Prediction of Alkanolamine pKa Values by Combined Molecular Dynamics Free Energy Simulations and ab initio Calculations

Javad Noroozi, William Smith

Submitted date: 26/11/2019 • Posted date: 06/12/2019
Licence: CC BY-NC-ND 4.0

Citation information: Noroozi, Javad; Smith, William (2019): Prediction of Alkanolamine pKa Values by Combined Molecular Dynamics Free Energy Simulations and ab initio Calculations. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11215025.v1

We use molecular dynamics free energy simulations in conjunction with quantum chemical calculations of gas phase reaction free energy to predict alkanolamines pka values.

File list (37)

AM1-BCC.pdf (10.16 KiB) view on ChemRxiv download file
Pka_amines_final.pdf (7.62 MiB) view on ChemRxiv download file
toc.pdf (824.29 KiB) view on ChemRxiv download file
1AP.png (215.25 KiB) view on ChemRxiv download file
2AEE.png (190.58 KiB) view on ChemRxiv download file
2AP.png (156.74 KiB) view on ChemRxiv download file
2DIPA.png (170.78 KiB) view on ChemRxiv download file
2MPA.png (144.35 KiB) view on ChemRxiv download file
3AP.png (163.81 KiB) view on ChemRxiv download file
3-DMAP.png (229.24 KiB) view on ChemRxiv download file
AEA.png (194.42 KiB) view on ChemRxiv download file
AEPD.png (181.16 KiB) view on ChemRxiv download file
AMP.png (165.10 KiB) view on ChemRxiv download file
AMPD.png (179.68 KiB) view on ChemRxiv download file
Prediction of Alkanolamine $pK_a$ Values by Combined Molecular Dynamics Free Energy Simulations and ab initio Calculations

Javad Noroozi† and William R. Smith⁎,‡,¶

†Department of Chemical Engineering, University of Waterloo, Waterloo ON N2L 3G1, Canada
‡Department of Mathematics and Statistics, University of Guelph, Guelph ON N1G 2W1, Canada
¶Faculty of Science, University of Ontario Institute of Technology, Oshawa ON L1H 7K4, Canada

E-mail: bilsmith@uoguelph.ca
Abstract

Knowledge of aqueous protonation constants ($pK_a$) of chemical species is of significant importance in CO$_2$ reactive absorption system design. Their theoretical prediction has mainly relied on implicit solvent models, and the performance of explicit solvent simulations based on classical force fields have rarely been studied. In this paper, we report the results of simulations in explicit TIP3P water with the General Amber Force Field (GAFF) and with the SMD continuum solvent method for the deprotonation $pK_a$ values of 29 conformationally diverse alkanolamine species commonly used in CO$_2$ capture. In both cases, we employ the Tissandier value for the hydration free energy of the proton (J. Phys. Chem. A, 1998, 102, 7787). The ideal–gas reaction free energies and their uncertainties were obtained from electronic structure calculations using five different compound methods (CBS–QB3, CBS–APNO, G3, G3B3, G4). The hydration free energies of the neutral and protonated forms of the amines were calculated using the semi–empirical AM1–BCC charge method, in addition to several partial atomic charge sets based on the RESP fitting method using electrostatic potentials computed at different ab initio theory/levels in the gas phase as well as in the presence of the solvent reaction field. We incorporated the Galvani surface potential of the ions in the ($pK_a$) calculations.

Although the individual species hydration free energies show significant sensitivity to the charge model, the resulting $pK_a$ values from different charge models are quite similar. Moreover, we found that the protonated amine hydration free energies show slightly less sensitivity to the partial charge method than in the case of the neutral amine. While the predicted $pK_a$ values based on the RESP charges yield reasonable agreement with the experimental data, they are prone to occasional disagreement for molecules of complex geometry. The best performance was achieved using the semi–empirical AM1–BCC charges, which showed a mean absolute error of less than 0.72 $pK_a$ units in comparison with experimental data. Our results suggest that the AM1–BCC charge method may be used to model electrolyte solutions encountered in the CO$_2$ reactive absorption process.
1 Introduction

Post–combustion capture (PCC) using aqueous amine solutions is one of the most mature technologies currently used for CO$_2$ capture from large point sources$^1$. Knowledge of chemical reaction behaviour in the solvent is crucial for efficient solvent design of PCC technology.

The p$K_a$ values of the amine species is one of the fundamental quantities that affects reaction equilibria/kinetics, as well as the CO$_2$ absorption capacity and the heat of solvent regeneration$^{2,3}$. Theoretical prediction of this quantity has mainly relied on a thermodynamic–cycle–based approach utilizing the dielectric continuum solvation model (DCSM)$^{4-6}$, and its variant that includes a small number of explicit solvent molecules in the first solvation shell (a “cluster–continuum model”) to account for hydrogen–bonding interactions$^{7-10}$. The DCSM approach has typically been tested on a limited set of small, rigid molecules, and its accuracy for large flexible molecules and ionic species has not yet been fully examined. While some DCSMs incorporate the effect of temperature on the solvation free energy (e.g., SM8T$^{11}$ and COSMO-RS$^{12}$), most have been tested for a limited number of solvents at 298.15 K.

Typical DCSM simulations consider only the lowest energy conformer of the solute (usually in the ideal gas (IG) phase), and treat it as a rigid molecule in the estimation of its free energy of transfer to the solvent.$^{13,14}$ As noted by Coote et al.$^{15}$ this implicitly assumes that the thermal contributions to the solute chemical potential ($i.e.$, those due to translational, rotational, and vibrational motions) are very similar in the IG and solution phases. While this may be a good approximation for the small rigid molecules that have been the target of most DCSM development, larger flexible molecules with more complex molecular geometries (such as amines and other hydrogen-bonding species) may undergo significant conformational and structural changes ($e.g.$, tautomerization) upon solvation, requiring the use of an ensemble of conformers for accurate solvation free energy estimation.$^{16,17}$ Furthermore, the identification of the most stable molecular conformations for a flexible molecule in solution is not a straightforward task, due to the fact that solution phase energies ($E_{\text{soln}}$), commonly used to rank the conformer relative stabilities may not be a proper metric, and estimated
solution phase Gibbs energies ($G_{\text{soln}}$) should instead be used in order to identify the most stable conformers. Recently, Haworth et al.\textsuperscript{16} showed that ignoring these effects can lead to erroneous p$K_a$ estimation for flexible molecules. Explicit solvent models, on the other hand, have the advantage that solvent effects on the solute conformations are inherently taken into account.

In spite of these shortcomings, DCSM simulations have been quite successful in predicting p$K_a$ values for a wide range of species, typically by employing strategies such as changing the level of quantum theory or using different optimal scaling factors to modify the original parameterization of the model to account for polarization in the solvent phase.\textsuperscript{18–20}

The SMx family of continuum solvation models of Cramer and Truhlar\textsuperscript{4} are widely used for p$K_a$ predictions.\textsuperscript{14,21–23} Pliego et al.\textsuperscript{14} assessed the performance of the SMD and the SM8 model for the estimation of p$K_a$ of diverse chemical species in methanol. They found RMS error values with respect to experiment of 3.6 and 2.7 p$K_a$ units for 23 amine species. Svendsen et al.\textsuperscript{23} reported an average absolute error of 1.65 p$K_a$ units for 25 amines in water using SM8T. Although these models have shown an overall good performance for certain classes of chemicals,\textsuperscript{10} they are unsuitable for systems involving explicit solute–solvent(water) interaction, especially involving ionic species and molecules forming intramolecular hydrogen bonds, which is difficult to describe accurately by a DCSM.

While DCSM simulations have been extensively used for p$K_a$ estimation, there are only a handful of studies using classical force–field (FF) methodology.\textsuperscript{24} This alternative approach uses explicit solvent molecular dynamics (MD) free energy simulation to model the complex intra- and intermolecular interactions using an underlying FF model. This methodology captures conformational effects when the solute is being annihilated/decoupled from its molecular environment by sampling its multiple conformations.

In a notable work, Brooks et al.\textsuperscript{24} used molecular dynamics simulations with empirical force field in conjunction with a DFT calculation of the IG free energy of deprotonation to predict the p$K_a$ values of drug–like molecules in a SAMPL6 challenge with an overall RMSE
of 2.4 p$K_a$ units. For ionic species, the solvation calculation requires both the bulk phase (intrinsic hydration free energy) and the air–water surface potential contribution (i.e., the Galvani potential)$^{25}$; however, this study omitted the surface contribution. Moreover, the value of the proton hydration free energy used in their work (-1106.7 kJ·mol$^{-1}$) refers to a standard state of 1 atm in the gas phase and 1 molar in solution, whereas the conventional hydration free energies calculated using the thermodynamic integration (TI) approach refer to a standard state of 1 molar in both gas and solution phases.

To the best of our knowledge there has been no systematic study using the classical force–field (CFF) approach to calculate either the solvation free energies (especially for the charged species) of complex molecular species involved in a protonation reaction or its p$K_a$ value. Although experimental solvation free energies exist for many neutral species involved in such a reaction, this is not the case for the charged species. The goal of the present study is to investigate the use of the CFF approach for a test set of 29 amine species comprising 13 primary amines, 10 secondary amines, and 6 tertiary amines using FFs generated using the General Amber Force Field (GAFF) approach with semi–empirical AM1–BCC charges and partial charges obtained from several different quantum chemical calculations of the molecular electron density computed using Hartree-Fock (HF), density functional theory (DFT) and the second-order Møller-Plesset theory (MP2). We calculated species solvation free energies and p$K_a$ values, and compared results with those of a popular implicit solvent method. We also compared with available experimental results to assess the quality of the methodologies and the employed force fields (FF).

The paper is organized as follows. The next section reviews the thermodynamic background for translating the relevant simulation quantities to the experimental p$K_a$ values. The subsequent section gives the details of our simulations, followed by a section presenting our results and their discussion. We conclude with a summary and conclusions section.
2 Thermodynamic Background

Whereas the chemical potential may be defined relative to different reference, standard states and concentration variables, the Henry-law-based (HL) chemical potential model using the infinite dilution reference state, the standard state as that of a hypothetical ideal solution of unit concentration, and the molality concentration variable \( m_i \) (in mol kg\(^{-1}\) solvent) is commonly used in macroscopic theoretical and experimental studies of electrolyte solutions. In contrast, molecular simulation results are most readily available using the ideal gas (IG) standard state and the density composition variable, \( \rho_i \) (in mol V\(^{-1}\), where \( V \) is the system volume). The translation between the two forms of chemical potential expressions leading to the expression of \( pK_a \) in terms of molecular simulation quantities is not straightforward, since it involves a careful consideration of standard state quantities. We thus first give a brief overview of the translation, and clarify some potential points of confusion regarding \( pK_a \) simulations that have appeared in the literature.

We consider \( pK_a \), the equilibrium constant for the deprotonation reaction

\[
\text{RNH}_3^+ = \text{RNH}_2 + \text{H}^+ \tag{1}
\]

where \( \text{RNH}_2 \) is a primary, secondary or tertiary alkanolamine.

For generality, we first derive an expression for \( pK_a \) in terms of molecular simulation quantities for the general case of a reaction that also involves \( \text{H}_2\text{O} \). This permits, for example, consideration of \( pK_a \) for the reaction

\[
\text{RNH}_3^+ + \text{H}_2\text{O} = \text{RNH}_2 + \text{H}_3\text{O}^+ \tag{2}
\]

Solute chemical potentials in the HL model using the molality concentration variable are
expressed as:

\[ \mu_i(T, P; \mathbf{m}) = \mu^\dagger_i(T; P) + RT \ln \left( \frac{m_i}{m_0} \right) + RT \ln \gamma_i(T, P; \mathbf{m}) \]  \tag{3} 

where \( \mathbf{m} \) denotes the system molality vector, \( \gamma_i \rightarrow 1 \) as \( m_i \rightarrow 0 \), and \( \mu^\dagger_i(T; P) \) is the standard state chemical potential of solute species \( i \) in a hypothetical ideal solution of \( m^0 = 1 \) molal. For the solvent, the Lewis–Randall (LR) chemical potential model using the pure solvent standard state and the mole fraction concentration variable is commonly adopted, and the solvent chemical potential is given by

\[ \mu_{\text{solv}}(T, P) = \mu^*_{\text{solv}}(T, P) + RT \ln a_{\text{solv}}(T, P; \mathbf{x}) \]  \tag{4} 

where \( \mathbf{x} \) denotes the system mole fraction vector, \( a_{\text{solv}} \) is the activity of the solvent, and \( \mu^*_{\text{solv}}(T, P) \) is the chemical potential of the pure solvent. In this model, \( a_{\text{solv}} \rightarrow 1 \) as \( x_i \rightarrow 1 \), or equivalently as the molality \( m_{\text{solv}} \rightarrow 1000/M_{\text{solv}} \), where \( M_{\text{solv}} \) is the solvent molecular weight.

It is most convenient to express all concentration variables in terms of the molality composition variable, which may be achieved by expressing \( \mu^*_{\text{solv}}(T, P) \) as

\[ \mu^*_{\text{solv}}(T, P) = \mu^\dagger_{\text{solv}}(T, P) + RT \ln \left( \frac{1000}{M_{\text{solv}}} \right) \]  \tag{5} 

and Eq. (4) then becomes

\[ \mu_{\text{solv}}(T, P) = \mu^\dagger_{\text{solv}}(T, P) + RT \ln \left( \frac{1000}{M_{\text{solv}}} \right) + RT \ln a_{\text{solv}}(T, P; \mathbf{m}) \]  \tag{6} 

\( \mu^\dagger_i(T, P) \) (for both solutes and the solvent) is related to the intrinsic solvation free energy,
\[ \mu^\text{res,NVT:}\infty_i[T, \rho(T, P)] \text{ by}^{26,27} \]

\[ \mu^\dagger_i(T, P) = \mu^0_i(T; P^0) + RT \ln \left( \frac{RT}{100 P^0} \right) + RT \ln \left( \frac{\rho^\text{solv}(T, P)}{1000} \right) + \mu^\text{res,NVT:}\infty_i[T, \rho^\text{solv}(T, P)] \]

(7)

where \( \mu^0_i(T; P^0) \) is the species ideal–gas (IG) chemical potential at \( T \) and standard state pressure \( P^0 \) (expressed in bar) and \( \rho^\text{solv} \) is the density of the pure solvent. \( \rho^\text{solv} \) denotes its expression in kg m\(^{-3}\).

The second and third terms in Eq. (7) bring the IG from its density at the standard state \( T \) and standard state pressure \( P^0 = 1 \text{ bar} \) to that of an ideal solution density of unit molality, and we refer to \( \mu^\text{res,NVT:}\infty_i[T, \rho^\text{solv}(T, P)] \) as the solute’s intrinsic solvation free energy. It is the molar free energy change when a solute molecule is transferred from the ideal gas phase to the pure solvent and both phases are at the same density. For a solvent, this quantity is equivalent to its intrinsic self–solvation free energy. \( \mu^\text{res,NVT:}\infty_i[T, \rho^\text{solv}(T, P)] \) is readily obtained by means of conventional free energy calculations in an MD simulation software package such as GROMACS.\(^\text{28}\)

The equilibrium condition for a chemical reaction \( j \) is

\[ \Delta G_j \equiv \sum_{i=1}^{N_s} \nu_{ij} \mu_i = 0 \]

(8)

where \( N_s \) is the total number of species and \( \nu_{ij} \) is the stoichiometric coefficient of species \( i \) in reaction \( j \) (conventionally positive for products and negative for reactants). Using Eqs. (3) and (6), Eq. (8) becomes

\[ \sum_{i=1}^{N_s} \nu_{ij} \mu^\dagger_i(T, P) + RT \nu^\text{solv,j} \ln \left( \frac{1000}{M^\text{solv}} \right) + RT \left( \sum_{\text{solutes}} \nu_{ij} \ln \left( \frac{m_i \gamma_i(T, P; m)}{m^0} \right) \right) + \nu^\text{solv,j} \ln[a^\text{solv}(T, P; m)] = 0 \]

(9)

Finally, \( pK \) for reaction (8) is obtained from the concentration–independent terms of Eq.
via:
\[ pK_j = \frac{\Delta G^*_j}{RT \ln(10)} \] (10)

where (using Eq. (7))

\[
\Delta G^*_j(T, P) = \sum_{i=1}^{N_s} \nu_{ij} \mu_i^1(T; P) + RT \nu_{solv,j} \ln \left( \frac{1000}{M_{solv}} \right) \\
= \sum_{i=1}^{N_s} \nu_{ij} \mu_i^0(T; P^0) + RT \bar{\nu}_j \ln \left( \frac{RT}{100 P^0} \right) + RT \nu_{solv,j} \ln \left( \frac{RT}{100} \right) \\
+ RT \nu_{solv,j} \ln \left( \frac{1000}{M_{solv}} \right) + \sum_{i=1}^{N_s} \nu_{i,j} \mu_i^{res,NVT;\infty}[T, \rho(T, P)] 
\] (12)

and \( \bar{\nu}_j = \sum \nu_{ij} \).}

Eq. (12) is a general expression for the equilibrium constant of any reaction \( j \) when the HL model using the molality concentration variable is used. For the deprotonation reaction of Eq. (1), with \( \bar{\nu}_j = 1 \) and \( \nu_{solv,j} = 0 \), the following well-known expression for \( \Delta G^*_j \) is obtained:

\[
\Delta G^*_j(T, P) = \Delta G^0_j(T; P^0) + RT \ln \left( \frac{RT}{100 P^0} \right) + RT \ln \left( \frac{\rho_{solv}(T, P)}{1000} \right) + \sum_{i=1}^{N_s} \nu_{i,j} \mu_i^{res,NVT;\infty}[T, \rho(T, P)] 
\] (13)

where

\[
\Delta G^0_j(T; P^0) = \sum_{i=1}^{N_s} \nu_{ij} \mu_i^0(T; P^0) 
\] (14)

In the literature, the term involving the density in Eqs. (12) and Eq. (13) is sometimes omitted. At \( P = 298.15 \) K and \( P = 1 \) bar, for reactions with \( \bar{\nu}_j \neq 0 \) and water as solvent, the density is very near 1000 kg m\(^{-3}\), and the term is negligible. However, when the solvent density is far from 1000 kg m\(^{-3}\), omission of this term can result in significant error for the \( pK_a \) value. The term has been thusly mistakenly omitted in the literature for aqueous amine \( pK_a \) calculations at higher temperatures,\(^{29,30}\) and for non-aqueous solvents such as ionic liquids and dimethyl sulfoxide (MDSO) at ambient conditions.\(^{31} \) For example, for
methanol solvent at 298.15K and 1 bar (ρ = 765.8 kg m⁻³) its omission introduces an error of 0.26 pKₐ units.

Another important point is that for an ionic species, Eq. (7) omits the contribution to \( \mu_{i}^{\text{res}} \) due to its crossing of the vacuum–solvent interface (i.e., the solvent’s Galvani potential, \( \phi_G \), which arises from surface polarization³²). Hence, the solvent–specific Galvani potential must be added to the intrinsic solvation free energy to obtain the absolute/real solvation free energy of an ion according to³³,³⁴

\[
\mu_{i}^{\text{res,ion}} = \mu_{i}^{\text{res,NVT}}[T, \rho(T, P)] + z_i \phi_G
\] (15)

where \( z_i \) is the ion’s valence.

Although the Galvani contributions cancel in a charge–balanced reaction, it must be included when an individual ion is of interest. In this study, we calculate pKₐ for the deprotonation reaction of Eq. (1) and use the value of the hydration free energy of H⁺ in water from Tissandier et al.³⁵ As noted by Palmer et al.,³³ numerous studies suggest that the Tissandier result is the absolute value, which incorporates the Galvani potential³⁶–³⁸. Hence, when the Tissandier value for H⁺ is used, the Galvani potential must be added to the intrinsic hydration free energy of RNH₃⁺ to obtain the correct pKₐ value for reaction (1). This has mistakenly been ignored in the SAMPL6 challenge study of Brooks et al.²⁴

3 Simulation Details

3.1 Conformational Search and Ideal Gas Reaction Free Energies

Initially, all possible molecular conformations were constructed for the 29 amine species studied and their protonated forms using the Spartan18 software package.³⁹ The resulting geometries were then further optimized using the fast semi–empirical PM6 model implemented in Spartan. When the conformational space was very large (more than 25 conformers), only the
first 25 with the lowest PM6 energies were submitted for further optimization and frequency calculations using five compound model chemistries using the Gaussian16 program:\textsuperscript{40} CBS–QB3, CBS-APNO, G3B3, G3 and G4. For each compound method, the conformer with the lowest Gibbs free energy (and no negative imaginary frequency) was used for the IG reaction free energy calculations at the standard state of $T = 298.15\, \text{K}$ and $P = 1\, \text{bar}$ and using standard expressions for the partition function under the harmonic oscillator rigid rotor approximation.

### 3.2 Partial Charges and Intrinsic Hydration Free Energy Calculations

All amine species were modeled using the General Amber Force Field (GAFF)\textsuperscript{41} with partial charges derived from RESP charge fitting,\textsuperscript{42} in addition to the semi–empirical AM1–BCC\textsuperscript{43} charge model in TIP3P water. The molecular structures of the amines in the test set are shown in Fig.1. For each molecule, starting with the PM6 optimized geometries of its conformers as described in the previous section, they were further optimized at the MP2–cc–pVTZ level to find the conformer with the lowest energy in the gas phase. Single-point calculations were then performed on this conformer using the HF/6–31G(d) and B3LYP/6–311++G(d,p) levels of theory. The electron density based on these two levels is widely used in conjunction with the GAFF intramolecular parameters,\textsuperscript{44–46} since they tend to overestimate the solute dipole moments in the gas phase and implicitly incorporate the polarization effect in the solution. Additionally, the electron density was calculated using the more sophisticated second–order Møller–Plesset theory employing the cc–pVTZ basis set in the presence of a solvent reaction field mimicked using the polarizable continuum model (PCM\textsuperscript{5}) with a dielectric constant of 78.39. For the latter, we allowed the MP2–cc–pVTZ gas phase minimum conformer geometry to further relax in the solution. All partial charge assignments were performed with the Antechamber package (version 17.3) of Ambertools,\textsuperscript{47} and Gromacs topologies were generated using the acype (version 2019) python interface.\textsuperscript{48}
Figure 1: Molecular Structures of the 29 alkanolamines investigated in this work.

All MD simulations were performed with the GROMACS 2018.3 program\textsuperscript{28} in a cubic box of one solute molecule solvated in 888 TIP3P water molecules using packmol.\textsuperscript{49} The resulting structure was then minimized to remove any bad contacts, and a short $NVT$ run was followed by a 12 ns $NPT$ simulation. The free energy simulations were then started from
these equilibrated structures in an NVT ensemble with box size corresponding to the TIP3P density at the standard state of $T = 298.15$ K and $P = 1$ bar. The equations of motion were integrated using the stochastic Langevin scheme, with a friction constant of 1.0 ps$^{-1}$. Lennard-Jones short-range interactions were smoothly switched off between 12 and 12.5 Å, and the electrostatic interactions were computed using the particle mesh Ewald (PME) method with a 12 Å real space cutoff, 1.0 Å grid spacing, 6th-order spline interpolation, and accuracy of $10^{-6}$.

For the constant pressure simulations, a Parrinello–Rahman pressure coupling constant of 2.0 ps was used. The free energy of decoupling the solute in the solvent environment was calculated using the statistically optimal Multi-state Bennett Acceptance Ratio (MBAR) method$^{50}$ with the Hamiltonian difference between the neighbouring states saved every 0.2 ps. We employed 6 equally spaced $\lambda$ values and linear decoupling of the electrostatics interaction, followed by 20 $\lambda$ values with equal spacing of $\Delta \lambda = 0.05$ to decouple the LJ interactions using the standard GROMACS soft-core potential function originally proposed by Beutler et al.$^{51}$ with parameters (in GROMACS notation) $sc\text{-}alpha = 0.5$, $sc\text{-}power= 1$ and $sc\text{-}sigma=0.3$. Each $\lambda$ window was subjected to a 12.5 ns simulation with the first 2.5 ns discarded for equilibration.

4 Results and Discussion

Various data used in our calculations is summarized in Table 1.

4.1 Intrinsic Hydration Free Energies of Amine Species

Since experimental free energies of solvation are not available for the species in this study, in Fig.2 we compare our GAFF results against the SMD solvation model of Truhlar et al.$^{55}$ which is parameterized with the goal of reproducing experimental solvation free energies at 298.15 K. The SMD simulations were performed using a Gaussian16 implementation similar
Table 1: Thermodynamic data used in this study. All values are at \( T = 298.15 \text{ K} \), \( P = 1 \text{ bar} \). \( \mu_{0}^{H^{+}} \) is based on the standard equations of thermodynamics for the ideal gas enthalpy and entropy of a monatomic classical ideal gas based on the Sackur–Tetrode equation, and \( \rho_{H_{2}O} \) and \( \phi_{G} \) refer to the TIP3P water model. \( \mu_{H^{+}}^{\text{res,}NVT;\infty} \) refers to a 1M molar concentration in the ideal gas and a 1M ideal solution.

| Property       | Value          | Source                |
|----------------|----------------|-----------------------|
| \( \mu_{0}^{H^{+}} \) | -26.2 kJ·mol\(^{-1}\) | McQuarrie\(^{52}\) |
| \( \rho_{H_{2}O} \) | 987.4 kg·m\(^{-3}\) | This work            |
| \( \phi_{G} \)    | -48.24 kJ·mol\(^{-1}\) | \(^{34,53,54}\)       |
| \( \mu_{H^{+}}^{\text{res,}NVT;\infty} \) | -1112.5 kJ·mol\(^{-1}\) | Tissandier et al.\(^{35}\) |

to that of Cox et al.\(^{13}\) Starting with the gas phase cc–pVTZ optimized minimum energy conformer geometries, the solute structures were further optimized at the M06–2X/c–pVTZ level in vacuum, followed by a single–point energy calculations at the M06–2X/6–31G(d)level. Second single–point energy calculations were performed on the vacuum geometry in the presence of the SMD water model to obtain the solute’s infinite dilution residual chemical potential and its absolute solvation free energy. The raw intrinsic hydration free energies from all the charge methods and from SMD are tabulated in the Supporting Information.

As shown in Fig.2, there is reasonable agreement (+/- 5 kJ·mol\(^{-1}\)) between the RESP–derived hydration free energies and those of the SMD model. However, for heavily sterically hindered amines with multiple hydroxyl group (each enclosed by an ellipse), the hydration free energies based on RESP charges deviate from the corresponding SMD values. RESP assigns charges to atoms based on a grid of electrostatic potential points, and is quite sensitive to the conformer geometry; the presence of strong intramolecular electrostatic interactions may thus lead to unreasonably distorted partial charges for the atoms involved in such interactions. To alleviate this, multiple conformation RESP fitting may be used.\(^{56}\) The AM1–BCC results (shown by blue circles) are in the best agreement with the SMD results for all solutes. A possible explanation for this is that AM1–BBC derives partial charges from the AM1 semi–empirical quantum mechanical calculation, after which empirical bond–charge corrections (BCCs) are applied to the partial charges. This makes the AM1–BCC
charges less sensitive to the conformation and to steric clashes. Previously, AM1–BCC has been shown to give superior performance for hydration free energy predictions of species in the FreeSolv Database.\textsuperscript{57,58}

![Parity plot comparison of the neutral amine hydration free energies (in kJ·mol\textsuperscript{-1}) against the corresponding SMD results. The black line shows the y=x parity line and the dotted green lines show y=x +/- 5 kJ·mol\textsuperscript{-1}.](image)

While there exist numerous studies addressing the effects of the charge model for neutral species on the resulting hydration free energies,\textsuperscript{59,60} this has been rarely studied for ionized species. A main reason is that such data for ionized species is generally inaccessible experimentally. However, we have seen that AM1–BCC performs well for neutral species with respect to SMD, which is considered to perform well with respect to experiment for such species. We will thus test both methods against experimental results for \(pK_a\) values and use the results to infer the quality of the AM1–BCC and SMD predictions for the ions.

We first benchmark different charge methods against the AM1–BCC approach. For a
charged solute, the intrinsic hydration free energy obtained from an atomistic simulations with periodic boundary conditions and Ewald summation differs from the absolute/real solvation free energy, which includes the contribution from the solvent’s Galvani surface potential, whereas a continuum solvent model (including SMD), inherently includes this contribution. To obtain the absolute hydration values for the protonated amines, the Galvani potential value for the TIP3P water model was added to the raw intrinsic values from GROMACS MBAR simulations. The resulting absolute hydration free energies obtained using RESP charges and using the SMD model are benchmarked against the AM1–BCC results in Fig. 3. Clearly, for protonated amines, the hydration free energy from different RESP charge sets agrees well with the AM1–BCC result, whereas for species with multiple hydroxyl groups, the SMD values (orange circles) show significant deviations. This could be attributed to the inability of SMD to describe complex hydrogen bonding interactions in these molecular species.
Figure 3: Parity plot comparison of the amine cation hydration free energies (in kJ·mol\(^{-1}\)) against the corresponding semi–empirical AM1–BCC results. The black line shows the y=x parity line and the dotted green lines show y=x +/- 5 kJ·mol\(^{-1}\).

4.2 Pka calculations

The gas–phase protonation reaction free energies or basicities for Eq. (1) at \(T = 298.15\) K and \(P = 1\) bar are summarized in Table 2 using five composite methods: CBS–QB3, CBS–APNO, G3B3, G3 and G4. The final column gives the average value from the methods, which we use in our calculations, and the standard deviation, which we use as the uncertainty. Only the free energies of the neutral amine and its protonated form were calculated from the molecular partition function and frequency calculations. For the species in our list, we found experimental reaction free energies (all in kJ·mol\(^{-1}\)) for MEA (896.8), 3–AP (917.3), EDA (912.5), PA (883.9) and 2-MPA (890.8) from a literature database\(^{61}\). Given that the experimental basicity may be subject to an uncertainty of several kJ, the agreement between
the calculated and experimental values is considered to be reasonable.

Table 2: $\Delta G^0$ in Eq. (14) for the reaction $\text{RNH}_3^+ = \text{RNH}_2 + \text{H}^+$ for 29 amines in kJ·mol$^{-1}$ at $T = 298.15$ K and $P = 1$ bar considered in this work. Subscripts in the final column denote the standard deviation of the indicated value. A superscript indicates the type of amine: $p$, $s$ and $t$ respectively denote a primary, secondary and tertiary alkanolamine.

| Amine                               | Abbreviation | CBS–QB3 | CBS–APNO | G3B3  | G3   | G4   | Average   |
|-------------------------------------|--------------|---------|----------|-------|------|------|-----------|
| monoethanolamine$^p$               | MEA          | 888.9   | 888.5    | 891.0 | 891.0| 891.0| 889.761.17|
| 3-amino-1-propanol$^p$             | 3-AP         | 916.9   | 919.2    | 919.9 | 920.0| 919.6| 919.121.28|
| ethylenediamine$^p$                | EDA          | 914.2   | 917.1    | 915.9 | 916.8| 915.5| 915.901.15|
| propanamine$^p$                    | PA           | 885.9   | 888.4    | 887.8 | 888.4| 888.0| 887.701.03|
| 2-methyl-1-propanamine$^p$         | 2-MPA        | 890.9   | 893.5    | 892.5 | 893.2| 892.6| 892.541.00|
| 2-amino-2-methylpropanol$^p$       | AMP          | 909.3   | 911.1    | 911.4 | 911.5| 911.4| 910.940.92|
| 2-amino-2-methyl-1,3-propanediol$^p$ | AMPD     | 921.3   | 922.6    | 923.8 | 923.7| 920.3| 922.341.52|
| 2-amino-2-ethyl-1,3-propanediol$^p$ | AEPD     | 924.5   | 926.3    | 929.6 | 930.1| 929.4| 927.982.45|
| 2-amino-1-propanol$^p$             | 2-AP         | 900.5   | 902.5    | 902.7 | 902.9| 902.6| 902.240.98|
| diethanolamine$^s$                 | DEA          | 942.0   | 944.7    | 945.8 | 945.6| 945.3| 944.681.55|
| 2-(2-aminoethoxy)ethanol$^p$       | 2-AEE        | 928.7   | 926.5    | 933.0 | 932.6| 932.5| 930.662.90|
| 2-(diisopropylamino)ethanol$^t$    | 2-DIPA       | 965.4   | 969.6    | 969.0 | 969.3| 969.5| 968.561.78|
| diisopropanolamin$^s$              | DIPA         | 958.2   | 958.6    | 961.3 | 960.8| 961.5| 960.081.56|
| methylkdiethanolamine$^t$          | MDEA         | 961.7   | 960.6    | 967.8 | 961.2| 966.3| 963.523.28|
| n-cyclohexylethanolamine$^s$       | n-CHEA       | 944.2   | 947.0    | 946.7 | 947.2| 948.5| 946.721.56|
| serinol(2-aminopropane-1,3-diol)$^p$ | SAPD      | 913.1   | 914.4    | 913.9 | 916.0| 915.8| 914.641.24|
| 2-(tert-butylamino)ethanol$^s$     | TBAE         | 942.0   | 945.0    | 945.0 | 945.0| 945.5| 944.501.41|
| tris(hydroxymethyl)aminomethane$^p$ | THMAM    | 919.6   | 920.9    | 922.0 | 923.4| 924.2| 922.021.14|
| 3-dimethylamino-1-propanol$^t$     | 3-DMAP       | 955.2   | 958.3    | 960.0 | 960.4| 959.6| 958.702.1 |
| N,N-dimethylisopropanolamine$^s$   | DMIPA        | 935.8   | 939.9    | 939.9 | 940.7| 939.8| 939.221.94|
| tert-butylkdiethanolamine$^t$      | t-BDEA       | 971.1   | 976.0    | 973.6 | 976.4| 976.6| 974.742.36|
| triethylamine$^t$                  | TREA         | 949.4   | 950.6    | 947.6 | 950.0| 950.2| 949.561.17|
| 1-amino-2-propanol$^p$             | 1-AP         | 895.2   | 894.6    | 897.2 | 895.5| 893.1| 895.121.48|
| 2-(methylamino)ethanol$^s$        | MAE          | 916.8   | 919.2    | 919.9 | 920.3| 920.0| 919.241.42|
| 2-(ethylamino)ethanol$^s$         | EAE          | 930.2   | 932.6    | 933.2 | 933.0| 933.4| 932.481.30|
| 2-(isopropylamino)ethanol$^s$     | IPAE         | 935.0   | 937.9    | 938.0 | 938.0| 938.3| 937.441.37|
| 2-((1-methylpropyl)amino)ethanol$^s$ | MPAE     | 939.1   | 942.8    | 942.0 | 942.6| 941.7| 941.461.48|
| 2-(isobutylamino)ethanol$^s$       | IBAE         | 936.6   | 939.4    | 939.2 | 939.2| 939.3| 938.741.20|
| 2-(ethylamino)-2-methyl-1-propanol$^s$ | EAMP   | 944.3   | 946.9    | 947.5 | 947.6| 947.3| 946.731.39|

For our $pK_a$ calculations, we used the average gas–phase values of Table 2 and the data in Table 1.

Our predicted $pK_a$ values for each charge model are compared with the SMD predictions and with available experimental results in Table 3. We ascribe the $pK_a$ uncertainties in
Eq. (10) to those of the corresponding ideal–gas contributions of Table 2. (This excludes any uncertainty in the Tissandier proton hydration free energy, and the relatively small uncertainties of the simulation results for the intrinsic hydration free energies.)

Table 3: $pK_a$ values for the reaction $\text{RNH}_3^+ = \text{RNH}_2 + \text{H}^+$ using different charge models and from SMD continuum solvent simulations at $T = 298.15$ K and $P = 1$ bar.

| amine | HF/6–31G* | B3LYP/6–311++G(d,p) | MP2–cc–pVTZ+PCM | SMD | AM1–BCC | $pK_a(\text{expt})^{62–67}$ |
|-------|-----------|---------------------|------------------|-----|---------|-----------------------------|
| MEA   | 7.42      | 7.67                | 8.19             | 7.62| 9.00    | 9.47                        |
| 3-AP  | 8.79      | 9.26                | 10.09            | 9.95| 10.12   | 10.00                       |
| EDA   | 8.20      | 8.36                | 8.95             | 9.43| 9.97    | 9.90                        |
| PA    | 5.61      | 5.59                | 8.18             | 9.46| 8.98    | 10.60                       |
| 2-MPA | 6.34      | 6.34                | 8.93             | 10.10| 8.87    | 10.50                       |
| AMP   | 7.56      | 7.02                | 8.34             | 8.33| 9.39    | 9.70                        |
| AMPD  | 6.89      | 6.34                | 7.16             | 6.81| 8.81    | 8.80                        |
| AEPD  | 6.58      | 6.06                | 6.26             | 7.49| 9.24    | 8.80                        |
| 2-AP  | 7.16      | 7.38                | 8.15             | 7.89| 9.35    | 9.40                        |
| DEA   | 9.49      | 9.46                | 8.71             | 7.18| 10.34   | 9.00                        |
| 2AEE  | 10.60     | 10.6                | 11.61            | 8.41| 10.33   | 9.42                        |
| 2DIPA | 9.50      | 8.75                | 8.51             | 12.09| 9.25    | 9.42                        |
| DIPA  | 9.99      | 9.79                | 8.81             | 9.46| 9.74    | 8.88                        |
| MDEA  | 11.55     | 11.90               | 11.14            | 11.65| 10.41   | 8.57                        |
| n-CHEA| 9.25      | 8.35                | 9.39             | 9.29| 9.09    | 10.10                       |
| SAPD  | 7.70      | 7.55                | 7.90             | 6.78| 9.34    | 8.55                        |
| TBAE  | 9.55      | 9.00                | 9.50             | 10.24| 9.59    | 9.70                        |
| THMAM | 4.34      | 3.14                | 3.95             | 2.48| 7.19    | 8.08                        |
| 3-DMAP| 10.72     | 10.90               | 9.94             | 11.36| 9.30    | 9.27                        |
| DMIPA | 10.42     | 10.35               | 11.26            | 12.51| 8.28    | 9.47                        |
| t-BDEA| 8.53      | 8.00                | 11.52            | 11.99| 9.55    | 9.03                        |
| TREA  | 8.96      | 8.28                | 10.29            | 12.90| 8.38    | 10.70                       |
| 1AP   | 6.76      | 6.76                | 7.60             | 7.98| 8.74    | 9.50                        |
| MAE   | 9.85      | 10.21               | 10.66            | 8.83| 9.17    | 9.80                        |
| EAE   | 9.34      | 9.71                | 10.06            | 11.82| 9.79    | 9.99                        |
| IPA E | 9.44      | 9.10                | 9.93             | 11.05| 9.33    | 9.93                        |
| MPA E | 9.41      | 8.73                | 9.86             | 11.82| 9.66    | -a                          |
| IBAE  | 8.86      | 8.80                | 10.42            | 9.55| 9.62    | -a                          |
| EAMP  | 9.24      | 9.20                | 9.68             | 11.65| 9.87    | -a                          |

$^a$ No experimental data found
$^b$ Average absolute deviation

$^{\text{AAD}}$ 1.64 1.85 1.27 1.67 0.72
All our raw simulation results are given in the Supplementary Information, and the quality of the agreement with experiment of the results of different approaches can be assessed by various statistical measures. We note in passing that the Pearson $R^2$ value is not a useful statistic for assessing the ability of a method to accurately predict experimental data, due to its well-known deficiencies.\textsuperscript{68,69} For example, $R^2$ measures the quality of a linear regression of $y$(predicted) on $y$(experiment), whereas the question at hand is whether or not that the quantities are equal. We consider here the two alternative measures of the Average Absolute Deviation (AAD) and the confidence interval arising from the Student t–test for the hypothesis that the predicted and experimental values are equal.\textsuperscript{70} Based on the AAD values alone, Table 3 indicates that the AM1–BCC charges are in the best agreement with the experimental data for the test set of species considered.

The validity of the Student t–test approach is based on the assumption that the predicted and experimental values each independently follow a normal distribution, which is not an unreasonable assumption. The results comparing each of the AM1–BCC and SMD methodologies by this approach are shown in Table 4. For the entire set of 26 species, the 95% confidence intervals for corresponding differences of the predicted minus the experimental values for both approaches enclose the zero value, indicating that there is insufficient information to reject the hypothesis that either approach adequately predicts the experimental results. Notwithstanding, the combination of its lower AAD value and tighter confidence interval suggests that the AM1–BCC method is slightly superior. When the same analysis is applied to the different types of alkanolamines are considered, the results of Table 4 suggest that the AM1–BCC approach is satisfactory for each type of alkanolamine, whereas the SMD approach performs poorly for the primary and tertiary alkanolamines.
Table 4: Comparisons with experiment of $pK_a$ values for the reaction $\text{RNH}_3^+ = \text{RNH}_2 + \text{H}^+$ using AM1–BCC and SMD continuum solvent simulations at $T = 298.15$ K and $P = 1$ bar. AAD is the average absolute deviation of the predicted from the experimental results, AD is the average deviation of the predicted minus the experimental values of the indicated method, 95% Confidence Interval is computed from the data using the student t distribution at the 95% level as described in the text, and the final column indicates whether or not the hypothesis that the predicted and experimental values are equal can be rejected at this confidence level.

| Method         | AAD | AD  | 95% Confidence Interval | Hypothesis: AD=0 |
|----------------|-----|-----|-------------------------|------------------|
| Entire Data Set (26 species) |     |     |                         |                  |
| AM1–BCC        | 0.72| 0.18| (−0.20, 0.55)           | Do not reject    |
| SMD            | 1.69| 0.12| (−0.82, 1.08)           | Do not reject    |
| Primary Alkanolamines (13 species) |     |     |                         |                  |
| AM1–BCC        | 0.63| 0.26| (−0.36, 0.87)           | Do not reject    |
| SMD            | 1.54| 1.54| (0.37, 2.71)            | Reject           |
| Secondary Alkanolamines (7 species) |     |     |                         |                  |
| AM1–BCC        | 0.67| 0.04| (−0.49, 0.58)           | Do not reject    |
| SMD            | 1.09| 0.08| (−1.25, 1.39)           | Do not reject    |
| Tertiary Alkanolamines (6 species) |     |     |                         |                  |
| AM1–BCC        | 1.00| 0.16| (−0.82, 1.13)           | Do not reject    |
| SMD            | 2.73| 2.73| (1.88, 3.53)            | Reject           |

A parity plot comparing the AM1–BCC results against the experimental values is shown in Fig. 4. For only 5 of the 29 molecules studied (propanamine(PA), 2-methyl-1-propanamine(2-MPA), triethylamine(TREA), diethanolamine(DEA) and methyldiethanolamine(MDEA)), is the error appreciably larger than 1.0 $pK_a$ units.
Figure 4: AM1–BCC calculated versus experimental pK\textsubscript{a} values for the 29 alkanolamine species considered in this work. Red denotes primary amines, blue denotes secondary amines and green denotes tertiary amines. The vertical error bars are obtained by propagating the uncertainty (horizontal error bars for the experimental data are not shown) in the ideal gas reaction free energy and the dashed line indicates a tolerance of 1.0 pK\textsubscript{a} unit.

The fact that DEA and MDEA, in addition to the PA and 2-MPA pair, differ only in a methyl group, suggests a common source of this discrepancy. The largest error in the predicted value is associated with MDEA and TREA. For the former the gas phase reaction free energy also shows significant uncertainty. Although less pronounced for the AM1–BCC charges, possibly due to \textit{ad hoc} bond–charge corrections, the significant deviations of the PA and 2–MPA values using HF/6–31G* and B3LYP/6–311++G(d,p) charges suggest that partial charges based on the gas phase electron density might be unsuitable for these two molecules. This is further supported by the fact that the inclusion of implicit polarization using MP2–cc–pVTZ+PCM or SMD significantly improves the prediction of their pK\textsubscript{a} values.

HF/6–31G* and B3LYP/6–311++G(d,p) gas phase charge methods yield very similar
results, and both generally tend to underestimate the experimental pK_a values. MP2–cc–PVTZ+PCM performs significantly better than the gas phase HF and B3LYP charge methods, indicating the importance of polarization effects in the solvation calculation. All methods based on RESP charges tend to perform poorly for the heavily hindered amines with multiple hydroxyl groups, possibly due to effects such as local over–polarization or steric clashes.

For three of the studied alkanolamines (i.e., 2-((1-methylpropyl)amino)ethanol (MPAE), 2-(isobutylamino)ethanol(IBAE) and 2-(ethylamino)-2-methyl-1-propanol(EAMP)), we were unable to locate experimentally measured pK_a values. However, we note that the AM1–BCC predicted value of 9.66 for IBAE is very close to the experimentally measured value of 9.92 for that of its isomer 2-(butylamino)ethanol (BAE). The corresponding AM1–BCC predicted value for EAMP (9.87) is higher than that of AMP(9.39), suggesting that steric hinderance by the addition of an alkyl group to the nitrogen atom of AMP increases its pK_a value. Similarly, based on the AM1–BCC predictions, the pK_a values of EAE (9.79), MPAE(9.66), IBAE(9.62), TBAE (9.59), IPAE(9.33) are higher than that of MEA(9.00), further indicating that the addition of alkyl chains to the nitrogen atom of MEA increases the pK_a value. This trend was also observed in the recent experimental study of Narku–Tetteh et al., who found that the longer alkyl chain lengths of secondary alkanolamines resulted in higher equilibrium CO_2 loading and pK_a values.

Since the AM1–BCC pK_a results agree well with experiment and the AM1–BCC results for neutral species agree well with the SMD results, we conclude that the AM1–BCC results for the charged species are also accurate. In general, we believe that AM1–BCC is a very promising approach for predicting alkanolamine pK_a values and for studying the effects of different functional groups on their potential CO_2 capturing abilities.
5 Summary and Recommendations

We have implemented molecular simulation methodology for calculating p$K_a$ values at ambient conditions ($T = 298.15$ K and $P = 1$ bar) for the deprotonation reaction of Eq. (1) for 29 potential CO$_2$ solvents involving primary, secondary, and tertiary alkanolamines.

We used the ideal-gas (IG) free energy of -26.28 kJ·mol$^{-1}$ for H$^+$ relative to 0 Kelvin, and calculated the free energy of each amine and its protonated form. We investigated the use of five composition methods: CBS–QB3, neutral CBS–APNO, G3B3 and G4. We used TIP3P water as the solvent model and calculated the neutral and protonated amine force fields using GAFF in conjunction with both RESP and AM1–BCC charges. We calculated the solution phase hydration free energies of the amines and their protonated forms, and used the value of Tissandier et al. for the H$^+$ hydration free energy. Since this incorporates the hydration Galvani surface potential, we also incorporated it in our calculations for the cations, using the TIP3P value of -48.25 kJ·mol$^{-1}$.$^{53,72}$

For the hydration free energies of the neutral molecules, we found that the AM1–BCC results agree more closely with the SMD values than those of the RESP methods. The different charge methods generally agree; however, for species with multiple hydroxyl groups, the RESP hydration free energies significantly differ from the AM1–BCC results. For the ions, the agreement of the RESP and AM1–BCC models is better than in the case of the neutral molecules. A possible explanation is that for the ions, the intermolecular interaction with the solute is dominated by the total Coulombic charge, and the effects of the intramolecular interactions on the electronic density are less pronounced than for the neutral molecules.

We found that the AM1–BCC charge method yielded the best overall AAD for the experimental p$K_a$ of 0.72 p$K_a$ units. Based on these results, we suggest that AM1–BCC also gives the best results for the ions. We also considered a statistical analysis of our data by calculating the 95% confidence interval for the mean difference between the predicted and experimental p$K_a$ values, based on the use of the Student t distribution, an approach based only on the reasonable assumption that the predicted and experimental data are normally
distributed. We applied this approach to our AM1–BCC and our SMD data sets, and found that the hypothesis that the predicted and experimental values are equal could be rejected only for the SMD $pK_a$ predictions of the primary and tertiary alkanolamines.

We conclude that larger test sets of data are required to better distinguish between the abilities of different approaches to adequately predict experimental $pK_a$ data.

The accuracy of the predictions may be improved by using various method/basis set combinations to overcome the limitation of the RESP derived charges (i.e., its reliance on the on the gas phase HF/6-31G* calculations to derive the condensed phase charges). Developing fixed charge force fields to calculate the phase transfer properties such as hydration free energy is the subject of ongoing research.\textsuperscript{58,73,74} For example, the physically motivated IPolQ partial charge method of Cerutti and \textit{et al.}\textsuperscript{75} uses the average partial charges of the gas phase and liquid phase to implicitly include the effect of the solute polarization upon transfer between the phases. Finally, we believe that the $pK_a$ prediction using fixed charge explicit solvent can be further improved with optimization of the GAFF LJ parameter which originally designed to work with the HF/6-31G* based partial charges.

Our approach can be extended to consider the temperature dependence of $pK_a$ in both aqueous and nonaqueous solvents (\textit{e.g.}, methanol). Using our recent reaction equilibrium simulation approach,\textsuperscript{27} the $pK_a$ results of this paper can be used to predict CO\textsubscript{2} solubility for a range of tertiary amines.

6 Acknowledgements

Support for this work was provided by the Natural Sciences and Engineering Research Council of Canada (Strategic Program Grant No. STPGP 479466-15), and by the SHARCNET (Shared Hierarchical Academic Research Computing Network) HPC consortium (www.sharcnet.ca) and Compute Canada (www.computecanada.ca)
References

(1) Dutcher, B.; Fan, M.; Russell, A. G. Amine–based CO₂ capture technology development from the beginning of 2013: A Review. *ACS Appl. Materials & Interfaces* 2015, 7, 2137–2148.

(2) Xiao, M.; Liu, H.; Idem, R.; Tontiwachwuthikul, P.; Liang, Z. A study of structure–activity relationships of commercial tertiary amines for post–combustion CO₂ capture. *Appl. Energy* 2016, 184, 219–229.

(3) Bernhardsen, I. M.; Knuutila, H. K. A review of potential amine solvents for CO₂ absorption process: Absorption capacity, cyclic capacity and pKa. *Int. J. Greenhouse Gas Control* 2017, 61, 27–48.

(4) 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.

(5) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* 2005, 105, 2999–3094.
(6) 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.

(7) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants. *J. Phys. Chem. A* 2006, 110, 2493–2499.

(8) Pliego, J. R.; Riveros, J. M. Theoretical Calculation of pKa Using the cluster–continuum model. *J. Phys. Chem. A* 2002, 106, 7434–7439.

(9) 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.

(10) Zhang, S. A reliable and efficient first principles–based method for predicting pKa values. III. Adding explicit water molecules: Can the theoretical slope be reproduced and pKa values predicted more accurately? *J. Comput. Chem.* 2012, 33, 517–526.

(11) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. Extension of a temperature-dependent aqueous solvation model to compounds containing nitrogen, fluorine, chlorine, bromine, and sulfur. *J. Phys. Chem. B* 2008, 112, 3024–3039.

(12) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2011, 1, 699–709.

(13) Cox, C. E.; Phifer, J. R.; Da Silva, L. F.; Nogueira, G. G.; Ley, R. T.; OLoughlin, E. J.; Barbosa, A. K. P.; Rygelski, B. T.; Paluch, A. S. Combining MOSCED with molecular simulation free energy calculations or electronic structure calculations to develop an efficient tool for solvent formulation and selection. *J. Computer–aided Mol. Design* 2017, 31, 183–199.

(14) Miguel, E. L.; Silva, P. L.; Pliego, J. R. Theoretical prediction of pKa in methanol:
Testing SM8 and SMD models for carboxylic acids, phenols, and amines. *J. Phys. Chem. B* 2014, 118, 5730–5739.

(15) Ho, J.; Klamt, A.; Coote, M. L. Comment on the correct use of continuum solvent models. *J. Phys. Chem. A* 2010, 114, 13442–13444.

(16) Haworth, N. L.; Wang, Q.; Coote, M. L. Modeling flexible molecules in solution: A pKa Case Study. *J. Phys. Chem. A* 2017, 121, 5217–5225.

(17) Bochevarov, A. D.; Watson, M. A.; Greenwood, J. R.; Philipp, D. M. Multiconformation, density functional theory–based pKa prediction in application to large, flexible organic molecules with diverse functional groups. *J. Chem. Theory Comput.* 2016, 12, 6001–6019.

(18) Xu, L.; Coote, M. L. Methods to Improve the Calculations of SMD Solvation Free Energies and Associated Aqueous pKa Values: Comparison between Choosing an Optimal Theoretical Level, Solute Cavity Scaling and Using Explicit Solvent Molecules. *J. Phys. Chem. A* 2019,

(19) Xu, L.; Coote, M. L. Improving the Accuracy of PCM-UAHF and PCM-UAKS Calculations Using Optimized Electrostatic Scaling Factors. *Journal of chemical theory and computation* 2019,

(20) others,, et al. Empirically Fitted Parameters for Calculating pK a Values with Small Deviations from Experiments Using a Simple Computational Strategy. *Journal of chemical information and modeling* 2016, 56, 1714–1724.

(21) Sutton, C. C.; Franks, G. V.; da Silva, G. First principles pKa calculations on carboxylic acids using the SMD solvation model: effect of thermodynamic cycle, model chemistry, and explicit solvent molecules. *J. Phys. Chem. B* 2012, 116, 11999–12006.
(22) Gupta, M.; da Silva, E. F.; Svendsen, H. F. Explicit solvation shell model and continuum solvation models for solvation energy and pKa determination of amino acids. *J. Chem. Theory Comput.* **2013**, *9*, 5021–5037.

(23) Gupta, M.; da Silva, E. F.; Svendsen, H. F. Postcombustion CO$_2$ capture solvent characterization employing the explicit solvation shell model and continuum solvation models. *J. Phys. Chem. B* **2016**, *120*, 9034–9050.

(24) Prasad, S.; Huang, J.; Zeng, Q.; Brooks, B. R. An explicit–solvent hybrid QM and MM approach for predicting pKa of small molecules in SAMPL6 challenge. *J. Computer–aided Molec. Des.* **2018**, *32*, 1191–1201.

(25) Leung, K. Surface Potential at the Air- Water Interface Computed Using Density Functional Theory. *J. Phys. Chem. Lett.* **2009**, *1*, 496–499.

(26) Nezbeda, I.; Moučka, F.; Smith, W. R. Recent progress in molecular simulation of aqueous electrolytes: Force fields, chemical potentials and solubility. *Molec. Phys.* **2016**, *114*, 1665–1690.

(27) Noroozi, J.; Smith, W. R. An efficient molecular simulation methodology for chemical reaction equilibria in electrolyte solutions: Application to CO$_2$ reactive absorption. *J. Phys. Chem. A* **2019**, *123*, 4074–4086.

(28) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high–throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(29) Gupta, M.; da Silva, E. F.; Svendsen, H. F. Modeling temperature dependency of amine basicity using PCM and SM8T implicit solvation models. *J. Phys. Chem. B* **2012**, *116*, 1865–1875.
(30) others., et al. The pKa values of amine based solvents for CO₂ capture and its temperature dependence. An analysis by density functional theory. *Int. J. Greenhouse Gas Control* **2017**, *58*, 62–70.

(31) Xue, X.-S.; Wang, Y.; Yang, C.; Ji, P.; Cheng, J.-P. Toward prediction of the chemistry in ionic liquids: An accurate computation of absolute pKa values of benzoic acids and benzenethiols. *J. Org. Chem.* **2015**, *80*, 8997–9006.

(32) Harder, E.; Roux, B. On the origin of the electrostatic potential difference at a liquid–vacuum interface. *J. Chem. Phys.* **2008**, *129*, 12B613.

(33) Misin, M.; Fedorov, M. V.; Palmer, D. S. Hydration free energies of molecular ions from theory and simulation. *J. Phys. Chem. B* **2016**, *120*, 975–983.

(34) Zhang, H.; Jiang, Y.; Yan, H.; Cui, Z.; Yin, C. Comparative assessment of computational methods for free energy Calculations of ionic hydration. *J. Chem. Inf. Modeling* **2017**, *57*, 2763–2775.

(35) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. The proton’s absolute aqueous enthalpy and Gibbs free energy of solvation from cluster–ion solvation data. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.

(36) 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.

(37) Beck, T. L. The influence of water interfacial potentials on ion hydration in bulk water and near interfaces. *Chem. Phys. Lett.* **2013**, *561*, 1–13.

(38) Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. Absolute hydration free energies of ions, ion–water clusters, and quasichemical theory. *J. Chem. Phys.* **2003**, *119*, 2702–2708.
(39) Spartan18, Wavefunction Inc., Irvine CA. 2018.

(40) others,, et al. Gaussian 16, Revision A. 03, Gaussian. Inc., Wallingford CT 2016,

(41) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* 2001, 25, 1157–1174.

(42) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 1993, 97, 10269–10280.

(43) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 2002, 23, 1623–1641.

(44) Zhang, Y.; Zhang, Y.; McCready, M. J.; Maginn, E. J. Evaluation and refinement of the general AMBER Force Field for nineteen pure organic electrolyte solvents. *J. Chem. Eng. Data* 2018, 63, 3488–3502.

(45) Barbosa, N. S.; Zhang, Y.; Lima, E. R.; Tavares, F. W.; Maginn, E. J. Development of an AMBER-compatible transferable force field for poly (ethylene glycol) ethers (glymes). *J. Mol. Modeling* 2017, 23, 194.

(46) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.* 2009, 5, 350–358.

(47) others,, et al. AmberTools 16, University of California, San Francisco, 2016. *There is no corresponding record for this reference.*[Google Scholar]

(48) da Silva, A. W. S.; Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC research notes* 2012, 5, 367.
(49) Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(50) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comp. Phys.* **1976**, *22*, 245–268.

(51) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical physics letters* **1994**, *222*, 529–539.

(52) McQuarrie, D. Statistical mechanics University Science Books. *Sausalito, CA* **2000**,.

(53) Lamoureux, G.; Roux, B. Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. *J. Phys. Chem. B* **2006**, *110*, 3308–3322.

(54) Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. Effect of Electrostatic Force Truncation on Interfacial and Transport Properties of Water. *J. Phys. Chem. A* **1996**, *100*, 17011–17020.

(55) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small–molecule solvation free energies. *J. Phys. Chem. B* **2009**, *113*, 4538–4543.

(56) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357–1377.

(57) Mobley, D. L.; Guthrie, J. P. FreeSolv: A database of experimental and calculated hydration free energies, with input files. *J. Computer-aided Mol. Design* **2014**, *28*, 711–720.
(58) Riquelme, M.; Lara, A.; Mobley, D. L.; Verstraelen, T.; Matamala, A. R.; Vöhringer-Martinez, E. Hydration free energies in the FreeSolv Database calculated with polarized iterative Hirshfeld charges. *J. Chem. Inf. Modeling* 2018, 58, 1779–1797.

(59) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* 2007, 111, 2242–2254.

(60) Jämbeck, J. P.; Mocci, F.; Lyubartsev, A. P.; Laaksonen, A. Partial atomic charges and their impact on the free energy of solvation. *J. Comput. Chem.* 2013, 34, 187–197.

(61) Hunter, E. P.; Lias, S. G. Evaluated gas phase basicities and proton affinities of molecules: an update. *J. Phys. Chem. Ref. Data* 1998, 27, 413–656.

(62) BELL, R. Dissociation constants of organic bases in aqueous solutions. 1965.

(63) Puxty, G.; Rowland, R.; Allport, A.; Yang, Q.; Bown, M.; Burns, R.; Maeder, M.; Attalla, M. Carbon dioxide postcombustion capture: a novel screening study of the carbon dioxide absorption performance of 76 amines. *Env. Sci. & Technology* 2009, 43, 6427–6433.

(64) Rayer, A. V.; Sumon, K. Z.; Jaffari, L.; Henni, A. Dissociation constants (pKa) of tertiary and cyclic amines: Structural and temperature dependences. *J. Chem. Eng. Data* 2014, 59, 3805–3813.

(65) Taggeri, A.; Mohamedali, M.; Henni, A. Dissociation Constant (p Ka) and Thermodynamic Properties of Some Tertiary and Cyclic Amines from (298 to 333) K. *J. Chem. Eng. Data* 2015, 61, 247–254.

(66) Hamborg, E. S.; Versteeg, G. F. Dissociation constants and thermodynamic properties of amines and alkanolamines from (293 to 353) K. *J. Chem. Eng. Data* 2009, 54, 1318–1328.
(67) Yamada, H.; Shimizu, S.; Okabe, H.; Matsuzaki, Y.; Chowdhury, F. A.; Fujioka, Y. Prediction of the basicity of aqueous amine solutions and the species distribution in the amine–H2O–CO2 system using the COSMO-RS method. Ind. Eng. Chem. Res. 2010, 49, 2449–2455.

(68) Aggarwal, R.; Ranganathan, P. Common pitfalls in statistical analysis: The use of correlation techniques. Perspect Clin Res 2016, 7, 187–190.

(69) Ranganathan, P.; Pramesh, C. S.; Aggarwal, R. Common pitfalls in statistical analysis: Measures of agreement. Perspect Clin Res 2017, 8, 187–191.

(70) Brandt, S. Data Analysis: Statistical and Computational Methods for Scientists and Engineers; Springer: New York, 1999.

(71) Narku-Tetteh, J.; Muchan, P.; Idem, R. Effect of alkanol chain length of primary alkanolamines and alkyl chain length of secondary and tertiary alkanolamines on their CO2 capture activities. Separation and Purification Technology 2017, 187, 453–467.

(72) Zhang, H.; Jiang, Y.; Yan, H.; Yin, C.; Tan, T.; van der Spoel, D. Free-energy calculations of ionic hydration consistent with the experimental hydration free energy of the proton. J. Phys. Chem. Lett. 2017, 8, 2705–2712.

(73) Jia, X.; Li, P. Solvation Free Energy Calculation Using a Fixed-Charge Model: Implicit and Explicit Treatments of the Polarization Effect. The Journal of Physical Chemistry B 2019, 123, 1139–1148.

(74) Mecklenfeld, A.; Raabe, G. Comparison of RESP and IPolQ-Mod Partial Charges for Solvation Free Energy Calculations of Various Solute/Solvent Pairs. Journal of chemical theory and computation 2017, 13, 6266–6274.

(75) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of fixed partial charges
for amino acids accommodating a specific water model and implicit polarization. *The Journal of Physical Chemistry B* 2013, 117, 2328–2338.
SMD solvation model

AM1-BCC
MP2 cc-pVTZ+ PCM
B3LYP/6-311++G(d,p)
HF/6-31G*

GAFF

SMD solvation model
