


## RESEARCH ARTICLE

# $pK_a$ prediction in non-aqueous solvents

Jonathan W. Zheng<sup>1</sup>  | Emad Al Ibrahim<sup>1</sup>  | Ivari Kaljurand<sup>2</sup>  | Ivo Leito<sup>2</sup>  | William H. Green<sup>1</sup> <sup>1</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA<sup>2</sup>Institute of Chemistry, University of Tartu, Tartu, Estonia**Correspondence**William H. Green, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.  
Email: [whgreen@mit.edu](mailto:whgreen@mit.edu)**Funding information**

Machine Learning for Pharmaceutical Discovery and Synthesis Consortium; King Abdullah University of Science and Technology; Ministry of Education and Research, Republic of Estonia, Grant/Award Number: TK210; Estonian Research Council, Grant/Award Number: PRG690

**Abstract**

Acid dissociation constants ( $pK_a$ ) are widely measured and studied, most typically in water. Comparatively few datasets and models for non-aqueous  $pK_a$  values exist. In this work, we demonstrate how the  $pK_a$  in one solvent can be accurately determined using reference data in another solvent, corrected by solvation energy calculations from the COSMO-RS method. We benchmark this approach in 10 different solvents, and find that  $pK_a$  values calculated in six solvents deviate from experimental data on average by less than 1  $pK_a$  unit. We observe comparable performance on a more diverse test set including amino acids and drug molecules, with higher error for large molecules. The model performance in four other solvents is worse, with one MAE exceeding 3  $pK_a$  units; we discuss how such errors arise due to both model error and inconsistency in obtaining experimental data. Finally, we demonstrate how this technique can be used to estimate the proton transfer energy between different solvents, and use this to report a value of the proton's solvation energy in formamide, a quantity that does not have a consensus value in literature.

**KEYWORDS**acid dissociation, COSMO-RS, ions,  $pK_a$ , solvation

## 1 | INTRODUCTION

The acid dissociation constant, or  $pK_a$ , is implicated in the behavior of pharmaceutical drugs in the human body, environmental impact of molecules, and other applications of chemistry.<sup>1–3</sup> In recent years, several models<sup>4–9</sup> and open-source data compilations<sup>10–15</sup> have been developed for aqueous  $pK_a$  predictions. Such developments in non-aqueous solvents are comparatively fewer.<sup>16,17</sup>

In light of the fewer available data, a variety of approaches have been proposed to calculate non-aqueous  $pK_a$  values. One approach is to train a neural network in a single solvent by augmenting the training set with computed values.<sup>18</sup> Another approach involves training a deep learning model simultaneously across multiple solvents. This method was used on the iBonD dataset, which includes more than 30,000  $pK_a$  data distributed across 46 solvents.<sup>17</sup> This approach was

reported to score an overall MAE of 0.89  $pK_a$  units, though the errors associated with specific solvents were higher, with MAEs in DMSO and acetonitrile respectively exceeding 1.5 and 1.2  $pK_a$  units.<sup>19</sup> The full data corpus, although accessible through a website and much larger than all other non-aqueous  $pK_a$  compilations, is not available in a convenient format for data science applications. Therefore, the potential of further extending machine learning approaches is also limited by the low availability of open-source experimental data.

One other class of models is to relate the  $pK_a$  in the desired solvent to a calculable or measurable energy difference. In this way, the  $pK_a$  is accessed via a thermodynamic cycle. A variety of relations have been proposed, relating  $pK_a$  to computed Gibbs free energies of dissociation,<sup>18,20,21</sup> computed  $pK_a$  values,<sup>22</sup> and experimental  $pK_a$  values in the same solvent<sup>23</sup> or a reference solvent (typically water).<sup>24–26</sup> In all of these approaches, a linear model is constructed

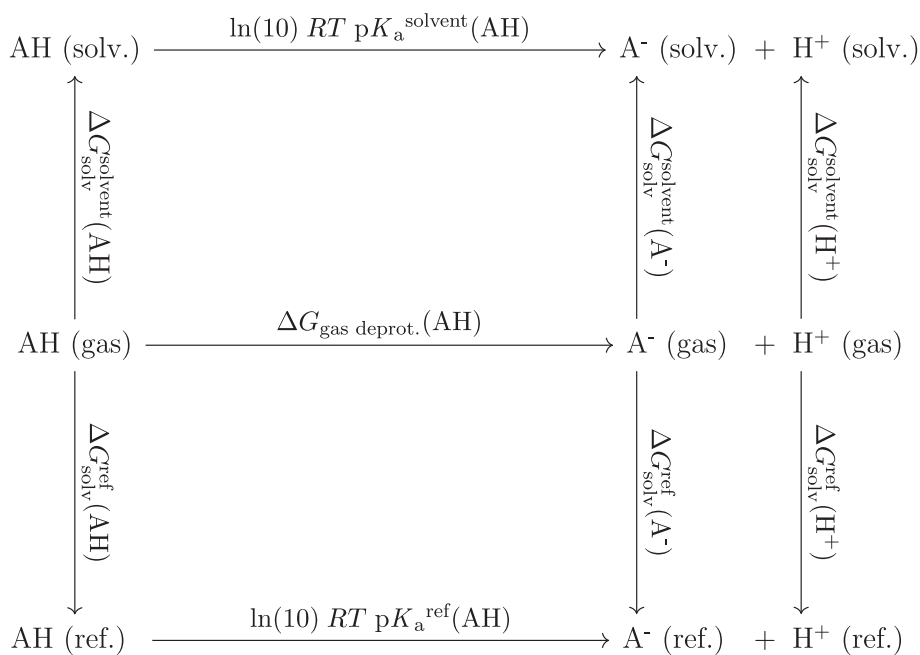
This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

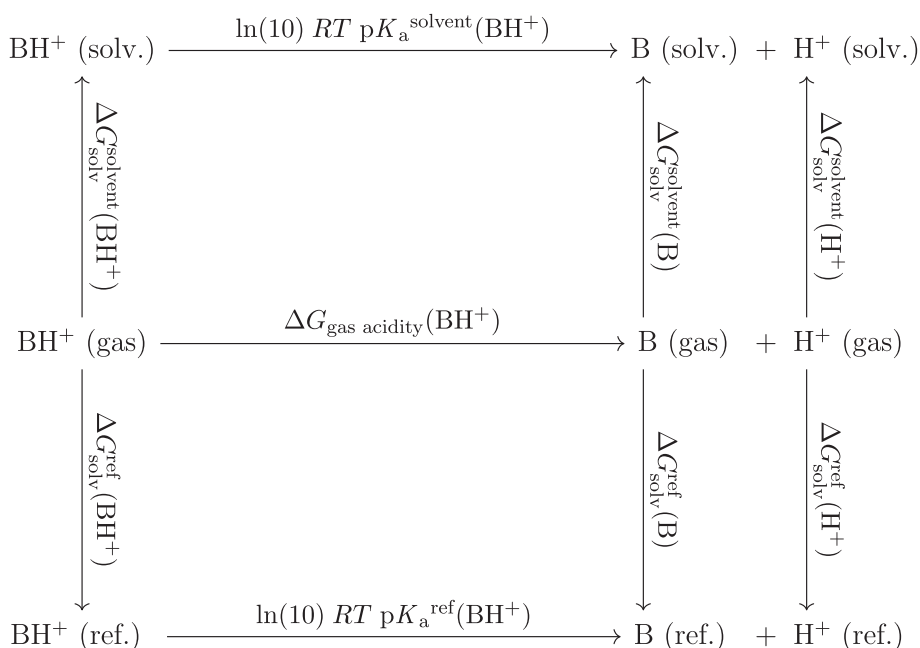
wherein empirical linear parameters are calculated. A varying degree of error is introduced in the choice of regression variable. Using the  $pK_a$  of the analyte in water as the regression variable has the benefit of low data uncertainty, as experimental aqueous  $pK_a$  values are often reported to less than 0.2  $pK_a$  units.<sup>12–15</sup> Therefore, we chose to use the reference solvent method in this work with water as the reference solvent.

It is sometimes useful to utilize the thermodynamic cycle of gas-phase dissociation followed by solvation, shown in Figures 1 and 2. This formulation links the  $pK_a$  to the solvation energy of the proton, which has only been measured in a handful of solvents<sup>26,27</sup> but is

required to establish the pH scale in different solvents. For non-aqueous  $pK_a$  data, such proton solvation free energies are either implicitly assumed, or explicitly assigned. For many solvents, this term can be recovered from the  $pK_a$  data if treated as a regression parameter. For example, Rossini et al.<sup>28,29</sup> estimated the proton's solvation energy in water, acetonitrile, methanol, and acetone by using a linear regression comparing computed and experimental  $pK_a$  values. Then, they used those proton solvation energies with the reference solvent method to compute  $pK_a$  values with RMSDs within 0.8  $pK_a$  units.<sup>25</sup> In their works, the authors used the electrostatic solvation model SOLVATE within the MEAD software suite.<sup>31,32</sup>



**FIGURE 1** Thermodynamic cycle for the dissociation of a neutral acid in different phases.



**FIGURE 2** Thermodynamic cycle for the dissociation of a cation acid in different phases.

In this work, we benchmark a procedure for calculating  $pK_a$  values in 10 different solvents by reference to data in water. We use the conductor-like screening model COSMO-RS, which has previously been demonstrated to successfully compute  $pK_a$  values in a handful of solvents such as acetonitrile<sup>20</sup> and acetone.<sup>21</sup> We show that this approach only requires one regression parameter, which can be used to approximate the proton's solvation energy in each new solvent. One considerable obstacle in predicting non-aqueous  $pK_a$  values is the issue of experimental error.  $pK_a$  values are often measured relative to each other and then “anchored” to a reference value, due to the ambiguity surrounding the solvation free energy of the proton, as well as experimental considerations especially in weakly-screening solvents. An important consideration is solvent purity, in particular water content, which can significantly affect  $pK_a$  measurements especially in low-polarity aprotic solvents. The choice of reference value can lead to significant inconsistencies among data sources; discrepancies of 3  $pK_a$  units, and sometimes more, are not uncommon.<sup>21</sup> Experimental error can also arise if ion pairing (wherein ions form pairs or clusters, especially in weakly-screening solvents) and homoconjugation (wherein anions may form adducts with neutral acids) are not accounted for.<sup>33</sup> There is also systematic error in solvation free energy computations for ionic compounds, which are present in this workflow.<sup>34</sup> Hence, model error must inextricably be considered alongside experimental error, and the relative contribution of each is not always clear. In this work, we attempt to correct for these issues by curating experimental data. We discuss a few examples wherein remaining inconsistencies in experimental data appear to be responsible for significant deviations from the model predictions, as well as several cases in which the situation is less clear.

## 2 | METHODS

### 2.1 | Dataset selection

We utilized  $pK_a$  data presented by Busch and collaborators, which includes a compilation of  $pK_a$  data from the iBonD dataset.<sup>22</sup> From their data spanning 18 solvents, we chose solvent systems with at least 10 data points and with representation of both acids and bases, resulting in a set of 10 solvents. We further modified the dataset by removing doubly-charged species, and selected only species with both aqueous and non-aqueous  $pK_a$  values. The selected compounds are all small molecules, consisting nearly entirely of substituted benzoic and phenolic acids, alkyl carboxylic acids, alkylamines, and pyridine and aniline derivatives.

We then critically curated the data to ensure consistency and accuracy of the data in all solvents. In acetonitrile, DMSO, DMF, pyridine, and acetone, we replaced some values from the Busch collection with trusted values from the literature. More information about the data curation can be found in the [Supporting Information](#).

**TABLE 1** Datapoints per solvent.

Solvent	# of acids	# of bases
Acetone	15	4
Acetonitrile	20	27
DCE	12	3
DMF	17	5
DMSO	30	24
Ethanol	19	14
Formamide	9	12
Methanol	24	27
Nitromethane	5	17
Pyridine	12	0

Table 1 shows the number of data available per solvent. All solvents include at least 10 data points. Some solvent systems (acetonitrile, DMSO, and methanol) include more than 45 datapoints and are split roughly half-half between acids and bases.

### 2.2 | Calculating $pK_a$ in different solvents

The  $pK_a$  in one solvent is calculated by applying solvent corrections to  $pK_a$  data in a reference solvent (water). For acids, this relationship is:

$$pK_a^{\text{solvent}} = pK_a^{\text{ref}} + \frac{1}{\ln(10)RT} (\Delta\Delta G_{\text{solv}}^{\ddagger}(\text{H}^+) + \Delta\Delta G_{\text{solv}}^{\ddagger}(\text{A}^-) - \Delta\Delta G_{\text{solv}}^{\ddagger}(\text{AH})), \quad (1)$$

and for bases,

$$pK_a^{\text{solvent}} = pK_a^{\text{ref}} + \frac{1}{\ln(10)RT} (\Delta\Delta G_{\text{solv}}^{\ddagger}(\text{H}^+) - \Delta\Delta G_{\text{solv}}^{\ddagger}(\text{BH}^+) + \Delta\Delta G_{\text{solv}}^{\ddagger}(\text{B})), \quad (2)$$

where  $\Delta\Delta G_{\text{solv}}^{\ddagger}(Z) \equiv \Delta G_{\text{solv}}^{\text{solvent}}(Z) - \Delta G_{\text{solv}}^{\text{ref}}(Z)$ . In these equations,  $RT$  is the product of the molar gas constant with temperature. The temperature is assumed to be 298 K.  $\Delta G_{\text{solv}}(Z)$  is the solvation energy of a solute at an arbitrary reference state.  $\text{A}^-$  refers to an anionic base and  $\text{AH}$  refers to its protonated conjugate acid. Likewise,  $\text{BH}^+$  refers to a cationic acid and  $\text{B}$  refers to its deprotonated conjugate base.  $pK_a^{\text{solvent}}$  refers to the dissociation constant of a molecule in a non-aqueous solvent of interest (i.e., the solvents in Table 3), and  $pK_a^{\text{ref}}$  refers to the dissociation of that same species in water. In this work, the  $\ddagger$  superscript denotes a reference value of water.

Because  $\Delta\Delta G_{\text{solv}}^{\ddagger}(\text{H}^+)$  is not known, we treated the term as a regression parameter,  $\delta_{\text{H}^+}^{\ddagger}$ , that minimizes the absolute residual of the fit for each solvent:

$$f_i(\delta) = \begin{cases} \text{p}K_{\text{a}i}^{\text{ref}} - \text{p}K_{\text{a}i}^{\text{solvent}} + \frac{1}{\ln(10)RT} (\delta - \Delta\Delta G_{\text{solv}}(\text{BH}_i^+) + \Delta\Delta G_{\text{solv}}(\text{B}_i)) & \text{p}K_{\text{a}i} = \text{basic} \\ \text{p}K_{\text{a}i}^{\text{ref}} - \text{p}K_{\text{a}i}^{\text{solvent}} + \frac{1}{\ln(10)RT} (\delta - \Delta\Delta G_{\text{solv}}(\text{A}_i^-) + \Delta\Delta G_{\text{solv}}(\text{AH}_i)) & \text{p}K_{\text{a}i} = \text{acidic} \end{cases}, \quad (3)$$

$$\delta_H^\ddagger = \text{argmin}_\delta \left( \sqrt{\sum_i^N f_i(\delta)^2} \right), \quad (4)$$

where  $i$  refers to the index of the neutral form of the acid or base under consideration,  $N$  is the number of datapoints per solvent,  $\delta$  is a proxy estimate for the proton transfer energy, and  $f_i(\delta)$  is the loss function that corresponds to the optimal  $\delta_H^\ddagger$  when minimized. In Equation (3), the  $\text{p}K_{\text{a}}$  terms are experimental values, whereas the  $\Delta\Delta G_{\text{solv}}$  terms are computed using COSMO-RS.

One advantage to this method is that it does not require any assumption of the absolute reference solvation energy  $\Delta G_{\text{solv}}^{\text{ref}}(\text{H}^+)$ , which is prone to significant error. Additionally, the solvation terms are likely to include cancellation of model error, as one  $\text{p}K_{\text{a}}$  calculation involves energy differences between different solvents (in each  $\Delta\Delta G_{\text{solv}}$  term) and also within the same solvent (by taking the difference of solvation energies between the acid and its conjugate base).

Each  $\Delta\Delta G_{\text{solv}}$  term describes the partitioning of the solutes between different phases—this energy difference is also termed the *transfer free energy*, and is directly proportional to  $\log(P)$ . The COSMO-RS method has been previously benchmarked against the SAMPL challenges for  $\log(P)$  of small neutral drug-like molecules, with root mean squared deviations of around 0.7  $\text{p}K_{\text{a}}$  units overall, and with model performance depending on the solvent.<sup>36,37</sup> One such study observed higher errors for  $\log(D)$ , which accounts for ionization.<sup>38</sup> The SM8 implicit solvation model has been evaluated on  $\log(P)$  for organic solutes between water and other solvents, with mean unsigned errors reported to exceed 3 log units.<sup>39,40</sup> Given the high uncertainty in computing transfer free energies, particularly for ionic solutes, the accuracy of this method is dependent on the degree of error cancellation.

If  $\text{p}K_{\text{a}}$  data for a compound is available in multiple solvents, it is possible to compute the  $\text{p}K_{\text{a}}$  using an ensemble of reference values rather than using a single reference. The fitting procedure can also be done simultaneously with multiple references utilizing all available data for a given solvent (see [Supporting Information](#)).

## 2.3 | QM details

A large set of initial conformers of each molecule was constructed using CREST<sup>41</sup> with GFN2-xTB version 6.6.0<sup>42</sup> using the analytical linearized Poisson-Boltzmann (ALPB) model<sup>43</sup> with water as the solvent. The 20 lowest-energy conformers for each species were

re-optimized using TURBOMOLE v7.7<sup>44</sup> at the BP86<sup>45,46</sup>/def2-TZVP<sup>47</sup> level of theory with the COSMO model at the conductor limit ( $\epsilon = \infty$ ). These 20 conformers were further filtered by pruning conformer geometries based on dihedral angles, removing conformers that matched dihedral angles with a maximum absolute deviation less than 10 degrees and a root mean square deviation within 20 degrees. The conformers were also screened using the Python package RDMC<sup>48</sup> to ensure that their 3D geometries correspond to their molecular graphs. The  $\sigma$ -profiles and single-point gas-phase energies of the remaining optimized conformers were then calculated using TURBOMOLE at the BP-TZVPD-FINE level. COSMO-RS was then used to calculate the solvation energies at 298 K via COSMOTHERM 2023.<sup>49,50</sup>

Optimized geometries were then manually inspected to ensure that they converged to reasonable structures. Geometries for all species were constructed independently using this approach (i.e., optimized conformers for neutral species were not used to generate initial geometries for ionic conjugate acids/bases). The optimized geometries are included in the [SI](#).

## 3 | RESULTS

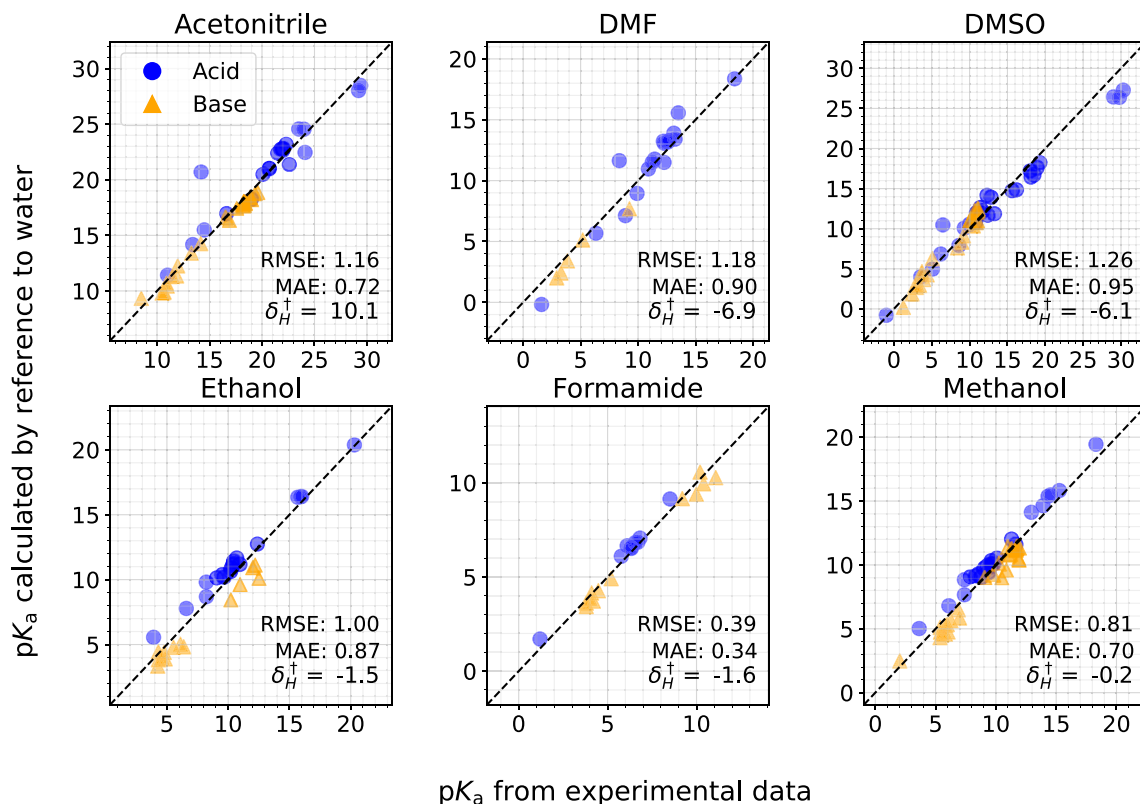
### 3.1 | Evaluation of model on parameterization data

The  $\text{p}K_{\text{a}}$  of acids and bases in six solvents (acetonitrile, DMF, DMSO, ethanol, formamide, and methanol) were calculated to MAEs within 1  $\text{p}K_{\text{a}}$  unit (Figure 3). Model errors were higher in acetone, nitromethane, pyridine, and dichloroethane. A summary of the error statistics is shown in Table 2. The six best-performing solvents all include  $R^2$  values exceeding 0.9, MAEs less than 1  $\text{p}K_{\text{a}}$  unit, and RMSEs less than 1.3  $\text{p}K_{\text{a}}$  units. The best performance was seen in formamide, with an RMSE of 0.39  $\text{p}K_{\text{a}}$  units.

The four worst-performing solvents fared worse in every metric, reflecting both worse agreement for the majority of predictions and the presence of more numerous outliers. The results for these four solvents, along with reasons for the high error, are described later in this manuscript.

### 3.2 | Evaluation of model on external test data

For each of the six solvents shown in Figure 3, we obtained additional experimental  $\text{p}K_{\text{a}}$  data, filtering out any data that were used to fit the  $\delta_H^\ddagger$  parameters.



**FIGURE 3**  $pK_a$  values calculated relative to a reference  $pK_a$ , for the six solvents with MAEs less than 1  $pK_a$  unit. Blue circles refer to acidic  $pK_a$  (involving anions) and orange triangles to basic  $pK_a$  (involving cations). The RMSE and MAE are shown in  $pK_a$  units, and the corresponding  $\delta_H^\ddagger$  in  $\text{kcal mol}^{-1}$ .

**TABLE 2** Error Statistics of  $\delta_H^\ddagger$  and  $pK_a$ .

Solvent	$\delta_H^\ddagger$ ( $\text{kcal mol}^{-1}$ )	RMSE	MAE	$R^2$
Acetonitrile	10.1	1.16	0.72	0.94
DMF	-6.9	1.18	0.90	0.92
DMSO	-6.1	1.26	0.95	0.96
Ethanol	-1.5	1.00	0.87	0.93
Formamide	-1.6	0.39	0.34	0.98
Methanol	-0.2	0.81	0.70	0.93
Acetone	1.3	2.18	1.88	0.77
Nitromethane	9.4	2.37	1.96	0.67
DCE	9.5	3.98	3.51	0.32
Pyridine	-13.6	2.12	1.43	0.43

### 3.3 | Description of test data

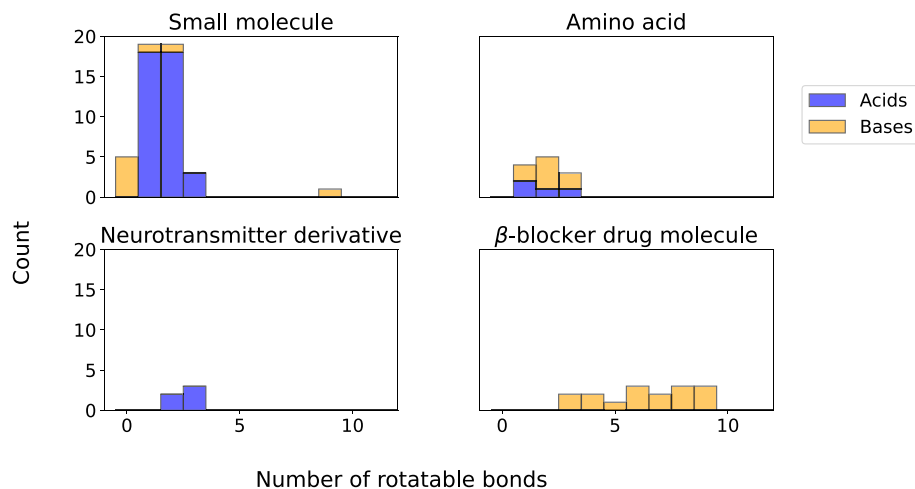
Data were selected from seven sources:

- Zevatskii (2009):<sup>51</sup> Various small acids and bases: 11 in water, six in methanol, seven in ethanol, and 11 in formamide.
- Cantu (2005):<sup>52</sup> Four basic drug molecules and their  $pK_a$  values in water, methanol, and acetonitrile.
- Sirén (2005):<sup>53</sup> Five acidic neurotransmitter derivatives and 12 basic  $\beta$ -blocker drugs with  $pK_a$  values in water and methanol.

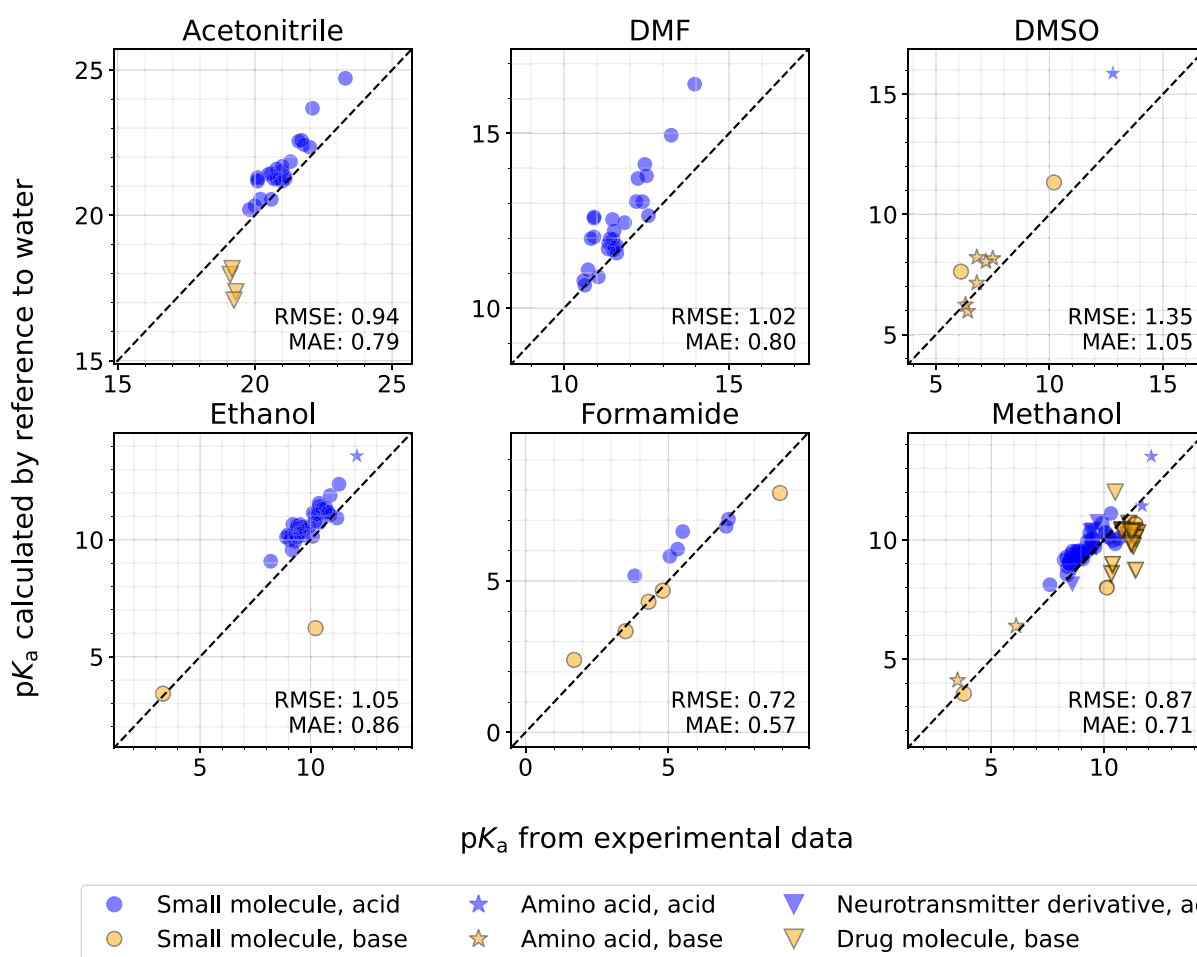
- Headley (1994):<sup>54</sup> Eight simple carboxylic acids with values in water, methanol, and ethanol.
- Ludwig (1986):<sup>55</sup> Benzoic acid derivatives and their  $pK_a$  values: 27 in water, 26 in methanol, 27 in ethanol, 27 in DMF, and 25 in acetonitrile.
- Hughes (1986):<sup>56</sup> Amino acids and two basic small molecules, totaling six values in water and 9 in DMSO. The remaining entries without aqueous data were supplemented by aqueous  $pK_a$  values from the IUPAC Digitized  $pK_a$  Dataset.<sup>11</sup>
- Ogston (1935):<sup>57</sup> Three amino acids and one small molecule, in water and methanol, and one amino acid in methanol.

There were two cases in which different values were available from two sources. The maximum difference in  $pK_a$  was 0.4  $pK_a$  units, so we used the average of their values in these two cases. We then further curated the data as was done for the parameterization set (see S1).

The test compounds include small acids and bases, much like the set of compounds used for parameterization, but also includes drug molecules with many rotatable bonds as well as amino acids, which we represent in their zwitterionic form in the QM calculations. Figure 4 shows the number of rotatable bonds for the molecules in the test set; all but one of the small molecules have fewer than four, whereas nearly all of the basic drug molecules have four or greater.



**FIGURE 4** Distribution of the number of rotors in the test sets, demarcated by the class of molecule. The small molecule with the large number of rotors is tributylamine.



**FIGURE 5** Test of  $pK_a$  values calculated relative to their experimental data, for the six solvents depicted in Figure 3. The MAE and RMSE are shown in  $pK_a$  units.

Using the same modeling methodology as described previously, and using the shift parameters obtained from the previous fitting (Table 2), we computed the  $pK_a$  values and compared them to the literature data.

### 3.3.1 | Test results

Figure 5 shows the parity between our predictions and the test data. The predicted  $pK_a$  values show satisfactory agreement in most

solvents, especially for small molecules. In fact, for solvents with only small molecules tested, we observed RMSEs of less than 1.1 pK<sub>a</sub> unit. The best performance was in formamide, with an RMSE of 0.72 pK<sub>a</sub> units. Also, in DMF, formamide, and DMSO, the pK<sub>a</sub> values come from only one source, which further minimizes the chance of systematic error from different acidity scales within the same solvent (such systematic error can be quite significant, on the order of several kcal mol<sup>-1</sup>, and is discussed later). Hence, the remaining uncertainty is due to the quantum chemical calculations as well as any inherent aleatoric uncertainty in the reference data.

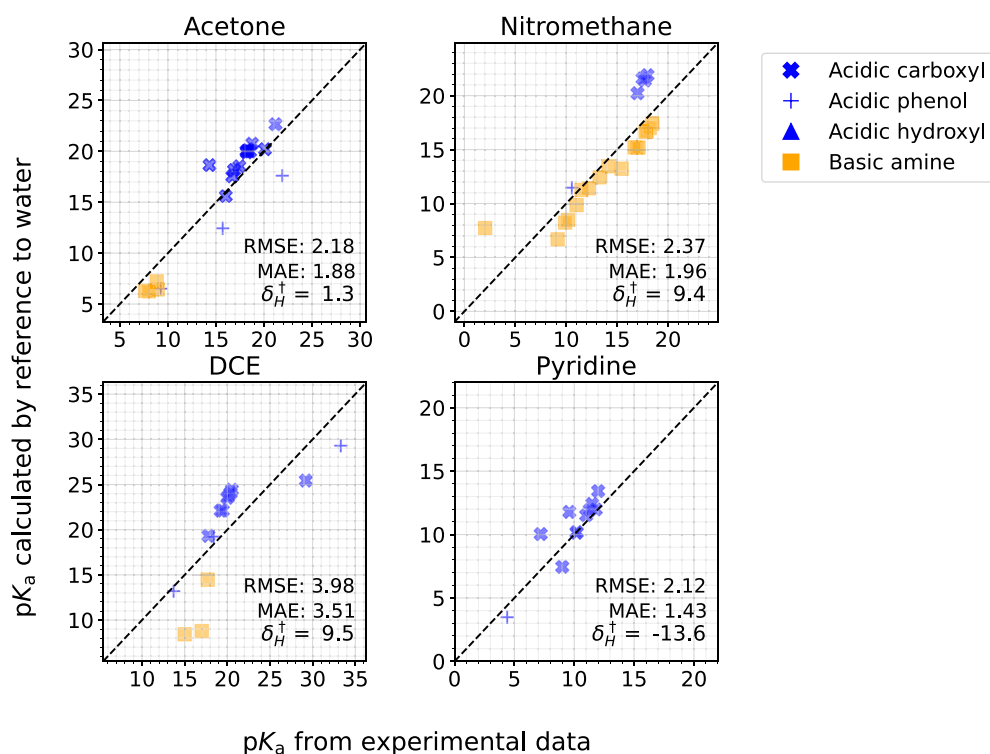
We focus here on a few outliers. First, in acetonitrile and methanol, several basic drug molecules showed a significant underprediction compared to the acids, which were generally overpredicted. This follows a pattern seen in Figure 3, where the same behavior was generally seen for acids and bases, albeit to a lesser extent. Similarly, in ethanol, there was a large underprediction in the pK<sub>a</sub> of tributylamine, which has a large number of conformers (the rightmost bar in the small molecule section of Figure 4). Hence, the molecules with the largest number of rotors tended to have much larger errors using this method. Note that in our computational workflow, we limited the number of conformers in the COSMO-RS calculation to 20 or fewer, as finding numerous thermodynamically significant conformers for very large molecules is a time-expensive task. However, it is not clear whether conformer effects are primarily responsible for these errors, as experimental inconsistencies, COSMO-RS limitations based on theory or parameterization, and incomplete sets of conformers are all implicated in systematic error. Future work should elucidate the effect of conformers on the efficacy of this method.

Despite the underprediction of pK<sub>a</sub> values for large molecules, we observed generally good agreement between our calculations and the reference data, including for zwitterionic amino acids and neurotransmitter derivative molecules. We thereby recommend using this method to determine acidities and basicities of small organic molecules and amino acids in acetonitrile, DMF, DMSO, ethanol, formamide, and methanol.

### 3.4 | Sources of error

As discussed previously, this method was also used to parameterize acetone, nitromethane, DCE, and pyridine, but the model agreement was observed to be worse. Figure 6 shows the poor quality of the model predictions during the parameterization step.

This occurs in some part due to inconsistencies in experimental data. pK<sub>a</sub> data in non-aqueous solvents are prone to systematic errors; they can deviate due to inconsistent calibration methods, as pK<sub>a</sub> values in a solvent are typically measured with reference to each other and then anchored to an absolute scale. Another potential source of error comes from the solvation model, which previous work has shown can systematically differ in solvation free energies by several kcal mol<sup>-1</sup> based on the ionization center.<sup>34</sup> Although we curated the data to reduce the occurrence of data-sourced errors, it is possible that several errors are still present. Furthermore, there are experimental considerations such as ion pairing, which occurs in low-polarity solvents such as pyridine and DCE, and may further contribute to error.<sup>33</sup> On the basis of our data curation, we labeled the experimental data in nitromethane and DCE as unreliable, but still include our



**FIGURE 6** pK<sub>a</sub> values calculated relative to a reference pK<sub>a</sub>, for the four solvents with MAEs greater than 1 pK<sub>a</sub> unit. The RMSE and MAE are shown in pK<sub>a</sub> units, and the corresponding  $\delta_H^\ddagger$  in kcal mol<sup>-1</sup>. Data are colored by functional site of the ionization center.

predictions herein in the absence of more reliable values for demonstrative purposes. For a more rigorous discussion of the data curation, we refer readers to the SI.

### 3.4.1 | Experimental data for nitromethane

We observed that carboxylic acid  $pK_a$  values were overpredicted compared to the amine and phenol values (Figure 6).

The carboxylic acids are benzoic acid derivatives from the same single source, anchored to the  $pK_a$  value of 10.5 for picric acid. The  $pK_a$  used in this work for benzoic acid is 17.73<sup>56</sup> and for picric acid (the only phenol in this solvent) the  $pK_a$  was 10.5.<sup>59</sup> However, values reported elsewhere on the same acidity scale have included 13.2 for benzoic acid and 7.94 for picric acid.<sup>60,61</sup> The high deviations for carboxylic acids therefore cannot be immediately explained.

The amine values are anchored to a  $pK_a$  of 17.2 for 1,3-diphenylguanidine. This is possibly a different scale than the one for the acids. The Izutsu  $pK_a$  compilation<sup>16</sup> reports values of 13.2 for benzoic acid when anchored to this value of 17.2 for the amines; but the value for benzoic acid used in our work was 17.73, which would imply that the benzoic acids are systematically misaligned from amines by 4.5  $pK_a$  units. This is approximately the same amount by which acidic carboxylic acids were overpredicted in Figure 6. Therefore, we expect that anchoring the experimental data to the same scale in nitromethane would improve the performance of the COSMO-RS method.

Even if the acidity scales were to be aligned, the nitromethane data still overall remain questionable, as the shift parameters estimated herein are vastly different from estimates obtained from comparisons of transfer energies from water to nitromethane and from water to acetonitrile (see SI). Hence, we assess the nitromethane  $pK_a$  data in general to be highly unreliable and do not recommend its usage.

### 3.4.2 | Experimental data for DCE

In DCE, generally poor agreement was observed, with amines significantly underpredicted compared to carboxylic acids. The  $pK_a$  values for the carboxylic acids and picric acid come from the same collection of papers<sup>62-66</sup> (except for acetic acid at  $pH = 29.2$ , the outlier on the right of Figure 6), and are anchored to the  $pK_a$  of hydriodic acid = 7.9 (corresponding to  $pK_a$  of picric acid = 13.7). The  $pK_a$  values for the three bases are also anchored to the same value.<sup>67</sup> Therefore, most of the data in DCE is aligned. We could not determine the sources for the other two phenols and for acetic acid.

Note, further, that theoretical values for the dissociation of picric acid have been reported to be as high as 45.0, if the experiment is done with a completely pure and dry sample, which would strongly disagree with the scale followed by all other values herein.<sup>33</sup>

In summary, the data appear to be almost, if not fully, anchored to the same scale. Therefore, the issue of consistent anchor values cannot solely explain the poor model performance. DCE is known to

be a solvent with considerable experimental challenges including ion pairing, and therefore, we consider this set of data to also be unreliable.

### 3.4.3 | Model error

An unfortunate limitation of this method is that experimental data and model error cannot be decoupled. Therefore, there also remains the possibility that error from the solvation model (e.g., limitations of modeling ions, insufficient sampling of conformers) is a significant contributor to error. Previous work<sup>34,68</sup> has shown that COSMO-RS predictions of ionic solvation energies deviate from experimental predictions by an offset that is systematically too high for cations and too low for anions. A single parameter added to the model predictions (with a different sign depending on the ion's charge) accounted for most of that offset. However, it is not clear how much of the offset arises from uncertainty in anchoring the proton's solvation energy versus from model accuracy. Because the value of that optimal parameter changes significantly depending on the model type, we believe it possible that a portion of the systematic deviation is from model error. In Equations (1) and (2), the difference in anion solvation energies between two solvents is *added* while the difference for cations is *subtracted*; so, any such oppositely-signed systematic model error would manifest as a systematic offset in  $\delta_H^\ddagger$  as well. Previous work has also shown that there is systematic error based on the functional group at the ionization center; this may contribute to the poor prediction quality in some of the solvents that appear to be dependent on the type of acid.<sup>34</sup>

We emphasize here that our workflow lumps together model error with experimental error. This has two important consequences: (1) unambiguously discerning the cause of outliers is difficult, if not impossible, and (2) values of  $\delta_H^\ddagger$  may not be understood as accurate, empirical determinations of  $\Delta\Delta G_{\text{solv}}^\ddagger(\text{H}^+)$  but rather as estimates. Recovering a reliable value of the proton transfer energy based on  $\delta_H^\ddagger$  implies that  $\Delta\Delta G_{\text{solv}}^\ddagger(\text{BH}_i^+)$  and  $\Delta\Delta G_{\text{solv}}^\ddagger(\text{A}_i^-)$  computed using COSMO-RS have negligible systematic error, and all experimental  $pK_a$  data are anchored to consistent anchor values, both internally (i.e., the same acidity scale for all compounds in a single solvent) and among solvents (i.e., the chosen anchor value for each solvent is based off of sensible extrathermodynamic assumptions, whether implicitly or explicitly). These assumptions are certainly not perfectly valid, so to probe the strength of them, we compare our computed values of  $\delta_H^\ddagger$  to proton transfer energies derived using separate extrathermodynamic assumptions.

## 3.5 | Comparison of $\delta_H^\ddagger$ to proton transfer energies in literature

We compared our computed values of  $\delta_H^\ddagger$  to experimental values of  $\Delta\Delta G_{\text{solv}}^\ddagger(\text{H}^+)$  from water. In the most ideal case, the values of  $\delta_H^\ddagger$



**TABLE 3** Comparison of  $\delta_H^\ddagger$  to consensus proton transfer energies derived from experiment ( $\text{kcal mol}^{-1}$ ).

Solvent	$\delta_H^\ddagger$	$\Delta\Delta G_{\text{sol}}^\ddagger(\text{H}^+)$ (TATB) <sup>35</sup>	(CPA) <sup>27</sup>	(Rossini) <sup>29,30</sup>
Acetonitrile	10.1	10.7	5.7	10.8
DMSO	-6.1	-4.6	-7.4	-0.5
Ethanol	-1.5	2.7	-	-
Methanol	-0.2	2.1	2.4	0.0
DMF	-6.9	-3.4	-	-
Formamide	-1.6	-	-	-

should exactly equal the difference in solvation energies of the proton ( $\Delta\Delta G_{\text{sol}}^\ddagger(\text{H}^+)$ ). However, this is complicated by the fact that any systematic error in COSMO-RS corrections will also contribute to  $\delta_H^\ddagger$ , as will any systematic errors in  $\text{pK}_a$  reference data. Furthermore, the choice of anchor value in non-aqueous solvents is based on consistency with extrathermodynamic assumptions. Hence, any proton transfer energy that is recovered from this approach will agree most strongly with the value corresponding to the extrathermodynamic assumption used to anchor the scale. That said, comparing our computed values to proton transfer energies reported in the literature could at least provide a loose estimate for how much of the term  $\delta_H^\ddagger$  originates from correcting for the proton solvation energies versus from systematic error.

Most efforts in the literature are concentrated around just a few solvents, limiting the scope of our comparison.<sup>69</sup> Additionally, depending on the extrathermodynamic assumption invoked in fixing the energy scale, different values derived from experiments have been reported for each solvent.<sup>40</sup> Values have been reported using the tetraphenylarsonium-tetraphenylborate (TATB) assumption,<sup>70,71</sup> the cluster-pair approximation (CPA),<sup>27,28,72,73</sup> and by comparing experimental and computed  $\text{pK}_a$  values (Rossini and this work).<sup>24,26,29,30</sup> We refer interested readers to the corresponding references for more information about these approaches.

Table 3 shows that our  $\delta_H^\ddagger$  parameters showed general agreement with experimental transfer energies of the proton in several solvents. There is no single method or solvent that agrees perfectly with our values; for instance, our values for acetonitrile and methanol agree well with the Rossini values, but disagree with the transfer energy reported in DMSO. The ordering of very positive values (acetonitrile) and very negative values (DMSO, DMF) are in agreement across all methods, while the intermediate values show considerably greater deviation.

The Rossini estimates were derived using a similar approach to the one presented herein, but were computed by adding solvation energy corrections to quantum-chemical calculations rather than to reference  $\text{pK}_a$  values. Our estimates for  $\delta_H^\ddagger$  appeared to match closely with Rossini's estimates except for in the case of DMSO, in which case our estimates match the TATB and CPA estimates far more closely. Values for ethanol and DMF could only be compared to energies obtained from the TATB assumption. These estimates were also

comparatively worse than those in other solvents, with differences of  $4.2 \text{ kcal mol}^{-1}$  for ethanol and  $3.5 \text{ kcal mol}^{-1}$  for DMF compared to estimates using the TATB assumption. Though there are substantial deviations among the estimates for proton transfer energies, our computed values of  $\delta_H^\ddagger$  roughly agree with the proton transfer energies estimated in the literature. These results suggest that the  $\delta_H^\ddagger$  regression parameter generally captures the proton transfer energy, and can be used to approximate such energies in solvents where other estimates are unavailable.

Because we observed a good quality fit for  $\delta_H^\ddagger$  and good test performance in formamide, we assign an estimate of the proton transfer energy from water to formamide as  $-1.6 \text{ kcal mol}^{-1}$ .

To our knowledge, there is no consensus value for the proton's transfer energy in formamide for any given extrathermodynamic assumption. Values reported in the IUPAC compilation of transfer free energies from water include 1.0,  $-3.3$ ,  $-4.1$ , and  $-1.7 \text{ kcal mol}^{-1}$ , which on average satisfactorily agree with the value we report herein; however, these values were considered "unsatisfactory", and do not reflect a consensus value.<sup>74</sup> The transfer energy reported above should be considered as an estimate, not a definitive value, and reflect the underlying assumptions used to construct the pH scale in the reference data. We further advise against using the  $\delta_H^\ddagger$  obtained in this work to estimate proton transfer energies for acetone, nitromethane, pyridine, and DCE, as in these solvents, our linear regression was unable to determine the fitting parameters satisfactorily.

## 4 | CONCLUSION

We have discussed how experimental acid dissociation constants in one solvent can be combined with COSMO-RS solvation energy calculations to compute the  $\text{pK}_a$  in a different solvent. This technique requires an estimate of the proton's solvation energy in each solvent, and in this work we propose a way to estimate the needed proton transfer energies via regression. Our computed proton transfer energies are in rough agreement with other estimates in the literature. During parameterization, we observed residuals within 1  $\text{pK}_a$  unit of MAE for 6 solvents using this method. Errors are larger for 4 other solvents—in those cases, a large portion of error is likely due to poor data quality. Further, we tested this method to predict the acidities and basicities of an external set of compounds in acetonitrile, DMF, DMSO, ethanol, formamide, and methanol. We observed generally good agreement, with MAE  $< 1.1 \text{ pK}_a$  unit in all of those six solvents. The  $\text{pK}_a$  values of small molecules were predicted particularly accurately, but values for bases with large numbers of rotatable bonds were consistently underpredicted by several  $\text{pK}_a$  units. One potential reason is that the computational workflow used herein restricted the number of conformers in the COSMO-RS method to 20 or fewer; however, it is difficult to disentangle conformer effects from uncertainty due to experimental error and limitations of the implicit solvation model method. In addition, it is possible that performing the conformer search in only the aqueous phase instead of other solvents

may affect prediction accuracy. The potential for conformer search in different solvents to affect computed solvation free energies has been observed in previous work for neutral solutes,<sup>75</sup> and is not yet well-understood for ionic solutes. Future work should investigate the extent of these effects for pK<sub>a</sub> prediction, particularly in the four solvents that were identified to have poor performance.

We also used this method to estimate the transfer free energy of the proton from water to formamide. We emphasize that the proton transfer free energy is implicitly or explicitly assumed when establishing the pH scale in non-aqueous solvents, and thus depends entirely on the anchor values used in anchoring the acidity scales. One disadvantage to this method is that at least one reference pK<sub>a</sub> is required for each desired solvent system. Furthermore, potentially time-expensive QM calculations are required to generate the conformers used in the COSMO-RS calculation, especially for very large molecules. However, we are optimistic that ongoing advances in data availability and processing power will reduce the impact of these obstacles. Because the method described herein relies only on calculations of solvation free energies, any solvation model that can consider ionic solutes can be used. Future efforts should investigate the performance of this method when employing other solvation models, such as the implicit Minnesota SMx models<sup>76–78</sup> or cluster-continuum solvation methods.<sup>79</sup>

A key challenge for using this approach is the lack of high-quality reference data in many solvents. We hope that large datasets of dissociation constants in diverse solvent systems with unified acidity scales will become available, enabling the computation of useful thermochemical properties in more solvents. One other barrier is the lack of consensus around the use of extrathermodynamic assumptions in accurately determining proton solvation energies. We hope that future research can glean insight into the merits of the different extrathermodynamic assumptions and their effects in anchoring energy scales.

## ACKNOWLEDGMENTS

The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing HPC resources that have contributed to the research results reported within this paper. J. W. Z. and W. H. G. acknowledge the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLPDS) for funding. E. A. acknowledges the support of the Ibn Rushd Postdoctoral Fellowship Program, administered by the King Abdullah University of Science and Technology (KAUST). Work by I. L., and I. K. was funded by the Estonian Ministry of Education and Research (TK210) and by the Estonian Research Council grant PRG690. We thank Sofja Tshepelevitsh for intellectual contributions to this work.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Supporting materials for: pK<sub>a</sub> Prediction in Non-Aqueous Solv at <https://zenodo.org/doi/10.5281/zenodo.11153563>.

## ORCID

Jonathan W. Zheng  <https://orcid.org/0000-0002-4863-1325>

Emad Al Ibrahim  <https://orcid.org/0000-0001-8595-5102>

Ivari Kaljurand  <https://orcid.org/0000-0001-6033-7842>

Ivo Leitao  <https://orcid.org/0000-0002-3000-4964>

William H. Green  <https://orcid.org/0000-0003-2603-9694>

## REFERENCES

- [1] D. T. Manallack, R. J. Prankerd, E. Yuriev, T. I. Oprea, D. K. Chalmers, *Chem. Soc. Rev.* **2013**, 42, 485.
- [2] L. Mamy, D. Patureau, E. Barriuso, C. Bedos, F. Bessac, X. Louchart, F. Martin-Laurent, C. Miegé, P. Benoit, *Crit. Rev. Environ. Sci. Technol.* **2015**, 45, 1277.
- [3] R. P. Bell, *The proton in chemistry*, Springer Science & Business Media, Ithaca, New York **2013**.
- [4] X. Pan, H. Wang, C. Li, J. Z. Zhang, C. Ji, *J. Chem. Inf. Model.* **2021**, 61, 3159.
- [5] J. Wu, Y. Wan, Z. Wu, S. Zhang, D. Cao, C.-Y. Hsieh, T. Hou, *Acta Pharmaceut. Sin. B* **2022**, 13, 2572.
- [6] M. Baltruschat, P. Czodrowski, *Chem. Inf. Sci.* **2020**, 9, 1.
- [7] K. Mansouri, N. F. Cariello, A. Korotcov, V. Tkachenko, C. M. Grulke, C. S. Sprankle, D. Allen, W. M. Casey, N. C. Kleinstreuer, A. J. Williams, *J. Cheminform.* **2019**, 11, 1.
- [8] O. Abarbanel, G. Hutchison, *J. Chem. Theory Comput.* **2024**, 20, 6946.
- [9] W. Luo, G. Zhou, Z. Zhu, Y. Yuan, G. Ke, Z. Wei, Z. Gao, H. Zheng, *J. Am. Chem. Soc. Au.* **2024**, 4, 3451.
- [10] T. Sander, J. Freyss, M. von Korff, C. Rufener, *J. Chem. Inf. Model.* **2015**, 55, 460.
- [11] J. Zheng, IUPAC/Dissociation-Constants. **2022** <https://doi.org/10.5281/zenodo.7236453>
- [12] E. P. Serjeant, B. Dempsey, *Ionisation Constants of Organic Acids in Aqueous Solution. IUPAC Chemical Data Series*, Vol. 23, Pergamon Press, Oxford **1979**, p. 160.
- [13] D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solutions*, Vol. 1, Butterworths, London **1965**.
- [14] D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solutions: Supplement*, Vol. 1, Butterworths, London **1972**.
- [15] G. Kortüm, W. Vogel, K. Andrussov, *Pure Appl. Chem.* **1960**, 1, 187.
- [16] K. Izutsu, *Acid-base dissociation constants in dipolar aprotic solvents*, Blackwell Scientific Publications, Oxford **1990**.
- [17] J.-P. Cheng, J.-D. Yang, X.-S. Xue, P. Ji, X. Li, Z. Wang, iBonD. <http://ibond.nankai.edu.cn/>
- [18] R. Roszak, W. Beker, K. Molga, B. A. Grzybowski, *J. Am. Chem. Soc.* **2019**, 141, 17142.
- [19] Q. Yang, Y. Li, J. D. Yang, Y. Liu, L. Zhang, S. Luo, J. P. Cheng, *Angew. Chem. Int. Ed.* **2020**, 59, 19282.
- [20] F. Eckert, I. Leitao, I. Kaljurand, A. Kütt, A. Klamt, M. Diedenhofen, *J. Comput. Chem.* **2009**, 30, 799.
- [21] N. Sülzner, J. Haberhauer, C. Hättig, A. Hellweg, *J. Comput. Chem.* **2022**, 43, 1011.
- [22] M. Busch, E. Ahlberg, E. Ahlberg, K. Laasonen, *ACS Omega* **2022**, 7, 17369.
- [23] E. L. Miguel, P. L. Silva, J. R. Pliego, *J. Phys. Chem. B* **2014**, 118, 5730.
- [24] H.-Z. Yu, Y.-M. Yang, L. Zhang, Z.-M. Dang, G.-H. Hu, *J. Phys. Chem. A* **2014**, 118, 606.
- [25] E. Rossini, R. R. Netz, E. W. Knapp, *J. Chem. Theory Comput.* **2016**, 12, 3360.
- [26] E. Rossini, A. D. Bochevarov, E. W. Knapp, *ACS Omega* **2018**, 3, 1653.
- [27] C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2007**, 111, 408.
- [28] M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, T. R. Tuttle, *J. Phys. Chem. A* **1998**, 102, 7787.

- [29] E. Rossini, E. W. Knapp, *J. Comput. Chem.* **2016**, *37*, 1082.
- [30] E. Rossini, E. W. Knapp, *J. Comput. Chem.* **2016**, *37*, 1082.
- [31] D. Bashford, K. Gerwert, *J. Mol. Biol.* **1992**, *224*, 107.
- [32] D. Bashford, D. A. Case, C. Dalvit, L. Tennant, P. E. Wright, *Biochemistry* **1993**, *32*, 8045.
- [33] A. Kütt, S. Selberg, I. Kaljurand, S. Tshepelevitsh, A. Heering, A. Darnell, K. Kaupmees, M. Piirsalu, I. Leito, *Tetrahedron Lett.* **2018**, *59*, 3738.
- [34] J. W. Zheng, W. H. Green, *J Phys Chem A* **2023**, *127*, 10268.
- [35] Y. Marcus, *Ions in Solution and their Solvation*, John Wiley & Sons, Hoboken, New Jersey **2015**.
- [36] J. Warnau, K. Wichmann, J. Reinisch, *J. Comput.-Aided Mol. Des.* **2021**, *35*, 813.
- [37] C. Loschen, J. Reinisch, A. Klamt, *J. Comput.-Aided Mol. Des.* **2020**, *34*, 385.
- [38] S. Tshepelevitsh, K. Hernits, I. Leito, *J. Comput.-Aided Mol. Des.* **2018**, *32*, 711.
- [39] M. H. Abraham, W. E. Acree, *J. Chromatogr. A* **2016**, *1430*, 2.
- [40] A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2007**, *3*, 2011.
- [41] P. Pracht, F. Bohle, S. Grimme, *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169.
- [42] C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652.
- [43] S. Ehlert, M. Stahn, S. Spicher, S. Grimme, *J. Chem. Theory Comput.* **2021**, *17*, 4250.
- [44] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaeei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, J. M. Yu, *J. Chem. Phys.* **2020**, *152*, 184107.
- [45] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822.
- [46] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
- [47] D. Rappoport, F. Furche, *J. Chem. Phys.* **2010**, *133*, 134105.
- [48] X. Dong, Reaction Data and Molecular Conformers (RDMC). **2024** <https://github.com/xiaoruiDong/RDMC>
- [49] COSMOtherm release 2023. **2023** <https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/solvation-chemistry/biovia-cosmotherm/>
- [50] A. Klamt, M. Diedenhofen, *J. Comput. Chem.* **2018**, *39*, 1648.
- [51] Y. E. Zevatskii, D. O. Ruzanov, D. V. Samoilov, *Russ. J. Gen. Chem.* **2009**, *79*, 1890.
- [52] M. D. Cantu, S. Hillebrand, E. Carrilho, *J. Chromatogr. A* **2005**, *1068*, 99.
- [53] H. Sirén, R. Kuldvee, T. Karla, T. Ekström, M. L. Riekkola, *J. Chromatogr. A* **2005**, *1068*, 89.
- [54] A. D. Headley, S. D. Starnes, L. Y. Wilson, G. R. Famini, *J. Organic Chem.* **1994**, *59*, 8040.
- [55] M. Ludwig, V. Baron, K. Kalfus, O. Pytela, M. Vecera, *Collect. Czechoslov. Chem. Commun.* **1986**, *51*, 2135.
- [56] D. L. Hughes, J. J. Bergan, E. J. J. Grabowski, *J. Org. Chem* **1986**, *51*, 2579.
- [57] A. Ogston, J. Brown, *Trans. Faraday Soc.* **1935**, *31*, 574.
- [58] Y. E. Zevatskii, D. Samoilov, *Russ. J. Org. Chem.* **2008**, *44*, 52.
- [59] E. Kaczmarczyk, L. Chmurzyński, *J. Mol. Struct.* **2000**, *526*, 41.
- [60] A. Kozachenko, E. Matrosov, M. Kabachnik, *Bull. Acad. Sci. USSR, Div. Chem. Sci.* **1976**, *25*, 1411.
- [61] A. Kozachenko, E. Matrosov, M. Kabachnik, *Bull. Acad. Sci. USSR, Div. Chem. Sci.* **1976**, *25*, 2275.
- [62] O. Pytela, J. Kulhánek, M. Ludwig, V. Říha, *Collect. Czech. Chem. Commun.* **1994**, *59*, 627.
- [63] O. Pytela, J. Kulhanek, M. Ludwig, *Collect. Czechoslov. Chem. Commun.* **1994**, *59*, 1637.
- [64] O. Pytela, J. Liska, *Collect. Czechoslov. Chem. Commun.* **1994**, *59*, 2005.
- [65] M. Ludwig, P. Štverka, *Collect. Czech. Chem. Commun.* **1996**, *61*, 1205.
- [66] J. Kulhánek, O. Pytela, *Collect. Czech. Chem. Commun.* **1997**, *62*, 913.
- [67] M. Bos, W. van der Linden, *Anal. Chim. Acta* **1996**, *332*, 201.
- [68] L. C. Kröger, S. Müller, I. Smirnova, K. Leonhard, *J. Phys. Chem. A* **2020**, *124*, 4171.
- [69] A. Malloum, J. J. Fifen, J. Conradie, *J. Mol. Liq.* **2021**, *322*, 114919.
- [70] J. Ho, *Aust. J. Chem.* **2014**, *67*, 1441.
- [71] R. Alexander, A. J. Parker, Solvation of Ions. XII, *J. Am. Chem. Soc.* **1967**, *89*, 5549.
- [72] L. Vlcek, A. A. Chialvo, J. M. Simonson, *J. Phys. Chem. A* **2013**, *117*, 11328.
- [73] L. Vlcek, A. A. Chialvo, *Fluid Phase Equilib.* **2016**, *407*, 58.
- [74] Y. Marcus, *Pure Appl. Chem.* **1983**, *55*, 977.
- [75] L. Pattanaik, A. Menon, V. Settels, K. A. Spiekermann, Z. Tan, F. H. Vermeire, F. Sandfort, P. Eiden, W. H. Green, *J. Phys. Chem. B* **2023**, *127*, 10151.
- [76] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378.
- [77] C. J. Cramer, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 760.
- [78] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2013**, *9*, 609.
- [79] J. R. Pliego, J. M. Riveros, *J. Phys. Chem. A* **2001**, *105*, 7241.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** J. W. Zheng, E. Al Ibrahim, I. Kaljurand, I. Leito, W. H. Green, *J. Comput. Chem.* **2025**, *46*(1), e27517. <https://doi.org/10.1002/jcc.27517>