Predicting Solvent Effects on S_N2 Reaction Rates – Comparison of QM/MM, Implicit and MM Explicit Solvent Models

Mackenzie Taylor,[1] Haibo Yu[2] and Junming Ho[1]*

[1] *School of Chemistry, The University of New South Wales, Sydney, NSW 2052, Australia*

[2] *Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, NSW 2522, Australia*

Abstract

Solvents are one of the key variables in the optimisation of a synthesis yield or properties of a synthesis product. In this paper, contemporary solvent models are applied to predict the rates of S_N2 reactions in a range of aqueous and non-aqueous solvents. High-level CCSD(T)/CBS//M06-2X/6-31+G(d) gas phase energies were combined with solvation free energies from SMD, SM12 and ADF-COSMO-RS continuum solvent models as well as molecular mechanics (MM) explicit solvent models with different atomic charge schemes to predict the rate constants of three S_N2 reactions in eight protic and aprotic solvents. It is revealed that popular solvent models struggle to predict their rate constants to within 3 log units of experimental values and deviations as large as 7.6 log units were observed. Amongst the implicit solvent models, the ADF-COSMO-RS model performed the best in predicting absolute rate constants with an average accuracy of 1.5 log units while the SM12 and CGenFF/TIP3P MM explicit solvent models were most accurate in the prediction of *relative* rate constants in different solvents due to systematic error cancellation. Free energy barriers obtained from umbrella sampling with explicit solvent QM/MM simulations led to excellent agreement with experiment provided that a validated level of theory is used to treat the QM region.

* Correspondence author email: junming.ho@unsw.edu.au

Introduction

Nucleophilic substitution reactions are amongst the most widely used reactions in synthetic organic chemistry and biochemistry. Prominent examples include methyl transfer reactions catalysed by methyltransferase enzymes[1] and in Williamson ether synthesis ("O-methylation"),[2] as well as carbon-carbon bond forming alkylation reactions that are ubiquitous in organic synthesis. Despite their importance, it remains very challenging to reliably predict the kinetics and thermodynamics of these reactions in a consistent fashion.

A reason underlying this difficulty is the lack of robust methods that can accurately describe solvent effects. Notably, the vast majority of synthetic and biochemical reactions occur in the solution phase, and solvents can have a profound influence over the outcome of the reaction. For example, the barrier for the S_N2 identity reaction ($Cl^- + CH_3Cl$) is increased by more than 25 kcal mol^{-1} going from the gas phase to water,[3] which corresponds to 18 orders of magnitude reduction in rate constant at room temperature. One way to quantify solvent effects is to consider a thermodynamic cycle which expresses the Gibbs free energy change of a solution phase process (ΔG_{soln}) in terms of the gas phase free energy change (ΔG_{gas}) and the solvation free energies (ΔG_S) of the reactants, products and/or transition state. Figure 1 illustrates an example cycle for the calculation of the solution phase barrier of a nucleophilic substitution S_N2 reaction. In Equation 1, $\Delta\Delta G_S$ refers to the net gain or loss in solvation free energy due to the reaction. An advantage of using a thermodynamic cycle is that it allows one to analyse a free energy change in terms of intramolecular and intermolecular contributions, where one can think of $\Delta\Delta G_S$ as the thermodynamic driving force due to the solvent. Additionally, when highly accurate QM methods (such as CCSD(T)/CBS) are employed for the calculation of ΔG_{gas} (Equation 1) and the assumptions of transition state theory hold true, the discrepancy between calculated and experimentally measured rate constants can be directly attributed to the quality of the calculated $\Delta\Delta G_S$.

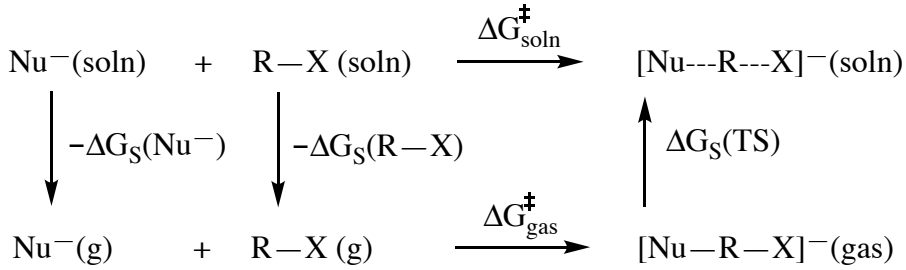

Figure 1. Thermodynamic cycle that expresses the solution phase S_N2 Gibbs free energy barrier in terms of the gas phase barrier and solvation free energies of reactants and transition state.

$$\begin{aligned} \Delta G^{\neq}_{\text{soln}} &= \Delta G^{\neq}_{\text{gas}} + \Delta G_S(\text{TS}) - \Delta G_S(\text{R-X}) - \Delta G_S(\text{Nu}^-) \\ &= \Delta G^{\neq}_{\text{gas}} + \Delta\Delta G_S \end{aligned} \tag{1}$$

At present, quantum mechanical (QM) continuum solvent models (also known as implicit solvent models) are the most popular approach for the prediction of ΔG_S due to their low-cost, moderate accuracy, and ease of use.[4] Examples include the polarisable continuum model and its variants (*e.g.* C-PCM,[5-6] the SM*x* (*x* = 8, 12, D)[7-8] and COSMO-RS[9-10] models. Many of these models have been parameterised to reproduce experimental data and hence are generally valid only at room temperature for selected neat solvents. While these models have reported accuracies of about 1 kcal mol^{-1} for the prediction of hydration free energies of neutral solutes, errors of many kcal mol^{-1} are common for ionic solutes.[11] This is presumably due to the implicit nature of these models which do not directly account for short-range specific interactions such as hydrogen bonding. Some improvements have been reported when the interactions between the solute and a small number of explicit solvent molecules are modelled using quantum chemical methods;[12-13] however, these cluster-continuum approaches introduce other complications such as the number of solvent molecules that should be added, and the treatment of conformational and anharmonic vibrational effects.[11] Hybrid implicit-explicit approaches such as 3D-RISM[14] has also been reported though they remain under development and are not yet broadly applicable. It is worth pointing out that in order to predict a rate or equilibrium constant to within an order of magnitude of experiment, the error in the calculated

ΔG_{soln} in Equation 1 must be smaller than 1.4 kcal mol^{-1} at room temperature. This level of rigor is also known as *chemical accuracy*, and previous work has shown that continuum solvent models often represent the bottleneck in terms of accuracy when combined with high-level QM methods.[15]

Another approach is to employ explicit solvent simulations where the solute-solvent interactions are modelled explicitly in molecular dynamics (or Monte Carlo) simulations and statistical free energy methods are used to extract the free energy of solvation (ΔG_S), or solvation free energy changes along a reaction path ($\Delta\Delta G_S$). As these methods require extensive configurational sampling to obtain converged values, they are generally implemented using approximate molecular mechanics (MM) force fields where inter and intra-molecular interactions are modelled classically. Nevertheless, these models have been very successful in the prediction of solvation free energies of neutral solutes in aqueous and organic solvents,[16-17] though they are less widely used to model ionic solutes or transition states.[3, 18-20]

An alternative to the use of a thermodynamic cycle is to perform direct dynamics using *ab initio* or hybrid quantum mechanics/molecular mechanics (QM/MM) potentials along the reaction path.[21-26] In this approach, the solution phase energy changes are directly determined (*i.e.* no separate gas phase calculation) using statistical free energy methods in conjunction with configurational sampling using molecular dynamics (MD) or Monte Carlo simulations. For example, Acevedo and Jorgensen have demonstrated successful applications of QM/MM and polarisable OPLS-AAP force field simulations of many organic reactions, including Menchutkin and ionic S_N2 reactions.[27-29] Similarly, Higashi and Truhlar have reported the use of density functional theory in conjunction with electrostatically embedded multiconfigurational molecular mechanics to study a model S_N2 reaction.[30] However, these methods require significantly more computational resources where construction of a reaction profile involving 20-50 sampling windows can easily consume many tens of thousands of CPU hours depending on the choice of QM treatment. This is approximately 3-4 orders of magnitude more resources compared to a QM implicit solvent calculation. Additionally, the quality of QM/MM

simulations depends on a number of parameters such as size of the QM region and pairing of QM and MM potentials that can have a significant impact on accuracy.[31-33] These considerations have limited the wide use of these approaches.

Thus, it is of interest to examine if computationally expensive QM/MM simulations translate to more accurate predictions, particularly when off-the-shelf continuum solvent models fail. In this context, there have been very few studies that directly compare the performance of these two approaches. Stirling and co-workers have compared the use of QM/MM molecular dynamics and static DFT continuum solvent calculations to evaluate the energetics of Ag-catalysed furan ring formation and concluded that the latter was adequate in uncovering the mechanistic features of the reaction.[34] The Shao group has also compared the performance of various continuum solvent models, MM and QM/MM explicit solvent models to evaluate the hydration free energies of neutral solutes. It was observed that the use of computationally expensive QM/MM potentials did not lead to a significant improvement in overall performance.[35] Henchman and co-workers have also employed an energy-entropy method to calculate the barriers of S_N2 reactions in water using QM/MM MD simulations and DFT implicit solvent calculations.[21] The authors observed that the implicit solvent model (IEF-PCM) incurred very large errors whilst the QM/MM values are in better agreement with experiment.

In this work, we sought to examine the ability of contemporary solvent models to predict solvent effects on the rates of S_N2 reactions. In particular, can these models reliably predict *solvent induced changes* in: (1) the reactivity of a nucleophile, and (2) the relative reactivity between different nucleophiles. This will enable chemists to make informed decisions concerning the choice of computational methods to guide the choice of solvents and reagents to optimise reaction yields and/or kinetics. To this end, we compared the performance of widely used QM continuum solvent models, namely Minnesota solvation models (SMD, SM12) and ADF-COSMO-RS as well as MM and QM/MM explicit solvent models for predicting absolute and relative rate constants of three S_N2 reactions depicted in Figure 2. These reactions were chosen because they have experimentally

measured rate constants in *eight* different solvents that span up to six orders of magnitude. These small rigid systems also permit the use of highly accurate quantum chemistry methods to calculate $\Delta G^{\ddagger}_{gas}$ as well as minimise complications associated with anharmonicity and conformational contributions so that one can directly assess the ability of the models to describe solvation effects.

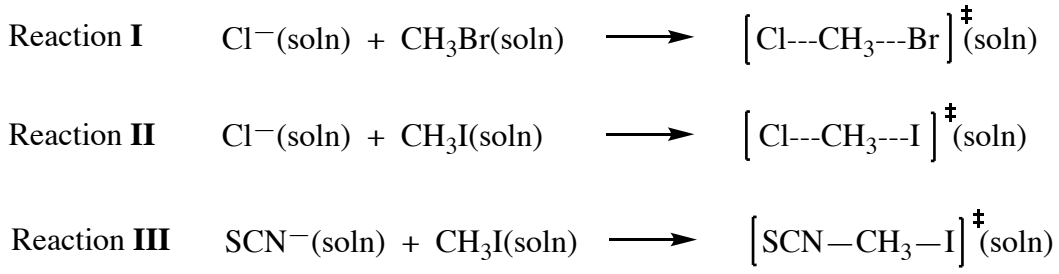

Figure 2. The three S_N2 reactions examined in this study.

Computational details

Gas phase calculations. All gas phase electronic structure calculations were performed using the Gaussian16[36] and ORCA (versions 4.2.0 and 5.0.1)[37-38] programs. Gas phase geometries and harmonic frequencies were calculated at the M06-2X/6-31+G(d) level of theory;[39] the LANL2DZdp[40] basis set was used for iodine. Harmonic frequency calculations confirm that the transition state structures were first order saddle points, and intrinsic reaction coordinate (IRC) calculations were performed to verify that they connect the correct reactants and products.

Single point calculations were performed using various levels of wavefunction and density functional theory (DFT) methods. This includes CCSD(T),[41-44] TightPNO-DLPNO-CCSD(T)[45-49] (both with CBS extrapolation) as well as DSD-PBEP86-D3(BJ),[50-51] B3LYP-D3(BJ),[52-54] M06-2X,[39] ωB97M-V,[55] and ωB97X-V[56] in conjunction with the def2-SVP, def2-TZVPP, def2-QZVPP, ma-def2-SVP, ma-def2-TZVPP, ma-def2-QZVPP, cc-pVTZ, cc-pVQZ, and aug-cc-pVTZ and aug-cc-pVQZ basis sets. For double-hybrid DFT and wavefunction calculations with Dunning's basis sets, the

$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$ orbitals (28 core electrons) of bromine and iodine were frozen, with the SK-MCDHF-RSC for iodine-containing systems. For the Aldrich's basis sets the $1s^2 2s^2 2p^6 3s^2 3p^6$ orbitals (18 core electrons) of bromine were frozen, and the $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$ orbitals (28 core electrons) of iodine were frozen, with the def2-ECP used for iodine-containing systems. CBS extrapolations were performed using the aug-cc-pVXZ (X = T, Q) basis sets;[57] the pseudopotential aug-cc-pVXZ-PP and ECP SK-MCDHF-RSC was used for the iodine containing systems.[57-58]

Continuum solvent calculations. The SMD, COSMO-RS and SM12 continuum solvent calculations were performed in the Gaussian16, ADF (version 2020.102) and Q-Chem (version 5.3.1) programs respectively, and applied at the levels of theory that are consistent with their parametrisation.[59] Gas phase geometries of each species were reoptimized using the SMD model (for the eight solvents) at the M06-2X/6-31+G(d) level of theory with the LANL2DZdp basis set for iodine atoms. Single point SM12 calculations were performed on the SMD optimised geometries at the M06-2X/6-31+G(d) level of theory; the MIDI! basis set was used for iodine atoms.[7] Single point COSMO-RS single point calculations were performed on SMD optimised geometries at the BP/TZP level of theory, with the ZORA/TZ2GP basis set for iodine. As these are fixed concentration free energies of solvation, a standard state correction (-1.89 kcal mol^{-1}) is applied when they are combined with the gas phase barriers (standard state 1 atm) to yield a standard state (1 M) solution phase Gibbs free energy barrier.

MM explicit solvent simulations. The molecular mechanics solvation free energy calculations were obtained using the CHARMM General Force Field (CGenFF)[60-61] to simulate the solute and organic solvents; the TIP3P model was used for water. Simulations were performed using NAMD 2.14[62] with periodic boundary conditions (PBC) at constant temperature (298.15 K) with the Langevin algorithm at a pressure of 1.0 bar using the Nose−Hoover Langevin Piston method (time step = 2.0 fs). All covalent bonds involving hydrogens were kept rigid with the RATTLE algorithm. The Particle Mesh Ewald (PME) algorithm was applied for long-range electrostatic interactions with a 12 Å distance cut-off. Each solute was fixed at their M06-2X/6-31+G(d) – SMD optimised geometries

(LANL2DZdp for iodine), with fixed atom forces turned on, and embedded in a 4 nm cubic periodic box with the appropriate number of solvent molecules to reproduce the density of that solvent. No counter ion was added. For this reason, only the MM non-bonded parameters (vDW and atomic charges) were needed for the solutes in these simulations. Intramolecular contributions to the solvation free energies were not considered as they expected to be negligible for the small rigid solutes considered in this work.

For solutes where there are no CGenFF parameters (*i.e.* transition states and SCN^- anion) or the parameters have high penalty scores (CH_3I and SCN^-), the Force Field Toolkit Plugin[63] in VMD[64] was used to determine their atomic charges. These charges are optimised to reproduce monohydrate interaction energies and dipole moment calculated at the HF/6-31+G(d) and MP2/6-31+G(d) levels of theory respectively with LANL2DZdp basis set for iodine. The QM interaction energies were scaled by 1.16 for neutral species, and 1.0 for charged species. All QM optimised distances were offset by -0.2 Å. Additionally, different atomic charge calculation schemes were used for the solutes, namely restrained and unrestrained electrostatic potential (RESP and ESP) charges obtained in the gas phase and in the SMD continuum solvent model at the M06-2X/6-31+G(d) level of theory with LANL2DZdp basis set for iodine. Overall, the use of CGenFF Lennard Jones parameters together with atomic charges optimised to reproduce monohydrate interaction energies produced the best results (see Figure 8 and Table S5 in the Supporting Information). Hence, all MM explicit solvent simulation results reported in Figures 1-3 employ this combination which is labelled CGenFF. Alchemical free energy perturbation (FEP)[65] was performed using 20 evenly distributed λ-windows, with 100,000 equilibration steps and 250,000 production steps per window. To prevent numerical instabilities as atoms were created or destroyed, a soft-core potential was used with alchVdWShiftCoeff = 5.0. These were carried out under the NPT ensemble (25 ^oC and 1 atm) and periodic boundary conditions. Both forward and reverse FEP was performed to check that hysteresis was smaller than 1 kcal mol^{-1}. All MM solvation free energies were obtained using the Bennett Acceptance Ratio (BAR) method.[66]

For ionic solutes, the solvation free energies further include the contribution due to the interaction of the solute with periodic images and also the surface potential of the solvent. The former is recovered through the use of the keyword "Alchdecouple OFF" in the NAMD simulations, whilst the correction due to the surface potential of the solvent was excluded as this contribution would cancel out in the calculation of the barrier. All solvation free energies (implicit and explicit solvent models) reported in this study correspond to fixed concentration intrinsic solvation free energies.

QM/MM explicit solvent simulations. The QM/MM umbrella sampling simulations were performed using ORCA/NAMD interface[67] and the weighted histogram analysis method (WHAM) was used to extract the free energy profile. The reaction coordinate was defined as $R = r(\text{Nu-C}) - r(\text{C-LG})$ (Nu=nucleophile and LG=leaving group) and there is a total of 30 windows spaced 0.1 Å apart. For each window, a classical MD simulation was performed with a harmonic biasing potential and fixed force constant (50 kcal mol^{-1} Å^{-2}) on the reaction coordinate. The starting structures for the QM/MM MD simulations were taken from the last frame of each of the classical simulation trajectories. During the QM/MM MD simulations, the solutes were described by the ωB97M-V/ma-def2-SVP level of theory whilst the solvent molecules were described by the TIP3P or CGenFF force fields. For each window, the force constant was set to be 200 kcal mol^{-1} Å^{-2} at a specific value of the reaction coordinate and the system was equilibrated for 5 ps followed by 5-10 ps production run, depending on the solvent. Checks were made to ensure that the PMF free energy profiles have converged with respect to the length of the trajectory – see Figure S1 for example. To convert the PMF to a 1 M standard state free energy barrier, a correction $\Delta G(\nu_0 \rightarrow \nu_{cage})$ which corresponds to approximately to the change of effective concentration of 1 M to ~55 M is applied as recommended by Warshel et al. This correction is about 2.4 kcal mol^{-1} at 298 K.[68]

Transition state theory. The rate constants were determined using the transition state theory expression shown in Equation (2).

(2)

Here, κ(T) is the tunnelling correction (assumed to be unity here), T is the temperature at 298.15 K, k_B and h are Boltzmann and Planck's constants, c^o is the standard state concentration 1 mol L^{-1}, m is the molecularity of the reaction (m=2 for S_N2 reactions) and is the Gibbs free energy of activation.

Results and discussion

Gas phase barriers. To accurately determine rate constants through a thermodynamic cycle, accurate gas phase reaction barriers must be calculated. Due to the relatively small size of the systems of interest, it is possible to use the 'gold standard' CCSD(T) extrapolated to the complete basis set (CBS) limit. However, the computational cost of CCSD(T) scales steeply with the size of the system so it is generally not applicable for larger systems. Consequently, it is of interest to understand the ability of approximate methods to reproduce 'gold standard' results. A range of DFT methods (DSD-PBEP96-D3(BJ), M06-2X, B3LYP-D3(BJ), ωB97X-V and ωB97M-V) and the wavefunction method TightPNO-DLPNO-CCSD(T) were investigated with a variety of basis sets. Figure 3 summarises the mean signed errors of selected levels of theory against the CCSD(T)/CBS//M06-2X/6-31+G(d) (LANL2Dzp for iodine) gas phase barriers of the three S_N2 reactions. (See Table S1 in the Supporting Information for complete set of data).

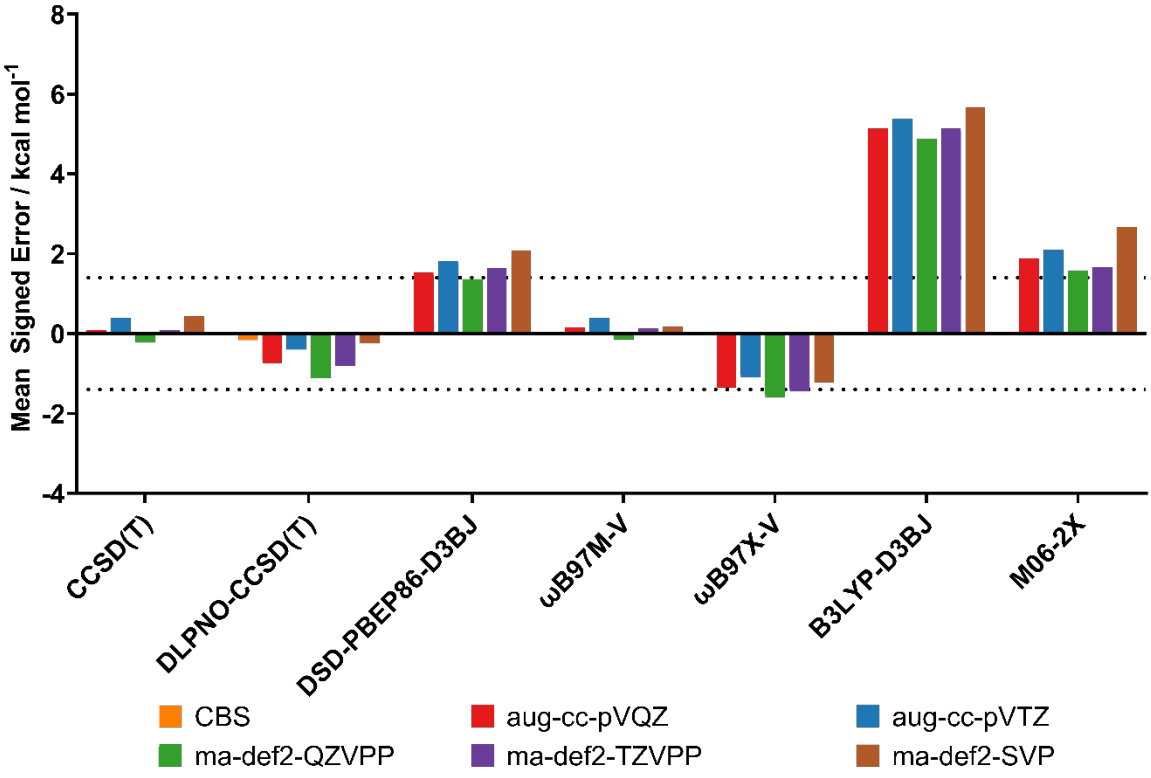

Figure 3. Mean signed deviation (MSD) of predicted gas phase activation free energy barriers at different levels of theory relative to CCSD(T)/CBS values. All calculations are based on M06-2X/6-31+G(d) (LANL2Dzp for iodine) optimised geometries and thermal corrections. The error bars represent the standard deviation over the three reactions.

For these reactions, it appears that the predicted barriers are relatively insensitive to the size of the basis sets for both CCSD(T) and DFT calculations. However, the inclusion of diffuse functions is important, and can change the calculated energy barrier by more than 2 kcal mol^{-1}. The ma-def2-TZVPP represents a cost-effective basis set, and in most cases the ma-def2-SVP basis set also gives good results. The DLPNO-CCSD(T) method accurately reproduced canonical CCSD(T) barriers to within 1.4 kcal mol^{-1} (chemical accuracy) for all three systems; however, it should be noted that the error in the DLPNO approximation is also size-extensive and very TightPNO threshold settings may be needed if it was applied to larger systems.[69] Of all the DFT methods using the ma-def2-TZVPP basis set, the range-separated hybrid functional ωB97M-V is in best agreement with CCSD(T)/CBS with a mean absolute deviation (MAD) of 0.5 kcal mol^{-1} while the other range separated functionals ωB97X-V displayed larger errors of 1.6 kcal mol^{-1}. Interestingly, the B3LYP-D3(BJ)

functional performed the worst with a MAD close to 5 kcal mol^{-1}, which is consistent with previous assessment studies.[70] Both DSD-PBEP86-D3(BJ) and M06-2X show reasonable agreement with CCSD(T)/CBS with a MAD of 1.4 kcal mol^{-1} and 1.6 kcal mol^{-1} respectively.

QM continuum solvent prediction of absolute rate constants. The CCSD(T)/CBS//M06-2X/6-31+G(d) gas phase barriers were combined with solvation free energies obtained from the SMD, SM12, COSMO-RS continuum solvent models and CGenFF/TIP3P MM explicit solvent model to obtain the solution phase barriers. The SMD optimised geometries of the transition state were nearly identical in the different solvents (see Table S3) though there are noticeable differences compared to the gas phase optimised geometry. Specifically, comparison of the gas phase and SMD (water) transition state geometries for the S_N2 reaction between the chloride with bromo- and iodomethanes (Reactions I and II; Figure 2) indicates a significant lengthening of the carbon-leaving group distance by as much as 0.07 Å (See Figure 4). This is to be expected since the ΔG of these reactions is largely determined by the relative solvation free energies of nucleophile (Cl^-; hydration free energy = -74.6 kcal mol^{-1}) and the leaving group (Br^- and I^-; hydration free energies = -68.6 and -59.9 kcal mol^{-1} respectively).[71] In accordance with Hammond's postulate,[72] the ΔG of these reactions should become increasingly positive (or less negative) in solvents compared to the gas phase, and the transition state becomes more product-like. In contrast, the S_N2 reaction between thiocyanate and iodomethane is accompanied by a very slight shortening of the carbon-leaving group distance (by at most 0.02 Å) which suggests that iodide and thiocyanate have similar solvation free energies, although there is no experimental data for the latter to confirm this.

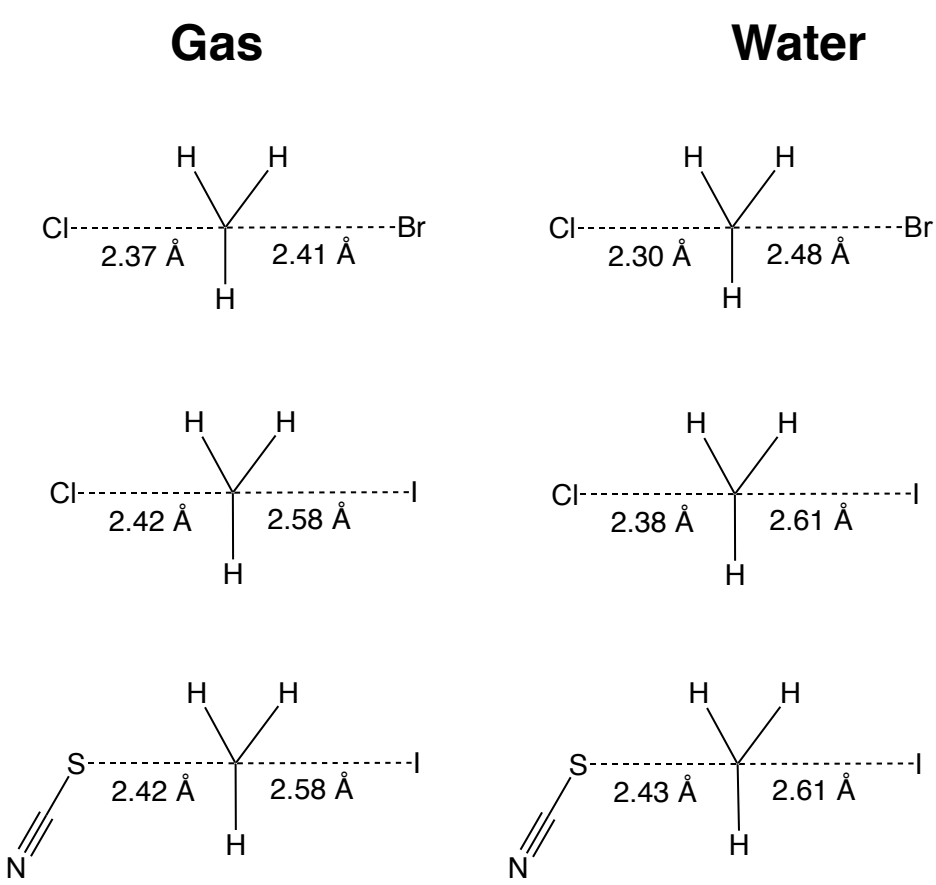

Figure 4. Bond distances for the transition state of the reaction between bromomethane and the chloride anion in the gas phase (left) and water (right) based on M06-2X/6-31+G(d) optimised geometries. A full list of bond distances in the eight solvents can be found in Table S3.

The signed errors in calculated solution phase rate constants (log *k*) are summarised in Figures 5 to 7; the log *k* values are provided in Table 1. For the SM12 models, three different charge schemes (Merz-Kollman, CHelpG and CM5) were compared. It appears that the SM12-MK scheme yielded the best agreement with experiment (see Table S4); all subsequent discussions of the SM12 model are based on this charge scheme. Overall, the ADF-COSMO-RS model is the best performing model with the lowest mean and maximum absolute deviations (MAD and AD_{max}) of 1.5 and 3.2 log units, respectively. The SM12-MK performed comparably well with a MAD of 1.6 log units; however, its performance is much more system-dependent. Notably, the SM12 model incurred errors of 5 or more log units in predicted rate coefficients for the S_N2 reaction between chloride anion and bromomethane (Reaction I) while its errors are significantly smaller (by three-fold or more) for the other two

reactions involving iodomethane. For example, the signed errors in the SM12 predicted rate constants for Reactions I and II in water are 5.5 and 0.1 log units respectively.

To better understand the origin of the inconsistent performance of the SM12 models, we used the thermodynamic cycle in Figure 1 together with the CCSD(T)/CBS//M06-2X/6-31+G(d) gas phase Gibbs free energy barrier and the experimental solvation free energies of bromomethane and iodomethane to indirectly estimate the "experimental" hydration free energies of the S_N2 transition states. This data is presented in Table 2. Interestingly, the SM12 model over-estimates the experimental solvation free energies of the chloride anion as well as its S_N2 transition state with iodomethane by about 7 and 9 kcal mol^{-1} respectively, resulting in significant error cancellation. On the other hand, the hydration free energy of the corresponding transition state with bromomethane (TS I) was accurately predicted by the SM12 model to within 1 kcal mol^{-1} of its experimental value. This inconsistent performance explains why SM12 incurred much higher errors for the reaction involving bromomethane.

Of the continuum solvent models considered, SMD was the worst performing model with an overall MAD of 2.9 log units. The large MAD is mostly due to the significantly higher errors incurred by this model for the predicted rate constants in formamide. In particular, the deviation from experiment is 7.6 log units for the reaction of chloride and iodomethane (Reaction II) whilst the SM12 model and ADF-COSMO-RS displayed errors of less than 1.5 log units for this reaction. On the basis of these results, the ADF-COSMO-RS solvent model is preferred over the Minnesota solvation models for the prediction of the *absolute* rate constants of simple S_N2 reactions in aqueous and organic solvents. This conclusion is also consistent with previous work which showed that COSMO-RS performed significantly better than SMD and SM8 models in the prediction of the free energy barriers of S_N2 reactions in methanol.[73]

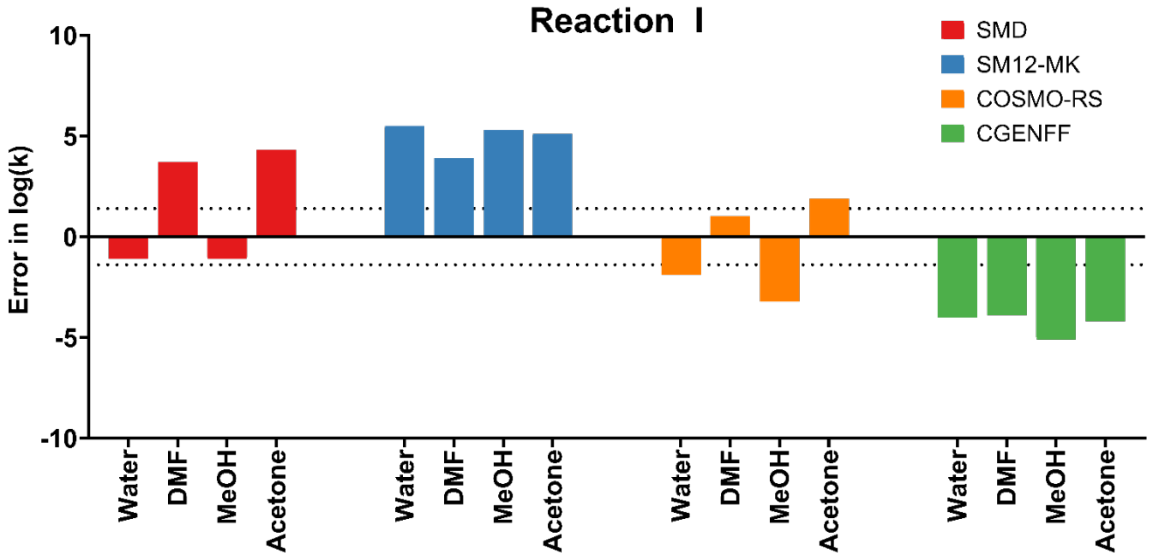

Figure 5. Signed error in predicted rate constants of Reaction I in various solvents. DMF is dimethylformamide. The CCSD(T)/CBS//M06-2X/6-31+G(d) gas phase barrier was used in these calculations. Horizontal dashed line represents an error of 1 log unit.

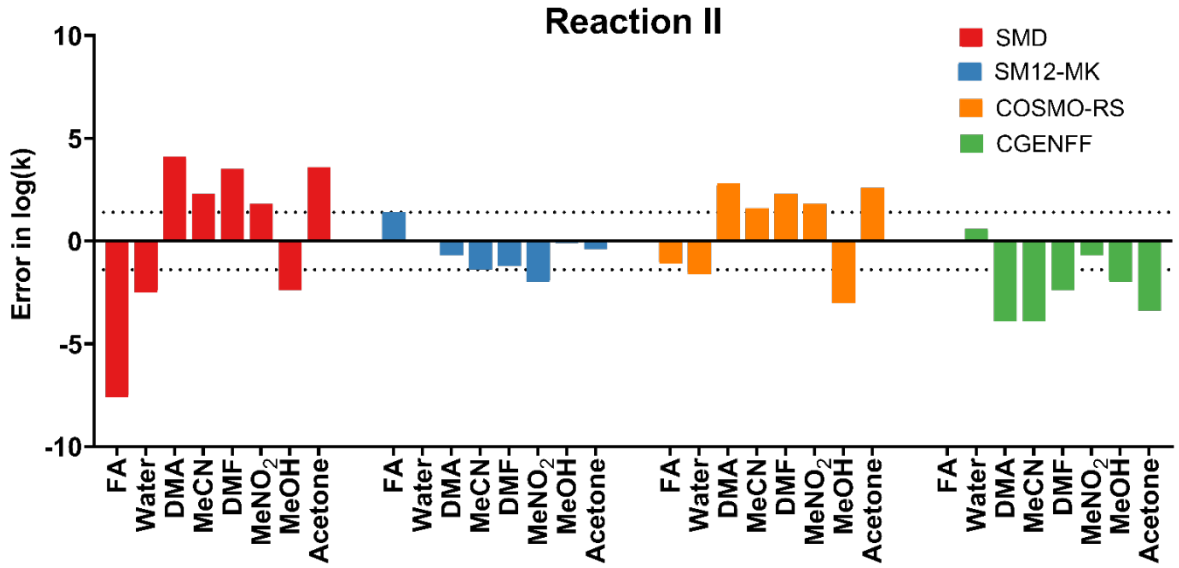

Figure 6. Signed error in predicted rate constants of Reaction II in various solvents. FA is formamide, DMA is dimethylacetamide, and DMF is dimethylformamide. The CCSD(T)/CBS//M06-2X/6-31+G(d) gas phase barrier was used in these calculations. Horizontal dashed line represents an error of 1 log unit.

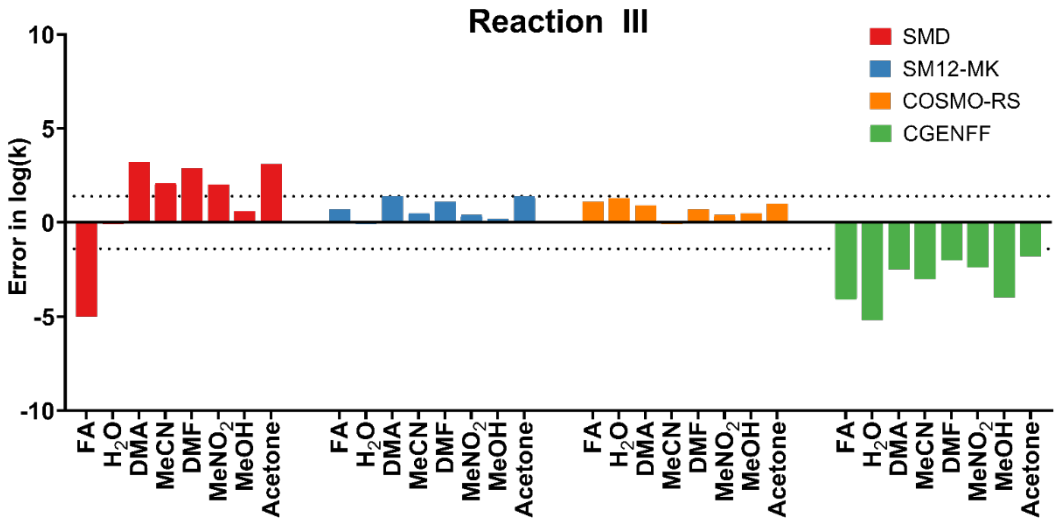

Figure 7. Signed error in predicted rate constants of Reaction III in various solvents. FA is formamide DMA is dimethylacetamide and DMF is dimethylformamide. The CCSD(T)/CBS//M06-2X/6-31+G(d) gas phase barrier was used in these calculations. Horizontal dashed line represents an error of 1 log unit.

Table 1. Predicted log *k* values in different solvents.

Solvent	Experiment[74]	SMD[a]	SM12-MK[a]	ADF-COSMO-RS[b]	CGenFF[c]
			Reaction I		
H_2O	-5.1	-4.0	-10.6	-3.2	-1.1
DMF	-0.4	-4.1	-4.3	-1.4	3.5
MeOH	-5.2	-4.1	-10.5	-2.0	-0.1
Acetone	0.5	-3.8	-4.6	-1.4	4.7
			Reaction II		
FA	-4.3	3.3	-5.7	-3.2	-4.3
H_2O	-5.5	-2.9	-5.4	-3.9	-6.0
DMA	0.9	-3.2	1.6	-1.9	4.8
MeCN	-0.9	-3.2	0.5	-2.5	3.0
DMF	0.4	-3.1	1.6	-1.9	2.8
Nitromethane	-1.3	-3.1	0.7	-3.1	-0.6
MeOH	-5.5	-3.1	-5.4	-2.5	-3.5
Acetone	0.7	-2.9	1.1	-1.9	4.1
			Reaction III		
FA	-2.8	2.2	-3.5	-3.9	1.3
H_2O	-3.5	-3.6	-3.4	-4.8	1.7
DMA	-0.8	-4.0	-2.2	-1.7	1.7
MeCN	-2.0	-4.1	-2.5	-2.1	1.0
DMF	-1.1	-4.0	-2.2	-1.8	0.9
Nitromethane	-2.0	-4.0	-2.4	-2.4	0.4
MeOH	-3.3	-3.9	-3.5	-3.8	0.7
Acetone	-0.8	-3.9	-2.2	-1.8	1.0
MSD		0.9	1.0	0.5	-2.9
MAD		2.9	1.6	1.5	2.9

AD_{max}	7.6	5.5	3.2	5.2

[a] Calculated at the M06-2X/6-31+G(d) level of theory. [b] Calculated at the BP/TZVP level of theory. [c] MM simulations used charges optimised to reproduce HF/6-31+G* monohydrate interaction energies.

Table 2. Experimental and calculated hydration free energies of selected species in kcal mol^{-1}.

Species	Expt	SMD	SM12-MK	ADF-COSMO-RS	CGenFF
CH_3Br	-0.8[a]	-1.2	-1.4	-1.0	-0.3
CH_3I	-0.9[b]	-1.2	-2.7	-1.3	0.1
Cl^-	-74.7[c]	-65.1	-81.6	-70.8	-80.4[e]
Cl^- + CH_3Br (TS I)	-52.7[d]	-45.0	-52.7	-51.7	-63.3[e]
Cl^- + CH_3I (TS II)	-51.7[d]	-45.8	-60.4	-50.3	-55.6[e]

[a] Experimental value from Reference [8]. [b] Experimental value from Reference [75]. [c] Experimental value from Reference [76]. [d]Estimated from Equation (1) using CCSD(T)/CBS gas phase barrier and experimental solvation free energies of Cl^-, CH_3Br and CH_3I. [e] Surface potential of TIP3P water (11.6 kcal mol^{-1})[77] was added as it has been reasoned that proton solvation free energy of -265.9 kcal mol^{-1} that is used to derive the experimental solvation free energy of Cl^- includes contributions from the surface potential.[78]

Absolute rate constants in MM explicit solvents. In addition to continuum solvent models, we have further investigated the use of molecular mechanics (MM) explicit solvent models to calculate ΔG_S and $\Delta\Delta G_S$ (see Computational Details). Recent work from the Sunoj and Harvey groups have shown that this approach was very effective in describing solvent effects on zwitterionic transition states and intermediates associated with the mechanism of the Morita-Baylis-Hillman reaction.[79-80] In that work, the authors fixed the geometry of the transition states and intermediates at the coordinates of the continuum solvent model optimised geometries so that bonded parameters for these species were not needed in the free energy perturbation simulations; the intramolecular thermal contributions to the free energy were recovered separately from continuum solvent calculations.[15, 81]

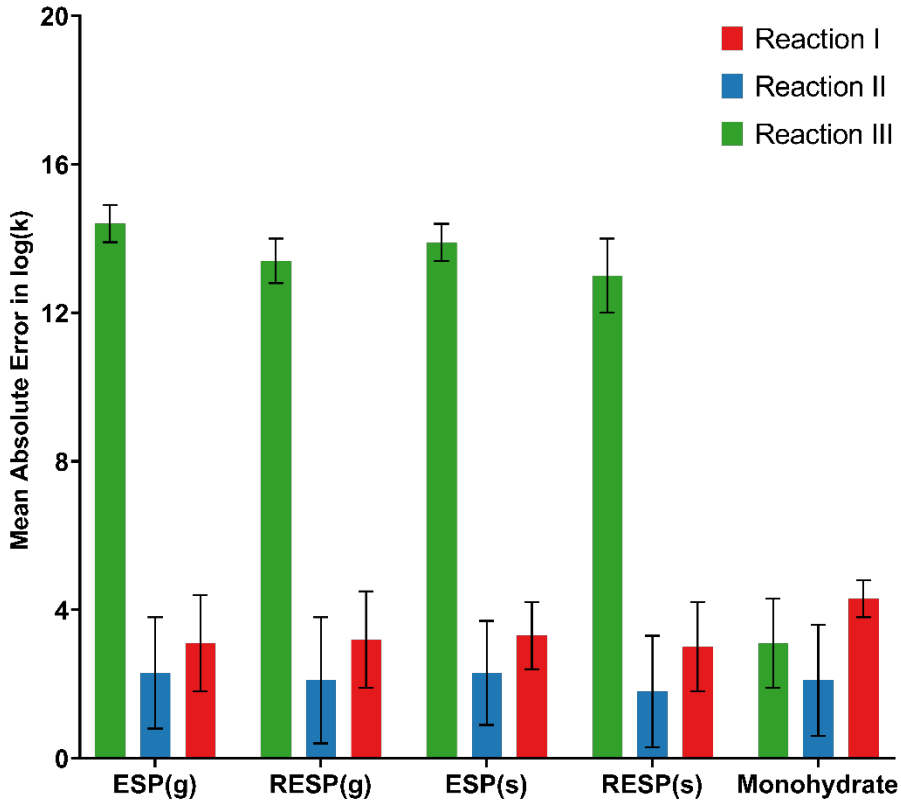

Figure 8 – Mean Absolute Errors using different charge schemes to define the solute charges in explicit solvent simulations. Error bars represent the standard deviation of errors in all relevant solvents.

A similar approach has been employed here, and the use of different atomic charge calculation schemes to describe the solute-solvent interaction has been examined. This includes gas phase and SMD derived restrained and unrestrained electrostatic potential (RESP and ESP) charges calculated with M06-2X/6-31+G(d), as well as charges optimised to reproduce monohydrate interaction energies and dipole moments at the more modest HF/6-31+G(d) level of theory, as is consistent with the parameterisation of the CHARMM force field (Figure 8). Solvation free energies calculated with these charge schemes can be found in Table S9. We have not included the contribution of the surface potential in the MM calculated solvation free energies of the ionic species as this will cancel out in the calculation of the reaction barriers.[11] Overall, the use of atomic charges optimised to reproduce monohydrate interaction energies (see Computational Details) appears to be most effective, and their results (labelled CGenFF) are also presented in Figures 5 to 7 and Table 1. Whilst the use of ESP

based charge schemes perform well for Reactions I and II, this approach incurs a large error for the third reaction. Comparatively, the charges optimised to reproduce monohydrate energies are more consistent for all three reactions. This charge scheme performed better than the direct QM charges presumably because the atomic charges are optimised to be consistent with the Lennard Jones parameters, resulting in error cancellation which is not seen when the two parameters are optimised separately. Interestingly, the MM explicit solvent model incurred the equal highest MAD value of 2.9 log units, which is two times larger than that of ADF-COSMO-RS. This is somewhat concerning since the MM explicit solvent models are typically 1-2 orders of magnitude computationally more demanding than a static QM continuum solvent model calculation. Nevertheless, the errors in these models are highly systematic for a given reaction. The CGenFF model consistently over-estimated the rate constants of all three reactions, as indicated by very similar MSD and MAD values. For example, the rate constant of Reaction I in water has an error of 4.0 log units. As shown in Table 2, this error is due to the significant over-estimation of the hydration free energy of TS I by 10.8 kcal mol^{-1} whilst CGenFF performed comparatively well for the hydration free energy of the Cl^- anion (overestimated by 5.7 kcal mol^{-1}). In contrast, for Reaction II in water, where the rate constant is predicted to within 0.6 log units of the experimental value, the hydration free energy of the transition state is overestimated by 4.8 kcal mol^{-1}, and therefore cancelation of errors between the Cl^- anion, the transition state, and the neutral species results in a highly accurate rate constant. The results indicate that MM explicit solvent models are not necessarily more accurate than continuum solvent models in predicting the rates of ionic S_N2 reactions. More extensive assessment of these models, particularly the choice of force fields and inclusion of explicit polarizability,[82] are needed to determine their broader performance in predicting solution phase barriers.

QM/MM potential of mean force simulations. Finally, we employed umbrella sampling explicit solvation simulations in conjunction with a QM/MM potential to predict the absolute rate constants of the three reactions. In this approach, the solution phase barrier is calculated directly from the

potential of mean force (PMF) along the reaction path which contrasts with the thermodynamic cycle-based approach discussed in the previous sections. For this reason, it is important to select a validated and economical level of theory to describe the QM region. Table S2 presents the gas phase barriers for the three reactions calculated using small basis set DFT calculations as well as various semi-empirical methods compared to CCSD(T)/CBS values. The data indicates that many of the semi-empirical methods such as HF-3c or B97-3c incur very large errors and are unsuitable for describing the QM region. On the other hand, the ωB97M-V/ma-def2-SVP calculation appears to be most accurate with an MAD of 0.9 kcal mol^{-1}. Due to the high cost of this QM treatment and the need for extensive configurational sampling, the umbrella sampling was only performed for approximately half of the reaction profile, *i.e.* from reactants to TS – see Computational Details.

Table 3 presents the 1 M standard state solution phase barriers obtained from QM/MM simulations. It is reassuring to see that the most physically rigorous model performed the best with a MAD and AD_{max} of 1.0 and 2.2 log units. The performance was particularly good for Reactions I and II where the deviation from experiment was typically less than 0.5 log units irrespective of the solvent. Reaction III has the highest MAD of 1.4 log units which is still smaller than the overall MAD of the best performing continuum solvent model (ADF-COSMO-RS). These results suggest that with the judicious selection of QM level of theory to treat the reaction centre, QM/MM PMF simulations can give highly accurate predictions.

Table 3. QM/MM rate constants. The QM level of theory is ωB97MV/ma-def2-SVP and CGenFF/TIP3P force field was used for the MM layer.

	log k(expt)	log *k*
Solvent	**Reaction I**	
H_2O	-5.1	-4.4
DMF	-0.4	-0.63
MeOH	-5.2	-5.7
Acetone	0.5	0.96
	Reaction II	
FA	-4.3	-4.6

H_2O	-5.5	-5.3
DMA	0.9	1.6
MeCN	-0.8	1.4
DMF	0.4	-0.1
Nitromethane	-1.3	-0.8
MeOH	-5.5	-4.2
Acetone	0.7	2.7
		Reaction III
FA	-2.8	-3.8
H_2O	-3.5	-4.8
DMA	-1.1	-2.5
MeCN	-2	-2.3
DMF	-1.1	-2.7
Nitromethane	-2	-4.1
MeOH	-3.3	-5.5
Acetone	-0.8	-2.2
MSD		-0.1
MAD		1.1
AD_{max}		2.2

Relative rate constants in different solvents. Based on the thermodynamic cycle presented in Figure 1, the relative rate constants for a solvent (X) relative to methanol are provided by the difference in ΔΔGs in the two solvents:

$$RT \ln\left(\frac{k_X}{k_{MeOH}}\right) = \Delta\Delta G_S^{MeOH} - \Delta\Delta G_S^X \quad (3)$$

Table 4 summarises the performance of the solvent models in predicting rate constants measured in various organic solvents relative to methanol. As shown, the SM12-MK model performed the best with the smallest MAD of 0.8 log units. This is followed by QM/MM and CGenFF explicit solvent models with MAD values of 1.0 and 1.4 log units respectively. On the other hand, the ADF-COSMO-RS which performed the best in the prediction of absolute rate constants has an MAD of 2.5 log units. This is not really surprising because the results in Figures 5 to 7 indicate that the errors in the SM12

and CGenFF models are highly systematic (MSD and MAD are similar magnitude; see Table 1) so there is very significant error cancellation in the calculation of relative rate constants in different solvents. As noted above, the performance of SMD is highly solvent-dependent, where it generally predicts aqueous phase rate constants in water and methanol accurately but incurs significantly higher errors in certain solvents such as formamide. As a result, this model fails to correctly predict shifts in rate constants between protic and aprotic solvents. For example, the SMD predicted rate constants for Reaction (I) in four protic and aprotic solvents are within 0.3 log units of each other whilst the experimental data indicates that there is approximately 6 log unit enhancement in rate going from water to acetone. As discussed elsewhere,[83] the failure of the SMD model in the prediction of classical rate enhancement going from protic to aprotic solvents is presumably due to the calculated solvation free energies dominated by the electrostatic term as many of the solvents have very high dielectric constants. Better discrimination between protic and aprotic solvents may be achieved by tuning the non-electrostatic contribution (e.g. cavitation free energy) through the introduction of solvent-specific atomic radii. These results indicate that the SM12-MK model can be used as a cost-effective approach for making predictions about solvent induced changes in nucleophilicity.

Table 4. Mean Absolute Deviation (MAD) of rate constants predicted in various solvents relative to MeOH for each reaction, $\log(k_x/k_M)$. Errors in log units.

Reaction	SMD	SM12-MK	ADF-COSMO-RS	CGenFF	QM/MM
I	3.4	0.6	3.6	1.0	0.8
II	4.7	1.0	4.2	1.6	1.3
III	2.3	0.6	0.4	1.4	0.8
MSD	2.1	-0.5	2.1	0.3	0.1
MAD	3.5	0.8	2.5	1.4	1.0
AD_{max}	6.4	1.9	5.8	2.6	2.1

In addition to the prediction of reaction rates in different solvents, the experimental data in Table 1 also reveals a number of interesting reactivity trends and it is of interest to see if the solvent models can correctly reproduce these trends:

(1) Comparison of Reactions I and II indicates that bromide is a better leaving group than iodide in protic solvents (water and methanol) while the opposite trend is observed in aprotic solvents (*N*,*N*-dimethylformamide and acetone). For example, the experimental log *k* values of Reaction I relative to Reaction II, i.e. $\log(k_I/k_{II}) = \log(k_I) - \log(k_{II})$, are 0.4 and -0.8 in water and DMF respectively.

This is presumably due to the preferential solvation of the S_N2 transition state of Reaction I in protic solvents (hydration free energy for TS in Reaction I is about 2.5 kcal mol^{-1} more negative than that of Reaction II; see Table 2) while the preference is reversed in aprotic solvents. As indicated from the data in Table 1, none of the implicit solvent models correctly reproduced this trend. The only exception is the QM/MM PMF simulation – see Table 3.

(2) Comparison of Reactions II and III indicates that chloride is a weaker nucleophile compared to thiocyanate in protic solvents (formamide, methanol and water) and their relative reactivity is reversed in aprotic solvents. This trend parallels the observed solvation free energies of chloride which is approximately 12 kcal mol^{-1} less exergonic in an aprotic solvent such as MeCN compared to water;[84] the weaker solvation of chloride in aprotic solvents renders it more reactive. Only the SM12 and CGenFF models correctly predicted the relative nucleophilicity of the chloride and thiocyanate anions in protic and aprotic solvents.

Conclusions

Solvents are one of the key parameters for the optimisation of reaction kinetics. Using three simple S_N2 reactions as test systems, it was shown that (with the exception of ADF-COSMO-RS) popular continuum solvent models such as SMD and SM12 failed to accurately predict *absolute* rate constants for these reactions in a variety of protic and aprotic solvents. Overall, the molecular mechanics explicit solvent model (CGenFF) displayed the largest error. It is of interest to examine other molecular mechanics force fields including polarisable force fields in the future. The performance of the solvent models is more favourable in the prediction of relative rate constants in different solvents

where the SM12 and CGenFF explicit solvent model performed best due to systematic error cancellation. These findings cast doubt over the ability of popular implicit solvent models to address chemical questions on reactivity (*e.g.* why is X^- a stronger nucleophile than Y^- in water) and reliably elucidate reaction mechanisms except in cases where there are very significant differences in reactivity, much higher than the intrinsic errors of the models reported in this work.

QM/MM potential of mean force simulations led to very accurate predictions of absolute rate constants provided that a validated QM level of theory was used to treat the QM atoms. However, the computational cost of these simulations are 3-4 orders of magnitude more expensive than an implicit solvent calculation. As indicated in this work, such expensive treatments are probably unnecessary if the aim was to screen solvents that optimise a reaction rate where the SM12 model performed very well. Nevertheless, recent developments in machine-learning assisted free energy simulations[25, 85] should significantly accelerate QM/MM simulations and facilitate their wide use in studying reaction mechanisms.

Supporting Information

Calculated gas phase and solution phase barriers, solvation free energies, electronic energies and cartesian coordinates of DFT optimized geometries.

Acknowledgments

J.H. thanks the Australian Research Council for financial support (Grant No. DP210102698) and the Australian National Computational Infrastructure, UNSW and Pawsey Supercomputer Centre for generous allocation of computational resources.

References

1. O'Hagan, D.; Schmidberger, J. W.,Enzymes That Catalyse Sn2 Reaction Mechanisms. *Nat. Prod. Rep.* **2010,** *27*, 900-918.
2. Williamson, A.,Xlv. Theory of Ætherification. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1850,** *37*, 350-356.

3. Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L.,Theoretical Examination of the Sn2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1985,** *107*, 154-163.
4. Tomasi, J.; Mennucci, B.; Cammi, R.,Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105*, 2999-3094.
5. Klamt, A.; Schurmann, G.,Cosmo: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799-805.
6. Barone, V.; Cossi, M.,Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998,** *102*, 1995-2001.
7. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.,Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013,** *9*, 609-620.
8. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.,Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009,** *113*, 6378-6396.
9. Pye, C.; Ziegler, T.; Lenthe, E.; Louwen, J.,An Implementation of the Conductor-Like Screening Model of Solvation within the Amsterdam Density Functional Package — Part II. COSMO for Real Solvents1. *Can. J. Chem.* **2009,** *87*, 790-797.
10. Klamt, A.,Conductor-Like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995,** *99*, 2224-2235.
11. Ho, J.,Predicting pKa in Implicit Solvents: Current Status and Future Directions. *Aust. J. Chem.* **2014,** *67*, 1441-1460.
12. Pliego, J. R.; Riveros, J. M.,The Cluster−Continuum Model for the Calculation of the Solvation Free Energy of Ionic Species. *J. Phys. Chem. A* **2001,** *105*, 7241-7247.
13. Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A.,Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *J. Phys. Chem. B* **2008,** *112*, 9709-9719.
14. Roy, D.; Kovalenko, A.,Performance of 3D-RISM-KH in Predicting Hydration Free Energy: Effect of Solute Parameters. *J. Phys. Chem. A* **2019,** *123*, 4087-4093.
15. Ho, J.; Ertem, M. Z.,Calculating Free Energy Changes in Continuum Solvation Models. *J. Phys. Chem. B* **2016,** *120*, 1319-1329.
16. Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D.,Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **2017,** *13*, 1034-1043.
17. Mobley, D. L.; Guthrie, J. P.,Freesolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput. Aided Mol. Des.* **2014,** *28*, 711-720.
18. Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L.,S_N2 Reaction Profiles in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1984,** *106*, 3049-3050.
19. Chandrasekhar, J.; Jorgensen, W. L.,Energy Profile for a Nonconcerted S_N2 Reaction in Solution. *J. Am. Chem. Soc.* **1985,** *107*, 2974-2975.
20. Chen, J.; Shao, Y.; Ho, J.,Are Explicit Solvent Models More Accurate Than Implicit Solvent Models? A Case Study on the Menschutkin Reaction. *J. Phys. Chem. A* **2019,** *123*, 5580-5589.
21. Ali, H. S.; Higham, J.; De Visser, S. P.; Henchman, R. H.,Comparison of Free-Energy Methods to Calculate the Barriers for the Nucleophilic Substitution of Alkyl Halides by Hydroxide. *J. Phys. Chem. B* **2020,** *124*, 6835-6842.
22. Blumberger, J.,Free Energies for Biological Electron Transfer from Qm/Mm Calculation: Method, Application and Critical Assessment. *Phys. Chem. Chem. Phys* **2008,** *10*, 5651-5667.
23. Uddin, N.; Choi, T. H.; Choi, C. H.,Direct Absolute Pka Predictions and Proton Transfer Mechanisms of Small Molecules in Aqueous Solution by QM/MM-MD. *J. Phys. Chem. B* **2013,** *117*, 6269-6275.

24. Tummanapelli, A. K.; Vasudevan, S.,Ab Initio Molecular Dynamics Simulations of Amino Acids in Aqueous Solutions: Estimating pKa Values from Metadynamics Sampling. *J. Phys. Chem. B* **2015,** *119*, 12249-12255.
25. Pan, X.; Li, P.; Ho, J.; Pu, J.; Mei, Y.; Shao, Y.,Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanical Accuracy Via a Semi-Empirical Reference Potential. II. Recalibrating Semi-Empirical Parameters with Force Matching. *Phys. Chem. Chem. Phys.* **2019,** *21*, 20595-20605.
26. Acevedo, O.; Jorgensen, W. L.,Advances in Quantum and Molecular Mechanical (Qm/Mm) Simulations for Organic and Enzymatic Reactions. *Acc. Chem. Res.* **2010,** *43*, 142-151.
27. Chen, X.; Regan, C. K.; Craig, S. L.; Krenske, E. H.; Houk, K. N.; Jorgensen, W. L.; Brauman, J. I.,Steric and Solvation Effects in Ionic S_N2 Reactions. *J. Am. Chem. Soc.* **2009,** *131*, 16162-16170.
28. Tirado-Rives, J.; Jorgensen, W. L.,QM/MM Calculations for the $Cl^- + CH_3Cl$ S_N2 Reaction in Water Using Cm5 Charges and Density Functional Theory. *J. Phys. Chem. A* **2019,** *123*, 5713-5717.
29. Vayner, G.; Houk, K. N.; Jorgensen, W. L.; Brauman, J. I.,Steric Retardation of S_N2 Reactions in the Gas Phase and Solution. *J. Am. Chem. Soc.* **2004,** *126*, 9054-9058.
30. Higashi, M.; Truhlar, D. G.,Combined Electrostatically Embedded Multiconfiguration Molecular Mechanics and Molecular Mechanical Method: Application to Molecular Dynamics Simulation of a Chemical Reaction in Aqueous Solution with Hybrid Density Functional Theory. *J. Chem. Theory Comput.* **2008,** *4*, 1032-1039.
31. Stanton, C. L.; Kuo, I. F. W.; Mundy, C. J.; Laino, T.; Houk, K. N.,Qm/Mm Metadynamics Study of the Direct Decarboxylation Mechanism for Orotidine-5'-Monophosphate Decarboxylase Using Two Different QM Regions: Acceleration Too Small to Explain Rate of Enzyme Catalysis. *J. Phys. Chem. B* **2007,** *111*, 12573-12581.
32. Chen, J.; Kato, J.; Harper, J. B.; Shao, Y.; Ho, J.,On the Accuracy of QM/MM Models: A Systematic Study of Intramolecular Proton Transfer Reactions of Amino Acids in Water. *J. Phys. Chem. B* **2021,** *125*, 9304-9316.
33. Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J.,How Large Should the Qm Region Be in QM/MM Calculations? The Case of Catechol-O-Methyltransferase. *J. Phys. Chem. B* **2016,** *120*, 11381-11394.
34. Fehér, P. P.; Stirling, A.,Assessment of Reactivities with Explicit and Implicit Solvent Models: Qm/Mm and Gas-Phase Evaluation of Three Different Ag-Catalysed Furan Ring Formation Routes. *New J. Chem.* **2019,** *43*, 15706-15713.
35. König, G.; Mei, Y.; Pickard, F. C.; Simmonett, A. C.; Miller, B. T.; Herbert, J. M.; Woodcock, H. L.; Brooks, B. R.; Shao, Y.,Computation of Hydration Free Energies Using the Multiple Environment Single System Quantum Mechanical/Molecular Mechanical Method. *J. Chem. Theory Comput.* **2016,** *12*, 332-344.
36. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
37. Neese, F.,The Orca Program System. *WIRES COMPUT MOL SCI* **2012,** *2*, 73-78.

38. Neese, F.,Software Update: The Orca Program System, Version 4.0. *WIREs Computational Molecular Science* **2018,** *8*, e1327.
39. Zhao, Y.; Truhlar, D. G.,The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008,** *120*, 215-241.
40. Hay, P. J.; Wadt, W. R.,Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for K to Au Including the Outermost Core Orbitals. *J. Chem. Phys.* **1985,** *82*, 299-310.
41. Čížek, J.,On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966,** *45*, 4256-4266.
42. Sinanoğlu, O.; Brueckner, K. A., *Three Approaches to Electron Correlation in Atoms: A Chemistry-Physics Interface*. Yale University Press: 1970.
43. Purvis III, G. D.; Bartlett, R. J.,A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982,** *76*, 1910-1918.
44. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M.,A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989,** *157*, 479-483.
45. Riplinger, C.; Neese, F.,An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013,** *138*, 034106.
46. Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F.,Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013,** *139*, 134101.
47. Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F.,Sparse Maps—a Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Phys.* **2016,** *144*, 024109.
48. Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F.,A New near-Linear Scaling, Efficient and Accurate, Open-Shell Domain-Based Local Pair Natural Orbital Coupled Cluster Singles and Doubles Theory. *J. Chem. Phys.* **2017,** *146*, 164105.
49. Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F.,Communication: An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018,** *148*, 011101.
50. Kozuch, S.; Martin, J. M. L.,DSD-PBEP86: In Search of the Best Double-Hybrid DFT with Spin-Component Scaled MP2 and Dispersion Corrections. *Phys. Chem. Chem. Phys* **2011,** *13*, 20104.
51. Kozuch, S.; Martin, J. M. L.,Spin-Component-Scaled Double Hybrids: An Extensive Search for the Best Fifth-Rung Functionals Blending Dft and Perturbation Theory. *J. Comput. Chem.* **2013,** *34*, 2327-2344.
52. Lee, C.; Yang, W.; Parr, R. G.,Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988,** *37*, 785.
53. Becke, A. D.,Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.
54. Grimme, S.; Ehrlich, S.; Goerigk, L.,Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.
55. Mardirossian, N.; Head-Gordon, M., ωb97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-Gga Density Functional with Vv10 Nonlocal Correlation. *J. Chem. Phys.* **2016,** *144*, 214110.
56. Mardirossian, N.; Head-Gordon, M., ωb97X-V A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys* **2014,** *16*, 9904-9924.

57. Neese, F.; Valeev, E. F.,Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated Ab Initio Methods? *J. Chem. Theory Comput.* **2011,** *7*, 33-43.
58. Helgaker, T.; Klopper, W.; Koch, H.; Noga, J.,Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997,** *106*, 9639-9646.
59. Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J.,On the Performance of Continuum Solvation Methods. *Acc. Chem. Res.* **2009,** *42*, 489-492.
60. Vanommeslaeghe, K.; MacKerell, A. D., Jr.,Automation of the Charmm General Force Field (CGenff) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012,** *52*, 3144-3154.
61. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr.,Automation of the Charmm General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012,** *52*, 3155-3168.
62. Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E.,Scalable Molecular Dynamics on Cpu and Gpu Architectures with Namd. *J. Chem. Phys.* **2020,** *153*, 044130.
63. Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C.,Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013,** *34*, 2757-2770.
64. Humphrey, W.; Dalke, A.; Schulten, K.,Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14*, 33-38.
65. Zwanzig, R. W.,High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954,** *22*, 1420-1426.
66. Bennett, C. H.,Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976,** *22*, 245-268.
67. Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z.,Namd Goes Quantum: An Integrative Suite for Hybrid Simulations. *Nature Methods* **2018,** *15*, 351-354.
68. Shurki, A.; Štrajbl, M.; Villà, J.; Warshel, A.,How Much Do Enzymes Really Gain by Restraining Their Reacting Fragments? *JACS* **2002,** *124*, 4097-4107.
69. Sandler, I.; Chen, J.; Taylor, M.; Sharma, S.; Ho, J.,Accuracy of Dlpno-Ccsd(T): Effect of Basis Set and System Size. *J. Phys. Chem. A* **2021,** *125*, 1553-1563.
70. Bento, A. P.; Solà, M.; Bickelhaupt, F. M.,Ab Initio and Dft Benchmark Study for Nucleophilic Substitution at Carbon (Sn2@C) and Silicon (Sn2@Si). *J. Comput. Chem.* **2005,** *26*, 1497-1504.
71. Pliego Jr, J. R.; Riveros, J. M.,Gibbs Energy of Solvation of Organic Ions in Aqueous and Dimethyl Sulfoxide Solutions. *Phys. Chem. Chem. Phys* **2002,** *4*, 1622-1627.
72. Hammond, G. S.,A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955,** *77*, 334-338.
73. Silva, N. M.; Deglmann, P.; Pliego, J. R.,Cmirs Solvation Model for Methanol: Parametrization, Testing, and Comparison with SMD, SM8, and COSMO-RS. *J. Phys. Chem. B* **2016,** *120*, 12660-12668.
74. Alexander, R.; Ko, E.; Parker, A.; Broxton, T.,Solvation of Ions. Xiv. Protic-Dipolar Aprotic Solvent Effects on Rates of Bimolecular Reactions. Solvent Activity Coefficients of Reactants and Transition States at 25°. *J. Am. Chem. Soc* **1968,** *90*, 5049-5069.
75. Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D.,Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006,** *2*, 128-139.
76. Pliego, J. R.; Riveros, J. M.,New Values for the Absolute Solvation Free Energy of Univalent Ions in Aqueous Solution. *Chem. Phys. Lett.* **2000,** *332*, 597-602.

77. Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D.,A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006,** *418*, 245-249.
78. Carvalho, N. F.; Pliego, J. R.,Cluster-Continuum Quasichemical Theory Calculation of the Lithium Ion Solvation in Water, Acetonitrile and Dimethyl Sulfoxide: An Absolute Single-Ion Solvation Free Energy Scale. *Phys. Chem. Chem. Phys.* **2015,** *17*, 26745-26755.
79. Morita, K.-i.; Suzuki, Z.; Hirose, H.,A Tertiary Phosphine-Catalyzed Reaction of Acrylic Compounds with Aldehydes. *Bull. Chem. Soc. Jpn.* **1968,** *41*, 2815-2815.
80. Baylis, A.; Hillman, M. In *German Patent 2155113, 1972*, Chem. Abstr, 1972; p 34174q.
81. Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.,Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B* **2011,** *115*, 14556-14562.
82. Inakollu, V. S. S.; Geerke, D. P.; Rowley, C. N.; Yu, H.,Polarisable Force Fields: What Do They Add in Biomolecular Simulations? *Current Opinion in Structural Biology* **2020,** *61*, 182-190.
83. Miguel, E. L.; Santos, C. I.; Silva, C. M.; Pliego Jr, J. R.,How Accurate Is the SMD Model for Predicting Free Energy Barriers for Nucleophilic Substitution Reactions in Polar Protic and Dipolar Aprotic Solvents? *J. Braz. Chem. Soc.* **2016,** *27*, 2055-2061.
84. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G.,Aqueous Solvation Free Energies of Ions and Ion−Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *J. Phys. Chem. B* **2006,** *110*, 16066-16081.
85. Pan, X.; Yang, J.; Van, R.; Epifanovsky, R.; Ho, J.; Huang, J.; Pu, J.; Mei, Y.; Nam, K.; Shao, Y. Machine-learning-assisted Free Energy Simulation of Solution-Phase and Enzymatic Reactrions *J. Chem. Theory Comput.* **2021,** *17*, 5745-5785.

TOC Graphic