and pressure for these simulations are set at 298.15 K and 1.01325 bar, and are maintained with a Martyna–Tobias–Klein algorithm[36] with time constants of 96 fs for both the thermostat and barostat. A cubic cell is used for all simulations. At the beginning of each simulation, the initial 48 ps of simulated dynamics is considered a system equilibration period, and this simulation output is discarded (including additional equilibration time beyond 48 ps was found to have a negligible effect on the solvation free energy predictions). For the remainder of each simulation, a finite difference method with a spacing parameter, h, of 0.001 is used at each time step to compute the quantity $\delta U/\delta\lambda$ using the formula:

$$\frac{\delta U}{\delta \lambda} = \begin{cases} U(\lambda + h) - U(\lambda) & \lambda = 0 \\ \frac{1}{2}(U(\lambda + h) - U(\lambda - h)) & 0 < \lambda < 1 \\ U(\lambda) - U(\lambda - h) & \lambda = 1 \end{cases} \tag{6}$$

3.4 Free energies of model chemistry transfer

The free energy of model chemistry transfer, $\Delta_{\mathrm{MM}\to\mathrm{QM}}G$, of each system is computed using the weighted average of a QM potential, U^{QM}, and a MM potential, U^{MM}, defined in eqn (7) by introducing a second coupling parameter, λ'.

$$U(\lambda') = \lambda' U^{\mathrm{QM}} + (1 - \lambda')U^{\mathrm{MM}} \tag{7}$$

The derivative of this potential is simply the vertical energy difference between the QM and MM potential energy surfaces:

$$\left\langle \frac{\delta U(\lambda')}{\delta \lambda'} \right\rangle_{\lambda'} = \left\langle U^{\mathrm{QM}} - U^{\mathrm{MM}} \right\rangle_{\lambda'} \tag{8}$$

The free energy, $\Delta_{\mathrm{MM}\to\mathrm{QM}}G$, is obtained by substituting eqn (8) into eqn (3):

$$\Delta_{\mathrm{MM}\to\mathrm{QM}}G = \int_0^1 \left\langle U^{\mathrm{QM}} - U^{\mathrm{MM}} \right\rangle_{\lambda'} \mathrm{d}\lambda' \tag{9}$$

Eqn (9) is evaluated using the formula in eqn (4) for each of the subsystems that appear in Fig. 1: the solute ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{u})$), the solvent ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v})$), and the solution containing the solute ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv})$). These three terms are then employed in eqn (1), which is used to determine the total contribution of model chemistry transfer to the free energy of solution, $\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G$. However, the magnitudes of the terms that describe the solvent, $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v})$, and the solution, $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv})$, are particularly large, and the convergence of $\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G$, which contains a difference of these two large terms (eqn (1)), may prove to be a computational challenge. One of the objectives of the present study is to investigate the practical feasibility of this integration approach.

A solvent bath of 63 water molecules is selected for each solute, which contains enough solvent molecules for each of the presently studied solutes to be surrounded by two full solvent shells. Since the number of solvent molecules is identical for each solution system, uv, it is only necessary to compute $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v})$ once. The determination of $\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G$ for each additional solute therefore requires only the computation of the $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{u})$ and $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv})$ terms.

Sampling the process of model chemistry transfer ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G$) in the *NPT* ensemble poses several difficulties. A longer equilibration time is required than would be required for sampling in the *NVT* ensemble; differences in the equilibrium densities of the MM and QM descriptions of the solvent could lead to an expansion or contraction of the system that does not converge in the limited sampling times that are available; and the choice of QM methods becomes limited to those for which analytical stress tensor calculations are available. We choose to assume that the difference in Gibbs free energy of model transfer of the solution, $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv})$, and the Helmholtz free energy of model transfer of the solution, $\Delta_{\mathrm{MM}\to\mathrm{QM}}A(\mathrm{uv})$, is equivalent to the analogous difference for the isolated solvent:

$$\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv}) - \Delta_{\mathrm{MM}\to\mathrm{QM}}A(\mathrm{uv}) \approx \Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v}) - \Delta_{\mathrm{MM}\to\mathrm{QM}}A(\mathrm{v}) \qquad (10)$$

Eqn (10) implies that the change of model chemistry leads to an expansion or contraction of the solution that is equal to the expansion or contraction experienced by the pure solvent system. According to this assumption, the overall QM correction to the MM free energy of aqueous solvation is approximately the same when sampled in the *NVT* ensemble *versus* the *NPT* ensemble:

$$\Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}G \approx \Delta_{\mathrm{MM}\to\mathrm{QM}}\Delta_{\mathrm{solv}}A \qquad (11)$$

Accurate QM methods such as B97M-V[37] and MP2[38] have been reported to slightly over-predict the density of water with Born–Oppenheimer simulations in the *NPT* ensemble. However, B97M-V has also been shown to perform well for other liquid properties with Born–Oppenheimer simulations in the *NVT* ensemble at the experimental equilibrium density[37] and offer high accuracy potential energy surfaces for water clusters[39] and non-covalent interactions,[40] suggesting that the B97M-V method might be a judicious model chemistry to satisfy the approximation in eqn (11) while introducing low overall error in $\Delta_{\mathrm{solv}}G^{\mathrm{QM}}$.

Each $\Delta_{\mathrm{MM}\to\mathrm{QM}}G$ term in eqn (1) is computed using 9 simulations, each having incrementally differing values of λ' (eqn (4)). Each simulation is 14.4 ps in length, with the first 0.144 ps of dynamics of each simulation discarded for equilibration (the autocorrelation functions of the difference in eqn (8) were found to reach approximately zero at a lag time of 0.144 ps). The average temperature in each simulation is maintained at 298.15 K by a Martyna–Klein–Tuckerman algorithm[41] with a time constant of 96 fs.

The equilibrium density of the MM description of the solutions is predicted from *NPT* simulations. The resulting equilibrium length of each unit cell (reported in Table S.1 in the ESI†) is employed for all *NVT* simulations used to compute $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{u})$ and $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{uv})$. Also reported in Table S.1 (ESI†) is the unit cell determined by *NPT* simulations for 63 water molecules without any solute, and these parameters are used in all *NVT* simulations used to compute $\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v})$. The density of each simulation box and the molar concentration of each box containing a solute are also reported in Table S.1 (ESI†).

3.5 MM method

All polyatomic solutes are described in the MM method by harmonic bonding potentials, harmonic angular potentials, sinusoidal dihedral potentials, and Lennard-Jones interactions. The parameters which define these potentials are taken from the generalized Amber force field[42] (GAFF) version 1.7, except in the case of acetonitrile, in which the parameters of Nikitin and Lyubartsev[43] are used. The latter parameters were determined based on a fit to thermodynamic data for the water–acetonitrile system, which was necessary due to the inability of the GAFF parameters to qualitatively reproduce the acetonitrile potential energy surface.[43] The Lennard-Jones parameters of Vrabec *et al.*[44] are used to describe argon solute, and the aSPC/Fw force field[45] is used to describe water molecules in aqueous solvent. The latter method is chosen because of its anharmonic bonding potential and its ability to reproduce a variety of experimental properties for liquid water with classical simulations. The Lorentz–Berthelot mixing rules are used for all Lennard-Jones parameters, and all 1–4 interactions are scaled by 1/2 and 1/1.2 for Lennard-Jones and electrostatic interactions, respectively. A cutoff length of 6.0 Å is used for all Lennard-Jones interactions.

For each polyatomic solute, CM5 charges[46] are used for the description of electrostatic interactions due to the demonstrated accuracy of resulting predictions of free energies of aqueous solvation.[47] These charges are determined using CM5PAC[48] together with the results of M06-L[49] density functional theoretical calculations obtained with the may-cc-pVTZ basis set[50] and the SMD implicit solvation model,[51] evaluated using the Gaussian 09 revision D.01[52] software suite. For the purpose of MM simulations of aqueous solvent, the original charges specified for aSPC/Fw are retained. Electrostatic interactions are evaluated with the smooth particle mesh Ewald method,[53] a value of 0.8 for the Ewald convergence parameter, a real-space cutoff of 9.0 Å, and 29 grid points in each dimension (for cells that range from 12.1 to 12.8 Å in length).

3.6 QM method

Evaluation of forces on the Born–Oppenheimer potential energy surface is performed with density functional theory using the Gaussian and plane waves method (GPW).[54,55] The density functional method used is based on the B97M-V[39] meta-generalized gradient approximation, which has been demonstrated to predict accurate energies for nonbonded complexes: for example, a root-mean-squared error (RMSE) of 1.05 kJ mol^{-1}[40] was found for a chemically diverse database of benchmark intermolecular binding energies[56] and the RMSE was 1.47 kJ mol^{-1}[39] for a set of benchmark binding energies of water clusters.[57] B97M-V is also notable for its accurate predictions of the geometries of nonbonded complexes.[58] In place of the original VV10[59] nonlocal van der Waals functional with which B97M-V has been optimized, the revised rVV10 functional of Sabatini *et al.*[60] is used in the present study (and not the recent "B97M-rV" re-optimized rVV10 parameters of Mardirossian *et al.*[61]). LIBXC 3.0[62] is used for the meta-generalized gradient evaluation, whereas the default CP2K implementation is used for the evaluation of the nonlocal component. The electronic density is described both by the TZV2P[55] atom-centered Gaussian-type basis set and an auxiliary plane wave basis expanded using a multi-grid method[63] with a cutoff of 280 Ry. Norm-conserving GTH pseudopotentials[64] optimized with the PBE functional[65] are applied to all atoms.

3.7 Benchmark binding energy calculations

Several 1 : 1 solute–water complexes are also studied with both the QM and MM methods. To confirm the accuracy of the QM method for these complexes, CCSD(T)/CBS calculations after the method of Marshall *et al.*[66] (eqn (12)) are implemented in Psi4 1.1.[67]

$$E_{\mathrm{CBS}} = E_{\mathrm{SCF}}^{\text{aug-cc-pVQZ}} + E_{\mathrm{MP2}}^{\text{aug-cc-pV[TQ]Z}} + \delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}\big|_{\text{aug-cc-pVTZ}} \tag{12}$$

3.8 Visualization

All molecular visualizations are produced with the Atomic Simulation Environment.[68]

4 Results

4.1 Predictions of MM free energies of solvation

The results for the predictions of $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ are summarized in Table 2, and are further decomposed into a Lennard-Jones component (part (ii) in Fig. 2), $\Delta_{\mathrm{LJ}}G^{\mathrm{MM}}$, and an electrostatic component (the sum path of parts (i) and (iii) in Fig. 2), $\Delta_{\mathrm{ES}}G^{\mathrm{MM}}$. The overall sampling error in these terms (reported in Table 2) is determined assuming no correlation among the propagated uncertainties of the $\langle \mathrm{d}U/\mathrm{d}\lambda \rangle_\lambda$ values predicted by each simulation trajectory. The sampling error assigned to the $\langle \mathrm{d}U/\mathrm{d}\lambda \rangle_\lambda$ value of each trajectory is a standard deviation computed using the effective number of independent observations for an autocorrelated dynamical variable, using the equation of Bayley and Hammersley.[69]

Among the four solutes in the test set, the $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ predictions that are found to be in the best agreement with experiment are those of argon and acetonitrile. In both of these cases, MM parameters are used that were previously optimized to solute-specific experimental data. Our prediction for *trans*-1,2-dichloroethene, by contrast, deviates from the experimental value by 13.1 kJ mol^{-1}, which is roughly twice the error as the prediction made by Shivakumar *et al.*[70] with GAFF and AM1 charges (as reported in Table S.2 in the ESI†). We interpret that this discrepancy for *trans*-1,2-dichloroethene arises due to our use of the combination of GAFF parameters and CM5 charges, and this is investigated in more detail in Section 4.3 by way of a comparative investigation of solute–water interactions using both the MM method and the higher-quality QM method. Our $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$ prediction for methanol is in better agreement with the predictions of Shivakumar *et al.*; however, our MM prediction for methanol still deviates substantially from experiment. The disparity in the results between the solutes for which the MM parameters have been optimized to solute-specific experimental data (argon and acetonitrile) and the solutes for which the MM parameters have been optimized to both experimental and computational data (methanol and *trans*-1,2-dichloroethene) illustrates the need to consider solute-specific QM corrections to MM predictions of free energies of aqueous solvation. The average unsigned error in $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$, 6.9 kJ mol^{-1}, of the four solutes considered here is comparable to those reported by Shivakumar *et al.* for much larger sets of solutes, where they employed a variety of similar methods.

4.2 Convergence of free energies of model chemistry transfer

The computed values of the free energy of model chemistry transfer are reported in Table 3 for the aqueous solvent ($\Delta_{\mathrm{MM}\to\mathrm{QM}}G(\mathrm{v})$), for

Table 2 The Lennard-Jones component, $\Delta_{\mathrm{LJ}}G^{\mathrm{MM}}$, electrostatic component, $\Delta_{\mathrm{ES}}G^{\mathrm{MM}}$, and total free energy of solvation, $\Delta_{\mathrm{solv}}G^{\mathrm{MM}}$, predicted by the MM method for 4 test solutes, and the values, $\Delta_{\mathrm{solv}}G^{\mathrm{exp.}}$, given by experiment (kJ mol^{-1})

Solute	$\Delta_{\mathrm{LJ}}G^{\mathrm{MM}\,a}$	$\Delta_{\mathrm{ES}}G^{\mathrm{MM}\,a}$	$\Delta_{\mathrm{solv}}G^{\mathrm{MM}\,a}$	$\Delta_{\mathrm{solv}}G^{\mathrm{exp.}\,b}$
Argon	9.7 ± 0.1	—	9.7 ± 0.1	8.38 ± 0.04
Methanol	8.4 ± 0.2	−18.1 ± 0.1	−9.7 ± 0.2	−21.3 ± 0.2
Acetonitrile	10.9 ± 0.1	−28.0 ± 0.1	−17.2 ± 0.2	−17.7 ± 0.1
trans-1,2-Dichloroethene	13.1 ± 0.2	−2.47 ± 0.02	10.7 ± 0.2	−2.36 ± 0.02

[a] Intervals refer to standard deviations computed with the effective number of independent observations.[69] [b] Reproduced from Table 1. Intervals refer to standard deviations of experimental determinations.

all four solutes ($\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{u})$), and the corresponding solutions ($\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$), as well as the complete $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ given by eqn (1). By inspection of Table 3, the quantities $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{u})$, $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$, and $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ exhibit very large magnitudes of order $\approx 10^5$–10^7 kJ mol^{-1}. Each $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ term is constructed from the integration of $\langle \mathrm{d}U/\mathrm{d}\lambda' \rangle_{\lambda'}$ (eqn (8)), which is simply the averaged vertical energy difference between the MM and QM potential energy surfaces (eqn (7)). The observed large magnitudes of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ terms (Table 3) therefore arise from the arbitrary choices of the zeros that are assigned to the MM and QM potential energy surfaces. However, the fluctuations in $\mathrm{d}U/\mathrm{d}\lambda'$ over the course of the simulations are independent of these choices, and these energy fluctuations are much smaller in magnitude than the instantaneous potential energy difference $U^{\mathrm{MM}} - U^{\mathrm{QM}}$. For example, the 9 simulations performed for liquid water each have standard deviations in their $\mathrm{d}U/\mathrm{d}\lambda'$ values of 52–61 kJ mol^{-1} over the course of each simulation (plots of the time evolution of $\mathrm{d}U/\mathrm{d}\lambda'$ for two different values of λ' are included in panels (a) and (b) of Fig. S.1 in the ESI†). However, due to the exhaustive sampling of $\mathrm{d}U/\mathrm{d}\lambda'$, the predictions of $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{u})$, $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$, and $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ shown in Table 3 all have estimated sampling uncertainties of ≤ 1.1 kJ mol^{-1}, despite the large fluctuations in instantaneous values of $\mathrm{d}U/\mathrm{d}\lambda'$. As a consequence, the aggregated QM correction to the MM free energy of solution, $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$, is found to be well-converged within a sampling error of ≤ 1.5 kJ mol^{-1} for all four solutes in the test set. The assumptions behind this estimate of convergence are explained here in further detail.

The convergence of the individual $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ values depends on the number of discrete values of the coupling parameter, λ' (equivalent to the number of individually simulated trajectories, $n_{\lambda'}$), as well as the sampled length of the molecular dynamics trajectory, $t_{\lambda'}$, that is applied at each incremental value of λ'. The values of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$ integral and $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ terms are converged with respect to $n_{\lambda'}$ to within ≤ 1.5 kJ mol^{-1} for values of $n_{\lambda'} = 9$ (Fig. 4a and 5a).

The term $\langle \mathrm{d}U/\mathrm{d}\lambda' \rangle_{\lambda'}$ has substantial non-linear behavior with respect to λ' in the computation of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ terms, although this is difficult to determine from visual inspection of the thermodynamic integration (Fig. S.3 in the ESI†) due to the large magnitude of the $\langle \mathrm{d}U/\mathrm{d}\lambda' \rangle_{\lambda'}$ terms. For this reason, we have chosen to visualize the effect of $n_{\lambda'}$ on the convergence of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ terms by separating the non-linear contributions from the linear contributions. Linear approximations are constructed for the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{u})$, $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$, and $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ thermodynamic integrations for methanol. These linear approximations use the $\langle \mathrm{d}U/\mathrm{d}\lambda' \rangle_{\lambda'}$ data points at $\lambda' = 0$ and $\lambda' = 1$ to effectively estimate the $\langle \mathrm{d}U/\mathrm{d}\lambda' \rangle_{\lambda'}$ values at all intermediate points, λ^*, with residuals, r:

$$r(\lambda^*) = \left\langle \frac{\mathrm{d}U}{\mathrm{d}\lambda'} \right\rangle_{\lambda'=\lambda^*} - \lambda^* \left\langle \frac{\mathrm{d}U}{\mathrm{d}\lambda'} \right\rangle_{\lambda'=0} - (1-\lambda^*) \left\langle \frac{\mathrm{d}U}{\mathrm{d}\lambda'} \right\rangle_{\lambda'=1} \qquad (13)$$

A mathematical interpretation of these residuals (Fig. 3) is that they represent the non-linear component of the thermodynamic integration at a given point, λ', and can be integrated to compute the difference between the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ integrals computed with $n_{\lambda'} = 8$ and those computed with $n_{\lambda'} = 2$. The substantial magnitude of the residuals (over 20 kJ mol^{-1} for the solvent) is interpreted as arising from the interactions of solvent molecules described by potential energy surfaces of the averaged MM and QM model chemistries. The non-linearity of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ integral in Fig. 3 is similar enough to the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$ integral that it leads to a partial cancellation of these dependencies that is favorable to the convergence of $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ (which involves taking a difference of $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ and $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$ (eqn (1))) with respect to $n_{\lambda'}$. However, the non-linear behaviors are qualitatively different enough and the dependence of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ predictions on $n_{\lambda'}$ large enough that the consideration of several intermediate λ' values is warranted when integrating between MM and QM descriptions. The $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{u})$ integrals have a non-linear dependence which is much smaller than those of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})$ and $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ integrals.

While the dependencies of $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ terms on $t_{\lambda'}$ (Fig. 4b and 5b) appear smaller than those with respect to $n_{\lambda'}$ (Fig. 4a and 5a), this does not necessarily mean that these predictions are fully converged with respect to $t_{\lambda'}$. The high computational cost of condensed phase DFT calculations limits the $t_{\lambda'}$ values that can be used in this study and therefore limits the exploration of configuration space that can be achieved by the method. It remains unclear whether these short sampling times can adequately represent the ergodic limit, even within our careful estimates of the statistical sampling uncertainty. However, given that the lifetime of a hydrogen bond in liquid water is known to be on the order of 1 ps[71,72] and orientational times of the first hydration shell for hydrophobic solutes on the order of 3 ps,[73] a $t_{\lambda'}$ value of 14.4 ps may be adequate for these

Table 3 Free energies of model chemistry transfer, $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$, for each isolated solute (u), for the system of 63 aqueous solvent (v) molecules, and for each solution (uv) of the solute molecule in 63 solvent molecules ($n_{\mathrm{traj}} = 9$, $t_{\mathrm{traj}} = 14.4$ ps), as well as the aggregated QM correction to the MM free energies of solvation, $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ (kJ mol^{-1})

Solute	$\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{u})^a$	$\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{v})^a$	$\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})^a$	$\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G^b$
Argon	−55394.87	−2843127.0 ± 1.0	−2898523.0 ± 1.1	−1.5 ± 1.5
Methanol	−63141.99 ± 0.04	−2843127.0 ± 1.0	−2906283.0 ± 0.9	−14.1 ± 1.3
Acetonitrile	−60354.64 ± 0.10	−2843127.0 ± 1.0	−2903461.4 ± 0.8	20.2 ± 1.3
trans-1,2-Dichloroethene	−111511.35 ± 0.04	−2843127.0 ± 1.0	−2954638.7 ± 0.7	−0.3 ± 1.2

[a] Intervals refer to standard deviations computed with the effective number of independent observations.[69] [b] Computed using eqn (1). Intervals refer to standard deviations computed with the effective number of independent observations.[69]

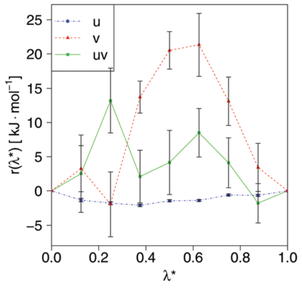
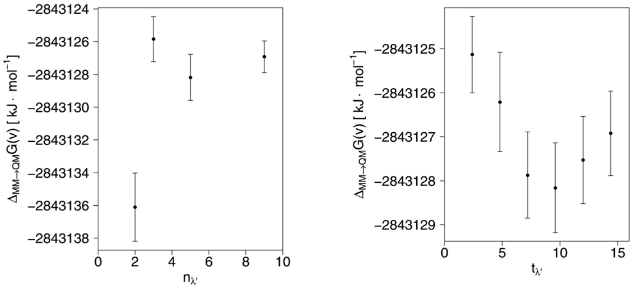


Fig. 3 The non-linear components, $r(\lambda^*)$, of the $\Delta_{MM\to QM}G$ thermodynamic integration of methanol, (u); 63 solvent molecules, (v); and the combined methanol and solvent, (uv), computed using eqn (13).



(a) (b)

Fig. 4 Convergence of $\Delta_{MM\to QM}G(v)$ with respect to (a) the number of trajectories, $n_{\lambda'}$, included in the thermodynamic integration ($t_{\lambda'}$ = 14.4 ps) and (b) the length of simulated dynamics, $t_{\lambda'}$, from each trajectory that is included ($n_{\lambda'}$ = 9). Intervals refer to propagated standard deviations computed with the effective number of independent observations (see Section 4.1 for details).

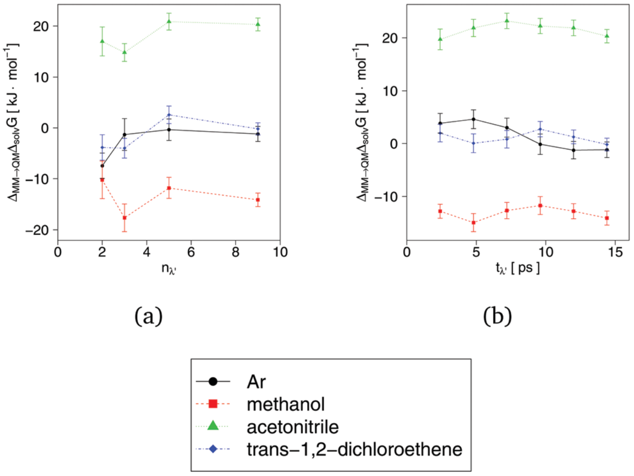


Fig. 5 Convergence of the four $\Delta_{MM\to QM}\Delta_{solv}G$ predictions with respect to (a) the number of trajectories, $n_{\lambda'}$, included in the thermodynamic integration ($t_{\lambda'}$ = 14.4 ps) and (b) the length of simulated dynamics, $t_{\lambda'}$, from each trajectory that is included ($n_{\lambda'}$ = 9). Intervals refer to propagated standard deviations computed with the effective number of independent observations (see Section 4.1 for details).

non-ionic solutes. Within the bounds of these assumptions, the numerical precision of the $\Delta_{MM\to QM}\Delta_{solv}G$ predictions that can be achieved is notable.

4.3 Predictions of QM free energies of solvation

The QM predictions of the aqueous free energies of solvation are calculated from eqn (2) and reported in Table 4 together with the MM predictions and experimental values. For argon and methanol, the inclusion of the QM correction significantly improves agreement with experiment compared to the MM prediction alone. However the QM correction worsens agreement with experiment for acetonitrile and has little effect on the MM prediction for *trans*-1,2-dichloroethene. Each of these cases is investigated in turn. As part of these investigations, the potential energy surface of each solute interacting with a single water molecule in the gas phase is explored with the QM method. These QM calculations are implemented with the same unit cell as the corresponding model chemistry transfer simulations (Table S.1 of the ESI†). The binding energy of each local minimum, $\Delta_{bind}E$, is defined as the difference in the potential energy of the bound conformation and the potential energy of the two molecules at infinite separation, excluding the zero-point vibrational energy and assuming a temperature of 0 K. The QM binding energies are compared to the binding energies predicted by the CCSD(T)/CBS method at the QM geometries, as well as those of the MM method. The MM binding energies use MM-optimized geometries in cases where the MM potential energy surface has an analogous local minimum, and the QM geometries in all other cases.

5 Discussion

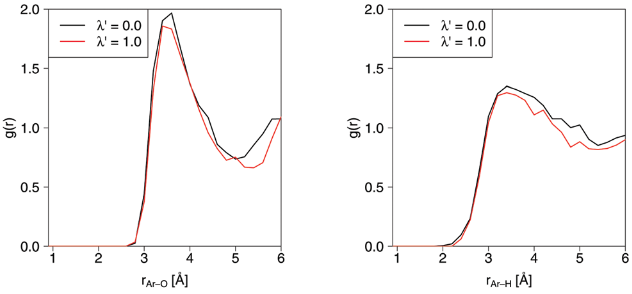
5.1 Argon

The aqueous solvation of argon is dominated by the hydrophobic effect, so much so that it has a positive experimental free energy of solvation. Interestingly, the pair correlation functions between the argon atom and the solvent oxygen and hydrogen atoms in Fig. 6 suggest that the solvent around argon may be slightly more structured with the QM method than it is with the MM method. This might be expected to result in a positive $\Delta_{MM\to QM}\Delta_{solv}G$ value rather than the observed value of -1.5 ± 1.5 kJ mol^{-1}. However, the contribution of the dispersion interactions between argon and the solvent, which lead to a lowering of the free energy of solvation, must also be considered. The potential energy surface of gas phase argon complexed with a single water molecule is found to have a single local minimum on both the QM and MM potential energy surfaces. The QM method predicts a binding energy (Table 5) which is 0.9 kJ mol^{-1} lower than the MM method.

Table 4 Comparison of MM predictions, QM predictions (eqn (2)), and experimental values of the free energies of aqueous solvation for four test solutes (kJ mol^{-1})

Solute	$\Delta_{solv}G^{MM\,a}$	$\Delta_{solv}G^{QM\,b}$	$\Delta_{solv}G^{exp.\,c}$
Argon	9.7 ± 0.1	8.2 ± 1.5	8.38 ± 0.04
Methanol	−9.7 ± 0.2	−23.8 ± 1.4	−21.3 ± 0.2
Acetonitrile	−17.2 ± 0.2	3.0 ± 1.3	−17.7 ± 0.1
trans-1,2-Dichloroethene	10.7 ± 0.2	10.4 ± 1.2	−2.36 ± 0.02

[a] Reproduced from Table 2. [b] Calculated using eqn (2). [c] Reproduced from Table 1.



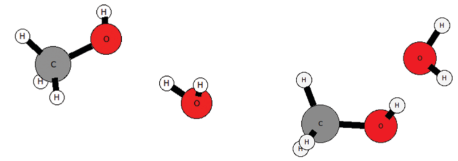
(a) (b)

Fig. 6 Pair correlation functions of (a) argon atoms with solvent hydrogen atoms and (b) argon atoms with solvent oxygen atoms during the initial, fully MM (λ' = 0) and final, fully QM (λ' = 1) model chemistry transfer simulations of methanol with 63 water molecules.

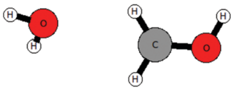
Table 5 Binding energies (kJ mol^{-1}) of the argon–water complex

Method	$\Delta_{\mathrm{bind}}E$
MM[a]	−0.8
QM[b]	−1.7
CBS[c]	−1.2

[a] See Section 3.5 for details. [b] B97M-rV/TZV2P. [c] CCSD(T)/CBS//B97M-rV/TZV2P.



conformation	**1** (C_1)	**2** (C_1)
$\Delta_{bind}E^{MM}$	-23.4	-24.2
$\Delta_{bind}E^{QM}$	-21.8	-22.6
$\Delta_{bind}E^{CBS}$	-21.6	-20.7



conformation	**3** (C_1)
$\Delta_{bind}E^{MM}$	*7.3**
$\Delta_{bind}E^{QM}$	-4.7
$\Delta_{bind}E^{CBS}$	-5.0

Fig. 7 Local minima of the B97M-rV/TZV2P potential energy surface of the methanol–water complex and corresponding QM and MM binding energies (kJ mol^{-1}). MM binding energies are computed with MM optimized geometries, except where noted with an asterisk.

This suggests that the dominant effect of the thermodynamic integration of the model chemistry transfer simulations was actually to correct for the slight underbinding of argon with water predicted by the MM method rather than to correct the ability of the MM method to model the hydrophobic effect.

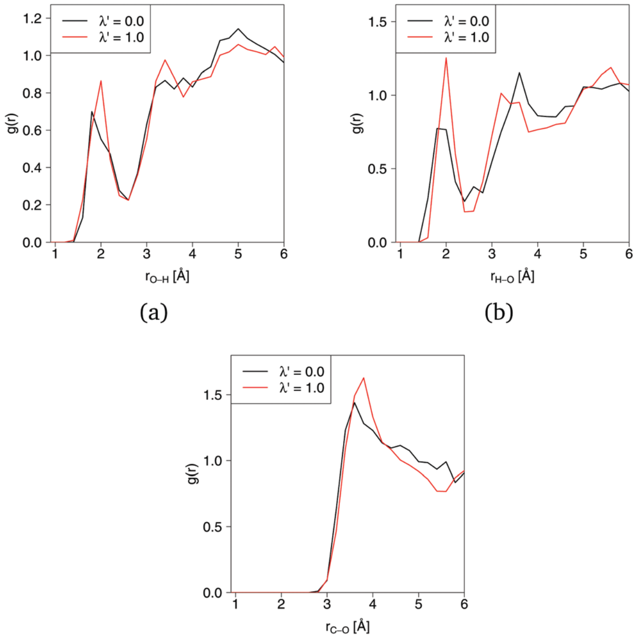
5.2 Methanol

The aqueous solvation of methanol is dominated by the hydrogen bonding. This is also true of the potential energy surface of a 1 : 1 complex of methanol with a single water molecule, which is found to have three local minima: one conformation (**1**) where methanol is accepting a hydrogen bond, another conformation (**2**) where methanol is donating a hydrogen bond, and a third conformation (**3**), significantly higher in energy than the first two, where the solvent molecule is closest to the methyl group of methanol. Whereas the MM method is in good agreement with the QM method for the two hydrogen bonding conformations, it predicts a repulsive interaction with no local minimum for conformation (**3**) where the QM methods predicts a slightly attractive interaction. The DFT-based QM method is found to be in good agreement with the more rigorous CBS calculations for all three conformations (Fig. 7).

The prediction of a negative $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ correction for methanol (-14.1 ± 1.3 kJ mol^{-1}) is interesting considering that the MM method actually slightly overbinds the hydrogen bonding conformations. Pair correlation functions from initial and final model chemistry transfer simulations of methanol and the 63 solvent molecules (Fig. 8) suggest that the negative $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ term may be a result of a QM solvent shell which is closer to the solute than the MM solvent shell. This is particularly evident in the pair correlation functions between the methanol carbon atom and solvent oxygen atoms (Fig. 8c), which show significantly increased structure in the QM (λ' = 1) simulation. It is likely that this increased solvent structure is due, in part, to interactions similar to the one in conformation (**3**), which the MM incorrectly predicts to be repulsive and which the QM method correctly predicts to be attractive. It therefore appears that the thermodynamic integration of the model chemistry transfer simulations does have an effect on the solvent cavity in the case of methanol, but that this effect has more to do with the methyl group than it does with hydrogen bonding. It is curious that there would be such a large $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}\Delta_{\mathrm{solv}}G$ correction associated with the MM and QM descriptions of a hydrophobic functional group when there appears to be no significant correction to the hydrophobic effect in argon; we suspect that this is due to the particular choices made for the description of the solutes in the MM method.

5.3 Acetonitrile

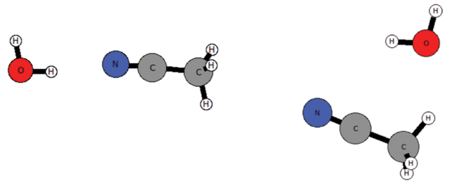
Like methanol, the experimental free energy of solvation of acetonitrile is highly negative due to hydrogen bonding, although acetonitrile can only accept a hydrogen bond while methanol may also donate one. Also, acetonitrile contains π-bonding, whereas methanol contains only σ-bonds. Three unique local minima are found on the potential energy surface of the QM acetonitrile–water complex (Fig. 9): one configuration in which acetonitrile acts as a hydrogen bond acceptor (**1**), one configuration which features an electrostatic interaction between the nitrile π-cloud and the water molecule (**2**), and one configuration which features an electrostatic interaction



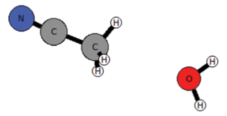
(a) (b)

(c)

Fig. 8 Pair correlation functions of (a) solute oxygen atoms with solvent hydrogen atoms, (b) solute (hydroxyl) hydrogen atoms with solvent oxygen atoms, and (c) solute carbon atoms with solvent oxygen atoms during the initial, fully MM ($\lambda' = 0$) and final, fully QM ($\lambda' = 1$) model chemistry transfer simulations of methanol with 63 water molecules.



conformation	**1** (C_s)	**2** (C_1)
$\Delta_{bind}E^{MM}$	-28.6	*-1.7**
$\Delta_{bind}E^{QM}$	-20.9	-15.3
$\Delta_{bind}E^{CBS}$	-19.2	-16.3



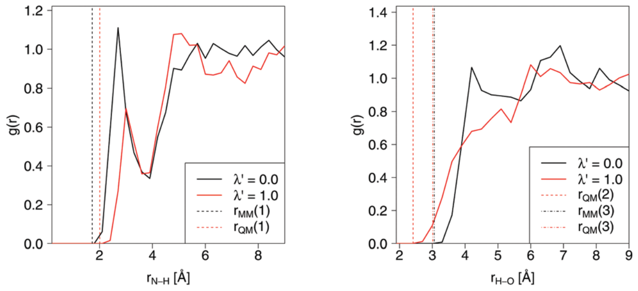
conformation	**3** (C_s)
$\Delta_{bind}E^{MM}$	-17.2
$\Delta_{bind}E^{QM}$	-12.2
$\Delta_{bind}E^{CBS}$	-9.1

Fig. 9 Local minima of the QM potential energy surface of the acetonitrile–water complex with corresponding QM and MM binding energies (kJ mol^{-1}). MM binding energies are computed with MM optimized geometries, except where noted with an asterisk.

between the electron-deficient methyl group and the water molecule (**3**).

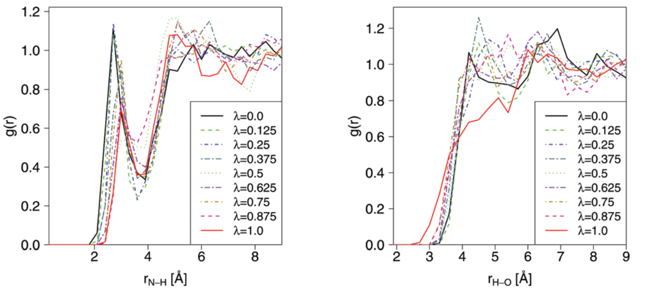
The MM method predicts stronger binding to water than does the QM method except in conformation (**2**), which features π-bonding. Unlike the MM models that are used for methanol and *trans*-1,2-dichloroethene, which are parameterized to fit a wide range of data, the Nikitin–Lyubartsev model for acetonitrile was parameterized to achieve agreement with experimental data specific to the water–acetonitrile system, and so the agreement of $\Delta_{solv}G^{MM}$ with experiment is not surprising. However, the MM method relies on electrostatic point charges, and therefore it cannot model π-bonding. The model apparently overbinds conformations **1** and **3** in order to compensate for the unphysical underbinding of conformation **2** that cannot be avoided due to the limitations of the functional form of the MM model. Inspection of the pair correlation functions shows that the cavity reorganization is substantial during the thermodynamic integration from the MM to QM potential energy surface (Fig. 10). The pair correlation functions between the solute nitrogen and solvent hydrogen atoms clearly show a decrease in the frequency and an increase spatial length of hydrogen bonding interactions in the QM potential energy surface relative to the MM potential energy surface, which is consistent with our analysis of conformation (**1**) of the acetonitrile–water potential energy surface. Conversely, the pair correlation functions between the solute hydrogen and solvent oxygen atoms show both an increase in close interactions in QM simulations relative to MM simulations, which is probably due to increased π-bonding interactions such as in conformation (**2**), but they also exhibit a decrease in interactions at around 3 Å interatomic distance, most likely due to less favorable interactions between solvent molecules and the methyl group of acetonitrile such as in conformation (**3**).

The $\Delta_{solv}G^{MM}$ prediction for acetonitrile is in good agreement with experiment. One might expect the solvent reorganization effects described above to cancel each other out, but in fact the QM prediction is 20.7 kJ mol^{-1} higher than experiment. Inspection of the pair correlation functions of the intermediate,



(a) (b)

Fig. 10 Pair correlation functions of (a) solute nitrogen atoms with solvent hydrogen atoms and (b) solute hydrogen atoms with solvent oxygen atoms during the initial, fully MM ($\lambda' = 0$) and final, fully QM ($\lambda' = 1$) model chemistry transfer simulations of acetonitrile with 63 water molecules. The nitrogen–hydrogen bond lengths of the MM and QM geometries of conformation (**1**) in Fig. 9 are marked in (a). The hydrogen–oxygen bond lengths of the MM and QM geometries of conformation (**3**) as well as the QM geometry of conformation (**2**) are marked in (b).



(a) (b)

Fig. 11 Pair correlation functions of (a) solute nitrogen atoms with solvent hydrogen atoms and (b) solute hydrogen atoms with solvent oxygen atoms during the initial, fully MM ($\lambda' = 0$), final, fully QM ($\lambda' = 1$), and the seven mixed-Hamiltonian model chemistry transfer simulations of acetonitrile with 63 water molecules.

mixed-Hamiltonian simulations (Fig. 11) reveal that while the weakening of the hydrogen bonding interactions happens gradually over the course of the nine simulations, the formation of the energetically favorable π-interactions do not form until the final, fully QM simulation. We suspect that this is due to the MM component of the mixed-Hamiltonian simulations predicting repulsive energies for the types of geometries that are sampled in the final, fully QM simulation. Even for $\lambda' = 0.875$, the relatively small contribution that the MM method makes to the Hamiltonian could be enough to prevent the geometries from being sampled if the MM method predicts them to be sufficiently high in energy. This is similar to the problem that occurs in the linear interpolation of two MM potentials with thermodynamic integration,[6,32,33] where intermediate ($0 < \lambda < 1$) trajectories can sometimes sample conformational space that is near equilibrium in one MM potential but highly repulsive in the other MM potential. Although we have found no evidence for "singularities" (that is, predictions of sharp increases in the intermediate potential and/or numerical instabilities), it does appear that the presence of any MM character to the Hamiltonian at all prevents effective sampling of important interactions.

An exploration of the non-linear components to the thermodynamic integration of the combined acetonitrile and solvent simulations (uv) shows large contributions from the later mixed-Hamiltonian simulations such as $\lambda^* = 0.625$, 0.75, and 0.875 (Fig. 12). These are simulations with mostly QM Hamiltonians, and they are also simulations in which the solvent has restructured to allow for less hydrogen bonding than exists in the purely MM simulation, but has not yet restructured to allow for attractive interactions between solvent molecules and the π-cloud, such as exist in conformation (2) in Fig. 9. The high $r(\lambda^*)$ values could therefore be interpreted as a consequence of the failure to sample geometries containing attractive interactions between the solvent and the π-cloud. The high $r(\lambda^*)$ values also mean that these simulations make anomalously large and positive contributions to the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G(\mathrm{uv})$ term, and thus the final $\Delta_{\mathrm{solv}}G^{\mathrm{QM}}$ prediction.

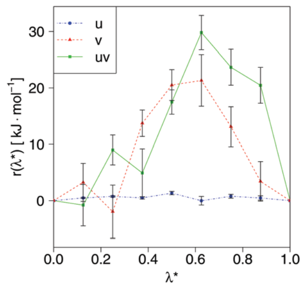
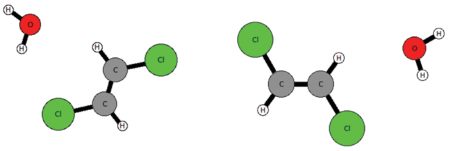


Fig. 12 The non-linear components, $r(\lambda^*)$, of the $\Delta_{\mathrm{MM}\rightarrow\mathrm{QM}}G$ thermodynamic integration of acetonitrile, (u); 63 solvent molecules, (v); and the combined acetonitrile and solvent system, (uv), computed using eqn (13).

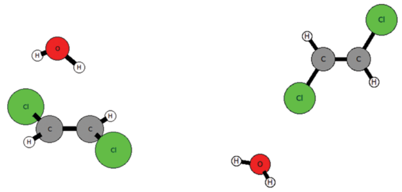
5.4 *trans*-1,2-Dichloroethene

The experimental aqueous solvation free energy of *trans*-1,2-dichloroethene is slightly negative, presumably due to moderate electrostatic interactions with the solvent. However, the anisotropic nature of the electronic density around the carbon and chlorine atoms due to π-density and the σ hole effect, respectively, pose a challenge to the computational modeling of these electrostatic interactions with the solvent. The potential energy surface of a 1 : 1 complex of *trans*-1,2-dichloroethene with a single water molecule is found to have four unique local minima (Fig. 13): two (**1**,**2**) configurations featuring electrostatic interactions between both a hydrogen and chlorine from *trans*-1,2-dichloroethene and the water molecule, one (**3**) configuration featuring an electrostatic interaction between the π-cloud and the water molecule, and one (**4**) configuration which features a weak halogen bond.

The QM method predicts stronger binding than the MM method in all of the local minimum energy conformations considered; the QM results are further supported by the CBS calculations (Fig. 13). The QM results also confirm that the MM



conformation	**1** (C_s)	**2** (C_s)
$\Delta_{bind}E^{MM}$	-8.9	-7.7
$\Delta_{bind}E^{QM}$	-13.1	-12.4
$\Delta_{bind}E^{CBS}$	-13.3	-12.7



conformation	**3** (C_2)	**4** (C_1)
$\Delta_{bind}E^{MM}$	*9.6**	*23.3**
$\Delta_{bind}E^{QM}$	-8.4	-5.3
$\Delta_{bind}E^{CBS}$	-6.1	-4.3

Fig. 13 Local minima of the B97M-rV/TZV2P potential energy surface of the *trans*-1,2-dichloroethene-water complex and corresponding QM and MM binding energies (kJ mol^{-1}). MM binding energies are computed with MM optimized geometries, except where noted with an asterisk.

method predicts unphysical repulsive interactions for conformations where the QM method predicts weakly attractive π-bonds and halogen bonds. This is in agreement with the findings of El Hage *et al.*,[74] who studied the effect of including atomic multipole expansions into force field descriptions of halogenated aromatics and found that including quadrupole terms on both the halogen atom and neighboring carbon atoms is necessary for obtaining accurate solvation free energies and a correct solvation shell structure. To test whether or not these qualitative differences in the MM and QM potential energy surfaces may have affected the structure of the solvent around the solute, we investigate pair correlation functions, $g(r)$, between solute and solvent atoms over the course of selected model chemistry transfer simulations (Fig. 14).

The chlorine–oxygen pair correlation function from the QM simulation (Fig. 14a) suggests an increase in the ordering of the solvent around the chlorine atoms, compared to the MM simulation. The carbon–oxygen and carbon–hydrogen correlation functions (Fig. 14b and c) possibly also suggest an increase in the ordering of the solvent above the plane of the solute, albeit a more subtle one. This is consistent with our analysis of the solute–water complex potential energy surface; the difference between the energies predicted by the QM and MM methods is greater for the halogen-bonding conformation (**4**) than it is for the conformation that includes a π-interaction (**3**). However, the $\Delta_{\mathrm{solv}}G^{\mathrm{QM}}$ prediction for *trans*-1,2-dichloroethene is higher than the experimental value by 13 kJ mol^{-1} (Table 4) despite the evidence that the QM potential energy surface is accurate and the solvent has undergone reorganization to reflect the qualitative differences in the potential energy surfaces. To explain this, we investigate the non-linear components of the model chemistry transfer thermodynamic integration, $r(\lambda^*)$ (Fig. 15).

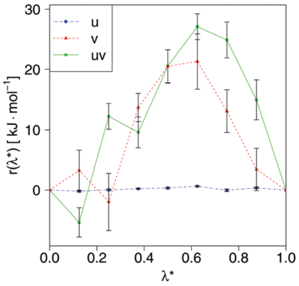
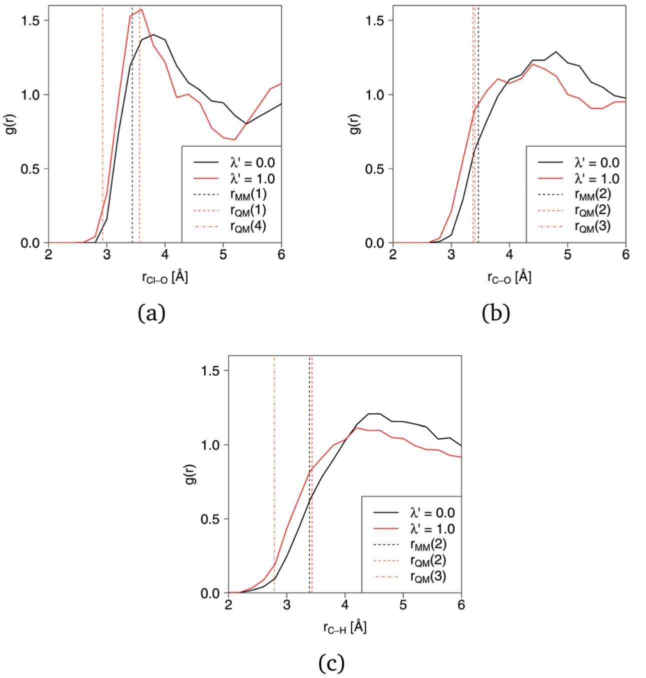


Fig. 15 The non-linear components, $r(\lambda^*)$, of the $\Delta_{\mathrm{MM}\to\mathrm{QM}}G$ thermodynamic integration of gas phase *trans*-1,2-dichloroethene, (u); 63 solvent molecules, (v); and the combined *trans*-1,2-dichloroethene and solvent system, (uv), computed using eqn (13).

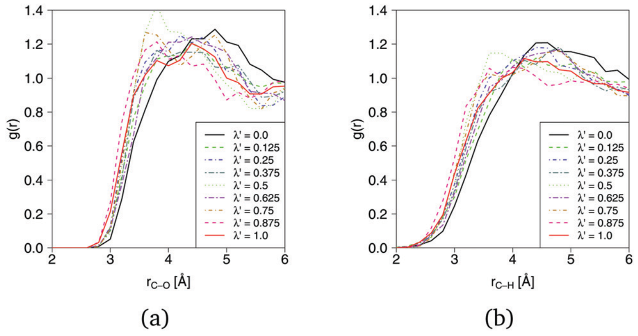
The $r(\lambda^*)$ values for the combined *trans*-1,2-dichloroethene and solvent simulations (uv) are substantially larger than those of the well-behaved example of methanol (Fig. 3). Especially notable are those of the four final mixed-Hamiltonian simulations, λ^* = 0.5, 0.625, 0.75, and 0.875. The solute–solvent carbon–oxygen and carbon–hydrogen pair correlation functions from the mixed-Hamiltonian simulations (Fig. 16) show an interesting feature of these simulations: the solvent appears to get even closer to the carbon atoms of the solute during these four final mixed-Hamiltonian simulations than it does during the final, fully QM simulation. One possible explanation for this is that, while the final QM simulation arrives at a solvent structure where the underbinding of MM description of the solute and the solvent is corrected, the intermediate simulations arrive a solvent structure which is overcorrected. In theory, sufficiently long sampling of these intermediate states should resolve this issue; however, the practical result of the unphysical



(a) (b)

(c)

Fig. 14 Pair correlation functions of (a) solute chlorine atoms with solvent oxygen atoms, (b) solute carbon atoms with solvent oxygen atoms, and (c) solute carbon atoms with solvent hydrogen atoms during the initial, fully MM (λ' = 0) and final, fully QM (λ' = 1) model chemistry transfer simulations of *trans*-1,2-dichloroethene with 63 water molecules. The chlorine–oxygen bond lengths of the MM and QM geometries of conformation (**1**) and the QM geometry of conformation (**4**) in Fig. 13 are marked in (a). The carbon–oxygen and carbon–hydrogen bond lengths of the MM and QM geometries of conformation (**2**) and the QM geometries of conformation (**3**) are marked in (b) and (c), respectfully.



(a) (b)

Fig. 16 Pair correlation functions of solute carbon atoms with (a) solvent oxygen atoms and (b) solvent hydrogen atoms during the initial, fully MM (λ' = 0), final, fully QM (λ' = 1), and the seven mixed-Hamiltonian model chemistry transfer simulations of *trans*-1,2-dichloroethene with 63 water molecules.

MM model seems to be that even with the computationally intensive simulation lengths used here, the sampling of these model chemistry transfer simulations is biased towards geometries with relatively high short-range repulsive forces which, in turn, results in an incorrectly high prediction of the free energy of solvation.

6 Conclusions

The QM corrections to MM free energies of solvation that we propose do not converge to the pertinent experimental free energies of solvation in two cases, acetonitrile and *trans*-1,2-dichloroethene. Moreover, the convergence of these QM corrections with respect to the number of intermediate states or the length of simulated dynamics (Fig. 5) does not suggest that a moderate increase in either quantity would be sufficient to achieve convergence with experiment. In both of these cases, the underlying MM methods which are being corrected do not fully account for atomic-level anisotropy of the electrostatic potential. However, we also provide two cases where the MM method does give a good qualitative description of solute–solvent interactions: argon and methanol. In both cases, the addition of the QM correction improves agreement with experiment. Interestingly, the QM correction for methanol is substantial; the lack of quantitative accuracy of the MM description of methanol seems to be unimportant, as it describes qualitatively correct solute–solvent interactions. However, in the case of acetonitrile, where the MM method accurately predicts the experimental free energy of solvation by means of unphysical solute–solvent interactions, the QM correction does not converge. As the use of methods such as soft core potentials, which expedite the convergence of the thermodynamic integration between two qualitatively different MM descriptions, are not possible for the thermodynamic integration between a MM method and a QM method, we instead recommend the exclusive use of MM methods which give a qualitatively correct description of all solute–solvent weak interactions for the thermodynamic integration between MM and QM methods.

Another possible limitation of the proposed methodology is that *ab initio* methods which accurately model Born–Oppenheimer surfaces typically can not be expected to accurately predict ensemble properties without the use of quantum nuclear methods.[37,75] The quantitative agreement with experiment that we achieve with our free energy of solvation predictions for argon and methanol suggest that the use of QM methods without a quantum treatment of nuclei, but with the experimental density enforced, is perhaps an acceptable compromise between accuracy and computational efficiency for free energy of solvation predictions. However, the unit cells used here for the QM correction of the solvent–solute system were taken from *NPT* simulations with the MM method, which again underscores the importance of the choice of MM method for achieving quantitative accuracy with the method we propose.

The small number of solvent molecules in our implementation should also be noted. Parameswaran and Mobley[76] have found that the number of solvent molecules have a negligible effect on the free energy of solvation predictions of non-ionic solutes. Although the smallest simulation box studied by Parameswaran and Mobley (18 Å) is larger than the simulation boxes used in our implementation (all less than 13 Å), 63 solvent molecules is sufficient for two solvation shells with the solutes considered here, and so we suspect any box size effects in our present implementation are minor. However, a much larger number of solvent molecules would be necessary for the consideration of ionic solutes with this method.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was supported in part by the University of Pittsburgh Center for Research Computing through the provision of computational resources used for the CBS calculations. The authors thank Giovanni Cangiani for HPC support and Ivano Tavernelli, Peter R. Tentscher, Daniela Trogolo, and Jennifer J. Guerard for helpful discussions.

References

1 W.-Y. Wen and J. A. Muccitelli, *J. Solution Chem.*, 1979, **8**, 225–246.

2 H. A. Bamford, D. L. Poster and J. E. Baker, *Environ. Toxicol. Chem.*, 1999, **18**, 1905–1912.

3 J. Tolls, J. van Dijk, E. J. M. Verbruggen, J. L. M. Hermens, B. Loeprecht and G. Schüürmann, *J. Phys. Chem. A*, 2002, **106**, 2760–2765.

4 S. Endo and K.-U. Goss, *Environ. Sci. Technol.*, 2014, **48**, 12477–12491.

5 C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161–2200.

6 M. R. Shirts and V. S. Pande, *J. Chem. Phys.*, 2005, **122**, 134508.

7 D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley and W. Sherman, *J. Chem. Theory Comput.*, 2010, **6**, 1509–1519.

8 W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, 1985, **83**, 3050–3054.

9 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.

10 M. R. Shirts, D. L. Mobley and J. D. Chodera, *Annu. Rep. Comput. Chem.*, 2007, **3**, 41–59.

11 J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson and V. S. Pande, *Curr. Opin. Struct. Biol.*, 2011, **21**, 150–160.

12 D. Marx and J. Hutter, *Ab initio molecular dynamics: basic theory and advanced methods*, Cambridge University Press, 2009.

13 V. Luzhkov and A. Warshel, *J. Comput. Chem.*, 1992, **13**, 199–213.

14 N. Vaidehi, T. A. Wesolowski and A. Warshel, *J. Chem. Phys.*, 1992, **97**, 4264–4271.

15 J. Gao and X. Xia, *Science*, 1992, **258**, 631–635.
16 T. Wesolowski and A. Warshel, *J. Phys. Chem.*, 1994, **98**, 5183–5187.
17 J. Gao and M. Freindorf, *J. Phys. Chem. A*, 1997, **101**, 3182–3188.
18 W. Yang, Q. Cui, D. Min and H. Li, *Annu. Rep. Comput. Chem.*, 2010, **6**, 51–62.
19 G. König and S. Boresch, *J. Comput. Chem.*, 2011, **32**, 1082–1090.
20 B. Lev, B. Roux and S. Y. Noskov, *J. Chem. Theory Comput.*, 2013, **9**, 4165–4175.
21 G. König, P. S. Hudson, S. Boresch and H. L. Woodcock, *J. Chem. Theory Comput.*, 2014, **10**, 1406–1419.
22 C. Sampson, T. Fox, C. S. Tautermann, C. Woods and C.-K. Skylaris, *J. Phys. Chem. B*, 2015, **119**, 7030–7040.
23 X. Lu, D. Fang, S. Ito, Y. Okamoto, V. Ovchinnikov and Q. Cui, *Mol. Simul.*, 2016, **42**, 1056–1078.
24 G. König, Y. Mei, F. C. Pickard, A. C. Simmonett, B. T. Miller, J. M. Herbert, H. L. Woodcock, B. R. Brooks and Y. Shao, *J. Chem. Theory Comput.*, 2016, **12**, 332–344.
25 P. Warneck, *Atmos. Environ.*, 2006, **40**, 7146–7151.
26 P. Warneck, *Chemosphere*, 2007, **69**, 347–361.
27 P. Warneck and J. Williams, *The Atmospheric Chemist's Companion: Numerical Data for Use in the Atmospheric Sciences*, Springer Science & Business Media, 2012.
28 J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.
29 E. Darve, *Free Energy Calculations*, Springer Berlin Heidelberg, 2007, pp. 119–170.
30 H. Paliwal and M. R. Shirts, *J. Chem. Theory Comput.*, 2011, **7**, 4115–4134.
31 T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber and W. F. van Gunsteren, *Chem. Phys. Lett.*, 1994, **222**, 529–539.
32 J. W. Pitera and W. F. van Gunsteren, *Mol. Simul.*, 2002, **28**, 45–65.
33 T. Steinbrecher, D. L. Mobley and D. A. Case, *J. Chem. Phys.*, 2007, **127**, 214108.
34 CP2K version 2.7 (svn:15815). CP2K is freely available from: http://www.cp2k.org/.
35 J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 15–25.
36 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.
37 L. Ruiz Pestana, N. Mardirossian, M. Head-Gordon and T. Head-Gordon, *Chem. Sci.*, 2017, **8**, 3554–3565.
38 M. Del Ben, M. Schönherr, J. Hutter and J. VandeVondele, *J. Phys. Chem. Lett.*, 2013, **4**, 3753–3759.
39 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2015, **142**, 074111.
40 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **145**, 186101.
41 G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.
42 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
43 A. M. Nikitin and A. P. Lyubartsev, *J. Comput. Chem.*, 2007, **28**, 2020–2026.
44 J. Vrabec, J. Stoll and H. Hasse, *J. Phys. Chem. B*, 2001, **105**, 12126–12133.
45 K. Park, W. Lin and F. Paesani, *J. Phys. Chem. B*, 2012, **116**, 343–352.
46 A. V. Marenich, S. V. Jerome, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2012, **8**, 527–541.
47 J. Z. Vilseck, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2014, **10**, 2802–2812.
48 K. Duanmu, B. Wang, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *CM5PAC version 2015*, University of Minnesota, *Minneapolis*, 2015.
49 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.
50 E. Papajak, J. Zheng, X. Xu, H. R. Leverentz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 3027–3034.
51 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
52 *Gaussian 09 Revision D.01*, Gaussian Inc., Wallingford CT, 2009.
53 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
54 G. Lippert, J. Hutter and M. Parrinello, *Mol. Phys.*, 1997, **92**, 477–488.
55 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.
56 D. E. Taylor, J. G. Ángyán, G. Galli, C. Zhang, F. Gygi, K. Hirao, J. W. Song, K. Rahul, O. A. von Lilienfeld, R. Podeszwa, I. W. Bulik, T. M. Henderson, G. E. Scuseria, J. Toulouse, R. Peverati, D. G. Truhlar and K. Szalewicz, *J. Chem. Phys.*, 2016, **145**, 124105.
57 B. Temelso, K. A. Archer and G. C. Shields, *J. Phys. Chem. A*, 2011, **115**, 12034–12046.
58 J. Witte, M. Goldey, J. B. Neaton and M. Head-Gordon, *J. Chem. Theory Comput.*, 2015, **11**, 1481–1492.
59 O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.*, 2010, **133**, 244103.
60 R. Sabatini, T. Gorni and S. de Gironcoli, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 041108.
61 N. Mardirossian, L. Ruiz Pestana, J. C. Womack, C.-K. Skylaris, T. Head-Gordon and M. Head-Gordon, *J. Phys. Chem. Lett.*, 2017, **8**, 35–40.
62 M. A. Marques, M. J. Oliveira and T. Burnus, *Comput. Phys. Commun.*, 2012, **183**, 2272–2281.
63 G. Lippert, J. Hutter and M. Parrinello, *Theor. Chem. Acc.*, 1999, **103**, 124–140.
64 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.
65 M. Krack, *Theor. Chem. Acc.*, 2005, **114**, 145–152.
66 M. S. Marshall, L. A. Burns and C. D. Sherrill, *J. Chem. Phys.*, 2011, **135**, 194102.
67 R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney,

T. D. Crawford and C. D. Sherrill, *J. Chem. Theory Comput.*, 2017, **13**, 3185–3197.

68 S. R. Bahn and K. W. Jacobsen, *Comput. Sci. Eng.*, 2002, **4**, 56–66.

69 J. M. H. G. V. Bayley, *J. R. Stat. Soc.*, 1946, **8**, 184–197.

70 D. Shivakumar, Y. Deng and B. Roux, *J. Chem. Theory Comput.*, 2009, **5**, 919–930.

71 C. J. Montrose, J. A. Bucaro, J. Marshall-Coakley and T. A. Litovitz, *J. Chem. Phys.*, 1974, **60**, 5025–5029.

72 F. W. Starr, J. K. Nielsen and H. E. Stanley, *Phys. Rev. Lett.*, 1999, **82**, 2294–2297.

73 N. Galamba, *J. Phys. Chem. B*, 2014, **118**, 4169–4176.

74 K. El Hage, T. Bereau, S. Jakobsen and M. Meuwly, *J. Chem. Theory Comput.*, 2016, **12**, 3008–3019.

75 L. Ruiz Pestana, O. Marsalek, T. E. Markland and T. Head-Gordon, *J. Phys. Chem. Lett.*, 2018, **9**, 5009–5016.

76 S. Parameswaran and D. L. Mobley, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 825–829.