



Cite this: DOI: 10.1039/c7cp02317k

# Explicit consideration of spatial hydrogen bonding direction for activity coefficient prediction based on implicit solvation calculations†

Wei-Lin Chen and Shiang-Tai Lin \*

The activity coefficient of a chemical in a mixture is important in understanding the thermodynamic properties and non-ideality of the mixture. The COSMO-SAC model based on the result of quantum mechanical implicit solvation calculations has been shown to provide reliable predictions of activity coefficients for mixed fluids. However, it is found that the prediction accuracy is in general inferior for associating fluids. Existing methods for describing the hydrogen-bonding interaction consider the strength of the interaction based only on the polarity of the screening charges, neglecting the fact that the formation of hydrogen bonds requires a specific orientation between the donor and acceptor pairs. In this work, we propose a new approach that takes into account the spatial orientational constraints in hydrogen bonds. Based on the Valence Shell Electron Pair Repulsion (VSEPR) theory, the molecular surfaces associated with the formation of hydrogen bonds are limited to those in the projection of the lone pair electrons of hydrogen bond acceptors, in addition to the polarity of the surface screening charges. Our results show that this new directional hydrogen bond approach, denoted as the COSMO-SAC(DHB) model, requires fewer universal parameters and is significantly more accurate and reliable compared to previous models for a variety of properties, including vapor–liquid equilibria (VLE), infinite dilution activity coefficient (IDAC) and water–octanol partition coefficient ( $K_{ow}$ ).

Received 10th April 2017,  
Accepted 28th June 2017

DOI: 10.1039/c7cp02317k

rsc.li/pccp

## Introduction

Molecular solvation with its focus on the interaction of a solute and its surrounding solvent molecules is important for understanding the spectroscopic, dynamic, and thermodynamic properties of materials.<sup>1,2</sup> By approximating the solvent as a dielectric continuum, implicit solvation methods, such as the generalized Born model,<sup>3</sup> the polarizable continuum model,<sup>4</sup> and conductor-like screening model,<sup>5</sup> have been very successful in providing the solvation free energy and the derived thermodynamic properties of both pure and mixed fluids.<sup>6,7</sup>

In implicit solvation methods, the dielectric response of the solvent due to the presence of a solute is reflected in the apparent charges at the solute–solvent boundary, which can be obtained from the electric displacement vector.<sup>8</sup> In the limiting case where the solvent is a perfect conductor, the screening charges can be more easily obtained based on the fact that the net electric potential is zero at the solute–solvent boundary.<sup>5,9,10</sup> Klamt quantified the distribution of the screening charges into a one-dimensional histogram, called the  $\sigma$ -profile, and developed

a statistical mechanical model, referred to as COSMO-RS,<sup>11</sup> that predicts the activity coefficient of a chemical species in a mixture, based on knowledge of the  $\sigma$ -profile (see Fig. 1 for an illustration of the  $\sigma$ -profile of water). Lin and Sandler developed a COSMO-SAC model<sup>12</sup> that resolved the Gibbs–Duhem inconsistency in the original COSMO-RS model. The COSMO-based activity coefficient models have been shown to provide reliable predictions for a variety of thermodynamic properties and phase equilibria, including vapor–liquid, liquid–liquid, and solid–liquid equilibria for mixtures containing water, organics,<sup>13–18</sup> electrolytes,<sup>19–23</sup> ionic liquids,<sup>24–27</sup> and polymers.<sup>28–30</sup>

Despite their success, the COSMO-based models are known to be less accurate for associating fluids. In the original treatment,<sup>11,31</sup> hydrogen bonding interactions are considered to occur between molecular surfaces associated with opposite and high screening charge densities (*i.e.*, at the two ends of the  $\sigma$ -profile). The strength of a hydrogen bond is quantified by both the screening charge density and an empirical interaction constant whose value was fitted to reproduce experimental phase equilibrium data. Lin *et al.*<sup>32</sup> later proposed limiting the hydrogen bonding interactions to surfaces associated with hydrogen bonding atoms (*i.e.*, O, N, F, and H connected to these atoms) in addition to the polarization of their screening charges. They showed that such treatment allows for more accurate predictions in

Department of Chemical Engineering, National Taiwan University, Taipei 10617, Taiwan. E-mail: stlin@ntu.edu.tw

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp02317k

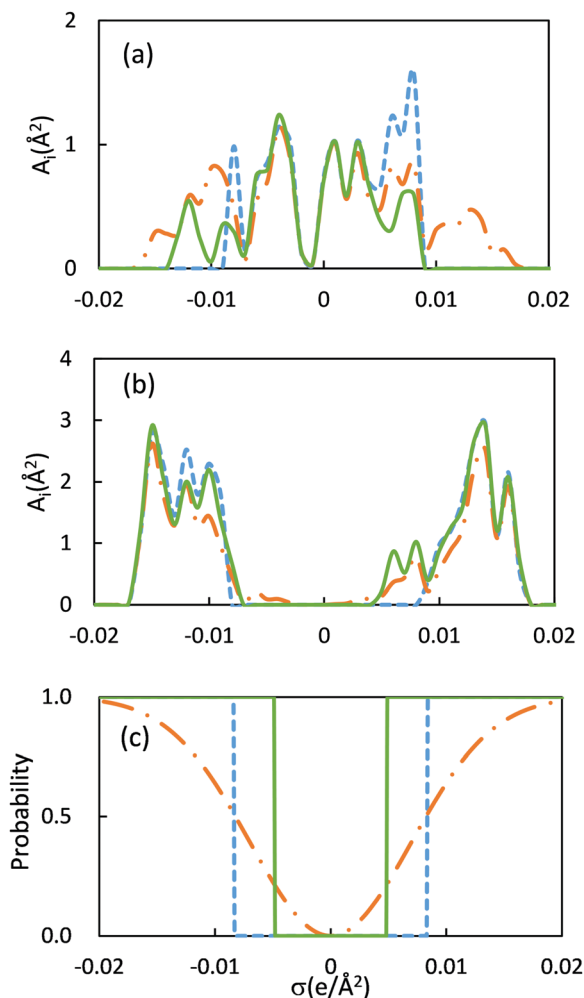


Fig. 1 The  $\sigma$ -profile of water in (a) nhb and (b) hb level and (c) the probability of forming a hydrogen bonding surface. The blue dashed line (—), orange dashed dotted line (---) and green solid line (—) represent the COSMO-SAC 2002, COSMO-SAC 2010 and COSMO-SAC(DHB) models, respectively.

vapor-liquid equilibrium for both pure<sup>32</sup> and mixed<sup>13</sup> fluids over a wide range of temperatures. On the other hand, Klamt *et al.*<sup>33</sup> introduced the temperature dependence in the hydrogen bond parameter of the COSMO-RS model. Hsieh *et al.*,<sup>14</sup> based on experimental observation that the strength of a hydrogen bond depends on the proton acceptor, proposed to further distinguish the type of hydrogen bonds to hydroxyls and non-hydroxyls. Furthermore, they found that there is no clear temperature dependency in the hydrogen bonding interaction parameter.

All the aforementioned efforts to improve the prediction accuracy of COSMO-based methods focus on introducing either temperature-dependent interaction parameters, or donor-acceptor-dependent interaction parameters. While these attempts have improved the description of associating fluids to a certain extent, these successes were achieved at the expense of introducing more parameters. Furthermore, the fact that the hydrogen bonds are directional, *i.e.*, the covalent bond of the proton donor needs to align with the lone pair of the acceptor, has never been explicitly

taken into consideration. Results from both experiment<sup>34,35</sup> and molecular simulations<sup>36,37</sup> have indicated such a directional feature between the hydrogen bonding donor-acceptor pairs. For example, consideration of distance ( $r$ ), the D-H...A angle ( $\theta$ ), the “secondary” H...A-R angle ( $\phi$ ), and the torsion angle ( $\psi$ ) between the donor-acceptor pair are important for describing hydrogen bonding interactions in semiempirical quantum mechanical methods. (Note that D stands for donor, H for hydrogen, A for acceptor, and R for an “acceptor base atom”.)<sup>38,39</sup> For implicit solvation methods, empirical corrections to the hydrogen bond are also found to be important to obtain the correct solvation free energy for amine and amide series.<sup>40</sup> Gallicchio *et al.*<sup>41</sup> further introduced hydrogen sites on the solute that empirically improve the hydration free energy. The COSMO-based methods (a branch of the implicit solvation method) estimate the hydrogen bond interactions between donor-acceptor pairs through the screening charge density.<sup>31</sup> However, the geometric constraint has not been introduced in these models. In this work, we develop a novel approach that allows for the inclusion of spatial constraints on  $\theta$  and  $\phi$  in forming a hydrogen bond in the COSMO-SAC model. We show that this new approach, denoted as the COSMO-SAC model with directional character of the hydrogen bond, or COSMO-SAC(DHB), results in a significantly more accurate prediction of a wide variety of phase behaviors of associating fluids with an even fewer number of parameters than the previous methods.

## Theory

### Solvation energy and activity coefficient

Solvation free energy,  $\Delta G_{i/S}^{*sol}$ , is the work needed for transferring a molecule  $i$  in a fixed position in an ideal gas to a fixed position in a solution  $S$ .<sup>1</sup> The activity coefficient of  $i$  in the solution can be determined from its solvation free energy as<sup>2</sup>

$$\ln \gamma_{i/S} = \frac{\Delta G_{i/S}^{*sol} - \Delta G_{i/i}^{*sol}}{RT} + \ln \frac{C_S}{C_i} \quad (1)$$

where  $C_i$  and  $C_S$  are the molar concentration of pure fluid  $i$  and solution  $S$ , respectively. Typically, the solvation free energy is calculated from a two-step process: first creating a molecular cavity with the size and shape of the solute molecule, and then transferring the solute from the ideal gas to the cavity. The first step is cavity formation and is dominated by the repulsive interactions between the solute and the solution. The second step is equivalent to “turning-on” the charges of the electrons and atomic nuclei and is dominated by the attractive interactions between the solute and the solution. Klamt<sup>11</sup> proposed a novel process that first transfers the solute to a solution in which all the molecules are perfectly screened (and therefore the solution behaves like a perfect conductor), and then all the screening charges are removed to restore the solution to its natural state. Based on this process, the solvation free energy becomes

$$\Delta G_{i/S}^{*sol} = \Delta G_{i/S}^{*cav} + \Delta G_{i/i}^{*is} + \Delta G_{i/S}^{*res} \quad (2)$$

where the superscripts represent cavity formation (cav), ideal solvation (is) and restoration (res), respectively. Lin and Sandler<sup>12</sup> showed that the difference in cavity formation energy of a solute in a solution S and in its pure liquid can be obtained from the so-called combinatorial model, such as the Staverman–Guggenheim model<sup>42</sup> as

$$\frac{\Delta G_{i/S}^{*cav} - \Delta G_{i/i}^{*cav}}{RT} = \ln \gamma_{i/S}^{SG} - \ln \frac{C_S}{C_i} \quad (3)$$

where the Staverman–Guggenheim combinatorial term accounts for the molecular size and shape differences between the species, *i.e.*,

$$\ln \gamma_{i/S}^{SG} = \ln \frac{\phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j \quad (4)$$

with  $\theta_i = (x_i q_i) / \left( \sum_j x_j q_j \right)$ ,  $\phi_i = (x_i r_i) / \left( \sum_j x_j r_j \right)$  and

$l_i = \left( \frac{z}{2} \right) [(r_i - q_i) - (r_i - 1)]$ , where  $x_i$  is the mole fraction of species  $i$ ;  $r_i$  and  $q_i$  are the normalized volume and surface area parameters for compound  $i$ ;  $z$  is the coordination number (usually set to 10); and the summation is over all of the species in the mixture. Substituting eqn (3) into eqn (1), the activity coefficient becomes

$$\ln \gamma_{i/S} = \frac{\Delta G_{i/S}^{*res} - \Delta G_{i/i}^{*res}}{RT} + \ln \gamma_{i/S}^{SG} \quad (5)$$

In COSMO-based models, the restoring energy in eqn (5) is determined based on the screening charges located at the surface of the molecular cavity. By dividing the surface into small segments and assuming that the interactions between segments occur when two segments are in close contact, the restoring energy can be determined from the sum of segment activity coefficient,  $\Gamma_{i/S}$ , as<sup>12</sup>

$$\frac{\Delta G_{i/S}^{*res}}{RT} = \frac{A_i}{a_{\text{eff}}} \sum_{\sigma_m} P_i(\sigma_m) \ln \Gamma_S(\sigma_m) \quad (6)$$

where  $A_i$  is the surface area of component  $i$ ,  $a_{\text{eff}}$  is the effective contact area when two molecules are in contact,  $\Gamma_S(\sigma_m)$  is the activity coefficient of a segment with a screening charge density of  $\sigma_m$  in the solution, and  $P_i(\sigma_m)$ , also known as the  $\sigma$ -profile, is the fraction of the segment with charge density  $\sigma_m$  on component  $i$ . In other words,  $P_i(\sigma_m)$  is the ratio of the surface area of segments whose charge density is  $\sigma_m$  [ $A_i(\sigma_m)$ ] to the total molecular surface area ( $A_i$ ) for molecule  $i$ . For a mixture, the  $\sigma$ -profile is determined from the area-weighted average of the pure components

$$P_S(\sigma) = \frac{\sum_i x_i A_i P_i(\sigma)}{\sum_i x_i A_i} \quad (7)$$

The segment activity coefficient can be determined from the following self-consistent equation derived from lattice fluid theory<sup>12,43</sup>

$$\ln \Gamma_j(\sigma_m) = - \ln \sum_{\sigma_n} P_j(\sigma_n) \Gamma_j(\sigma_n) \exp \left[ \frac{-\Delta W(\sigma_m, \sigma_n)}{kT} \right] \quad (8)$$

with the segment interaction

$$\Delta W(\sigma_m, \sigma_n) = C_{\text{ES}}(\sigma_m + \sigma_n)^2 \quad (9)$$

Note that subscript  $j$  can either be the pure liquid ( $i$ ) or a mixture (S). The coefficient  $C_{\text{ES}}$  is a constant that can be determined from  $a_{\text{eff}}$

$$C_{\text{ES}} = \frac{f_{\text{pol}} \times 0.3 \times a_{\text{eff}}^{3/2}}{2 \times \epsilon_0} \quad (10)$$

where  $f_{\text{pol}}$  is the polarization factor (whose value can be determined from QM calculations<sup>32</sup>) and  $\epsilon_0$  is the permittivity of free space. Therefore, in the basic form of the COSMO-based model, the effective contacting surface area,  $a_{\text{eff}}$ , is the only parameter. Its value was found to vary between 6 and 9 Å<sup>2</sup> depending on the model used for other components of the solvation energy.

### Hydrogen bonding interactions based on strength of donor-acceptor screening charges

The segment interaction of eqn (9) was derived based on the self-energy of a charge segment in a perfect conductor.<sup>5,11</sup> The positive interaction implies that work is needed to remove the screening charges.<sup>12</sup> This is suitable for non-associating fluids, but fails to describe fluids containing hydrogen bonds. Klamt<sup>11</sup> argued that extra stabilization is gained when a hydrogen bond is formed between two segments after their screening charges are removed. Therefore, he proposed to revise eqn (9) to the following

$$\Delta W(\sigma_m, \sigma_n) = C_{\text{ES}}(\sigma_m + \sigma_n)^2 + C_{\text{hb}} \max[0, \sigma_{\text{acc}} - \sigma_{\text{hb}}] \min[0, \sigma_{\text{don}} + \sigma_{\text{hb}}] \quad (11)$$

where  $C_{\text{hb}}(\sigma_m^t, \sigma_n^s)$  is the hydrogen bonding parameter;  $\sigma_{\text{acc}}$  and  $\sigma_{\text{don}}$  are the larger and smaller values of  $\sigma_m$  and  $\sigma_n$ . Based on eqn (11) the hydrogen bonding interaction (always negative) takes effect only when the charge density on both segments exceeds a cutoff value,  $\sigma_{\text{hb}} = 0.0084 \text{ e Å}^{-1}$ . The COSMO-SAC 2002 model<sup>44</sup> adopted this idea for hydrogen bonding interactions in associating fluids.

### Hydrogen bonding interactions based on type of donor-acceptor pairs

Based on experimental observation that the strength of hydrogen bonds varies with the donor-acceptor pairs,<sup>45</sup> which was not properly described using a simple cutoff (eqn (11)), Hsieh *et al.*<sup>14</sup> proposed to classify the hydrogen bonding surfaces according to the elements of the proton acceptor, *i.e.*,

$$P_i(\sigma) = P_i^{\text{nhb}}(\sigma) + P_i^{\text{OH}}(\sigma) + P_i^{\text{OT}}(\sigma) \quad (12)$$

where superscript OH represents the surface on the hydroxyl group and the OT represents the other hydrogen bonding surface (such as the N of amines, the O of a carbonyl group, the O of ethers, *etc.*), *i.e.*,

$$P_i^{\text{OH}}(\sigma) = [A_i^{\text{OH}}(\sigma)/A_i] \times P^{\text{HB}}(\sigma) \quad (13a)$$

$$P_i^{\text{OT}}(\sigma) = [A_i^{\text{OT}}(\sigma)/A_i] \times P^{\text{HB}}(\sigma) \quad (13b)$$

where  $A_i^{\text{OH}}(\sigma)$  and  $A_i^{\text{OT}}(\sigma)$  are the surface area with charge density  $\sigma$  of hydroxyl and non-hydroxyl groups of species  $i$ , respectively. The equation  $P^{\text{HB}}(\sigma) = 1 - \exp(\sigma/2\sigma_0^2)$  is an

empirical function that reduces the probability of forming a hydrogen bond for hb segments with low charge densities (see Fig. 1(c)) ( $\sigma_0$  is  $0.007 \text{ e } \text{\AA}^{-2}$ ).<sup>13</sup> The  $\sigma$ -profile for nonhydrogen bonding (nhb) segments becomes

$$P_i^{\text{nhb}}(\sigma) = A_i^{\text{nhb}}(\sigma)/A_i + [A_i^{\text{OH}}(\sigma)/A_i + A_i^{\text{OT}}(\sigma)/A_i] \times (1 - P^{\text{HB}}(\sigma)) \quad (13c)$$

Furthermore, the segment interactions are modified to the following

$$\Delta W(\sigma_m^t, \sigma_n^s) = C_{\text{ES}}(\sigma_m^t + \sigma_n^s)^2 - C_{\text{hb}}(\sigma_m^t, \sigma_n^s)(\sigma_m^t - \sigma_n^s)^2 \quad (14)$$

where  $C_{\text{hb}}$  is a constant whose value depends on the type of donor–acceptor segments

$$C_{\text{hb}}(\sigma_m^t, \sigma_n^s) = \begin{cases} C_{\text{OH-OH}} & \text{if } s = t = \text{OH and } \sigma_m^t \times \sigma_n^s < 0 \\ C_{\text{OT-OT}} & \text{if } s = t = \text{OT and } \sigma_m^t \times \sigma_n^s < 0 \\ C_{\text{OH-OT}} & \text{if } s = \text{OH}, t = \text{OT and } \sigma_m^t \times \sigma_n^s < 0 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

and the electrostatic term is also modified to be temperature dependent

$$C_{\text{ES}} = A_{\text{ES}} + \frac{B_{\text{ES}}}{T^2} \quad (16)$$

This approach is referred to as COSMO-SAC 2010 hereafter.

### Hydrogen bonding interactions with explicit consideration of hydrogen bond direction

As can be seen, previous efforts to improve the description of hydrogen bonding interactions have been focused on tuning the strength of interaction and recognizing the type of donor–acceptor pairs. Although these approaches are simple and effective, they do not take into account the directional character common to hydrogen bonds. For example, Fig. 2 illustrates the formation of a hydrogen bond between two methanol molecules. MD simulations<sup>36</sup> reveal that the distribution of  $\text{HO} \cdots \text{H}$  angle  $\phi$  centers around  $107^\circ$ , suggesting that a hydrogen bond is formed along the direction of the electron lone pair on proton acceptor. Furthermore, the distribution of  $\text{OH} \cdots \text{O}$ , solid angle  $\theta$ , suggests a preference for a linear hydrogen bond. Here we propose a new approach to include the directional character in the modelling of hydrogen bonding interactions in the COSMO-based models. The principle is to limit the hydrogen bonding surfaces to those that come into contact when a hydrogen bond is formed. For a proton acceptor X (X = N, O, or F), we define the hydrogen bond center (hbc) to be the point on the surface in the lone pair direction (see Fig. 3). For a donor, the hbc is simply the point on the surface along the covalent bond XH.

$$\vec{\text{hbc}}_k^{j\text{-th}} = \vec{P}_k + R_k \times \hat{D}_k^{j\text{-th}} \quad \text{for } j \leq n_k \quad (17)$$

where  $\vec{\text{hbc}}_k^{j\text{-th}}$  is the position of  $j$ -th hbc on the surface of atom  $k$ ,  $\vec{P}_k$  and  $R_k$  are the position and radius of atom  $k$ ,  $\hat{D}_k^{j\text{-th}}$  is the unit vector from atom  $i$  toward hbc, and  $n_k$  is the numbers of hbc on

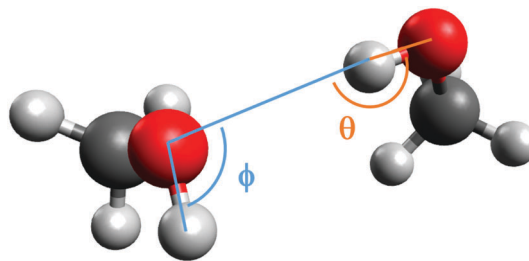


Fig. 2 The geometry of a methanol dimer showing the directional feature of the hydrogen bond: angle  $\phi$  centering around  $107^\circ$  and angle  $\theta$  around  $180^\circ$ .

atom  $k$ . The value of  $n_k$  and vector  $\hat{D}_k^{j\text{-th}}$  can be determined from the molecular geometry and the bond angles according to the VSEPR theory.<sup>46</sup> Table 1 summarizes the typical structures of

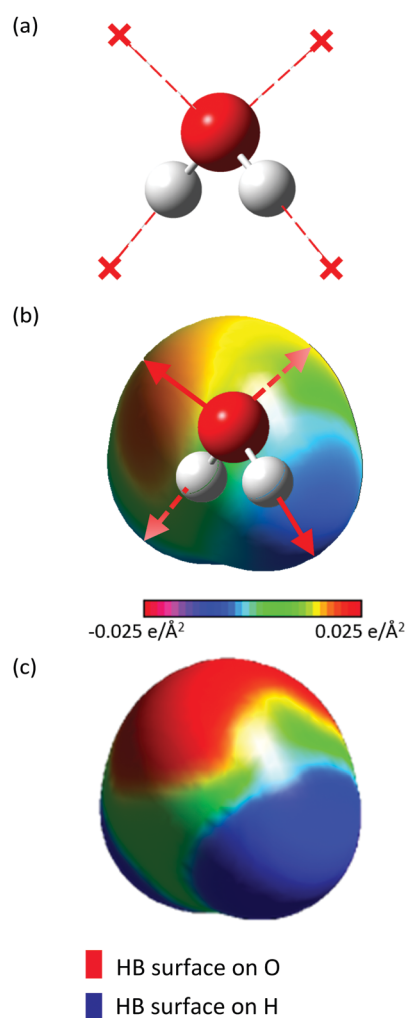
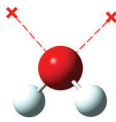
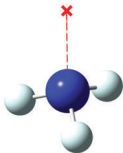
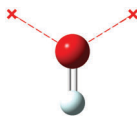
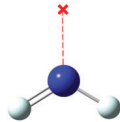
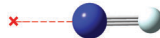



Fig. 3 (a) The hydrogen bond centers (hbc) of a water molecule are indicated by red crosses. (b) The spatial distribution of COSMO screening charge distribution on the water surface. The color indicates the intensity of the screening charges. The red arrow shows the vector from oxygen (or hydrogen) toward hbc. (c) The hydrogen bonding segments identified based on hbc and a cutoff radius. The red indicates the hydrogen bond acceptor surface and blue represents the donor surface.

**Table 1** Illustration of hydrogen bond centers on the surface of hydrogen bond acceptors and donor

Molecular type	AX <sub>2</sub> E <sub>2</sub>	AX <sub>3</sub> E <sub>1</sub>	AX <sub>1</sub> E <sub>2</sub>	AX <sub>2</sub> E <sub>2</sub>	AX <sub>1</sub> E <sub>1</sub>	AX <sub>1</sub>
Geometric structure						
Family example	Water alcohol ether	Amine	Ketone	Pyridine	Nitrile	Hydrogen bond donor
Bond angle	109.5	109.5	120	120	180	180
Number of hb centers ( <i>n<sub>k</sub></i> )	2	1	2	1	1	1

hydrogen bond acceptors and the corresponding hbc (*cf.* ESI,† for details on determination of  $\hat{D}_k^{\text{th}}$ ).

The hbc is used to determine the hydrogen bonding segments of a molecule. A segment is assigned to be an hb segment if its distance to an hbc is less than some hydrogen bonding cut-off distance ( $R_{\text{cut}}^{\text{HB}}$ ), *i.e.*,

$$A_i^{\text{hb}} = \sum_{\sigma_m} s(\vec{r}_{\sigma_m}) \quad \text{if} \quad \left| \vec{r}_{\sigma_m} - \text{hbc}_k^{\text{th}} \right| < R_{\text{cut}}^{\text{HB}} \quad (18)$$

where  $\vec{r}_{\sigma_m}$  is the position of segment  $\sigma_m$  and  $s(\vec{r}_{\sigma_m})$  is its area.

Fig. 3 illustrates the determination of hb segments for water, in which the oxygen atom belongs to type AX<sub>2</sub>E<sub>2</sub> (see Table 1). The two lone pairs and two OH bonds form a tetrahedral structure, as shown in Fig. 3(a). The two hbc on oxygen and hbc on each hydrogen are identified according to eqn (17). The hb segments are then determined according to eqn (18). This results in hb segments shown in the red and blue regions on oxygen and hydrogen atoms, respectively, as shown in Fig. 3(c).

It is worth mentioning that the special cases of sp<sup>2</sup> oxygen and sp<sup>3</sup> nitrogen on the empirical hydrogen bond potential function from the semi-empirical method are consistent with our proposed method. For sp<sup>2</sup> oxygen, not only 120 but also 180 are considered as possible hydrogen bonding sites. For sp<sup>3</sup> nitrogen, the possibility of planarization is considered.<sup>47</sup> The details and additional examples (methanol, *n*-butyl amine,

*n*-methyl formamide and acetonitrile) are provided in Fig. S6 of the ESI.†

With this identification of hb segments, the segment interactions are modified to be

$$\Delta W(\sigma_m^t, \sigma_n^s) = \left( A_{\text{ES}} + \frac{B_{\text{ES}}}{T^2} \right) (\sigma_m^t + \sigma_n^s)^2 + C_{\text{hb}} \max[0, \sigma_{\text{acc}} - \sigma_{\text{hb}}] \min[0, \sigma_{\text{don}} + \sigma_{\text{hb}}] \quad (19)$$

As a result, the hb interactions are determined based on their spatial location and the charge density, as shown in Fig. 1(c) and (b), respectively. We refer to this approach as the COSMO-SAC(DHB) model.

## Computational details

In this work, the performances of the three COSMO-SAC models are compared to determine the effectiveness of different treatments in hydrogen bonding interactions. The model parameters for each model are summarized and compared in Table 2. For COSMO-SAC 2002 and 2010, the parameters are taken directly from the literature. For the new COSMO-SAC(DHB) model, the three parameters ( $R_{\text{cut}}^{\text{HB}}$ ,  $\sigma_{\text{hb}}$  and  $C_{\text{hb}}$ ) for describing hydrogen bonding interactions are optimized based on vapor–liquid equilibrium (VLE) data for 586 binary mixtures containing 35 associating fluids (which include water, alcohol, amine and amines). All other parameters are set to be the same as those for COSMO-SAC 2010. The objective function used is the following:

obj of VLE

$$= \frac{1}{M} \left[ \sum_{i=1}^M (y_i^{\text{calc}} - y_i^{\text{expt}})^2 \right]^{1/2} + \frac{1}{M} \left[ \sum_{i=1}^M \left( \frac{p^{\text{calc}} - p^{\text{expt}}}{p^{\text{expt}}} \right)^2 \right]^{1/2} \quad (20)$$

where  $M$  is the number of VLE data points; the superscripts calc and expt indicate the calculated and experimental values, respectively.

The liquid–liquid equilibrium (LLE), infinite dilution activity coefficient (IDAC) and the octanol–water partition coefficient ( $K_{\text{ow}}$ ) all are used to examine the predictive power of the COSMO-SAC(DHB) model. The VLE, LLE and IDAC data were retrieved from the DECHEMA Chemistry Data Series,<sup>48–50</sup> while

**Table 2** Parameters in COSMO-SAC models considered in this work

Parameter	COSMO-SAC2002	COSMO-SAC2010	COSMO-SAC(DHB)
$f_{\text{pol}}$ (—)	0.6916	—	—
$A_{\text{ES}}$ (kcal mol <sup>−1</sup> Å <sup>−4</sup> e <sup>−2</sup> )	—	6525.69	6525.69
$B_{\text{ES}}$ (kcal mol <sup>−1</sup> Å <sup>−4</sup> e <sup>−2</sup> K <sup>−2</sup> )	—	$1.4859 \times 10^8$	$1.4859 \times 10^8$
$C_{\text{hb}}$ (kcal mol <sup>−1</sup> Å <sup>−4</sup> e <sup>−2</sup> )	85 580	—	31881.95
$C_{\text{OH-OH}}$ (kcal mol <sup>−1</sup> Å <sup>−4</sup> e <sup>−2</sup> )	—	4013	—
$C_{\text{OT-OT}}$ (kcal mol <sup>−1</sup> Å <sup>−4</sup> e <sup>−2</sup> )	—	932	—
$C_{\text{OH-OT}}$ (kcal mol <sup>−1</sup> Å <sup>−4</sup> e <sup>−2</sup> )	—	3016	—
$\sigma_{\text{hb}}$ (e Å <sup>−2</sup> )	0.0084	—	0.0049
$\sigma_0$ (e Å <sup>−2</sup> )	—	0.007	—
$R_{\text{cut}}^{\text{HB}}$ (Å)	—	—	1.5934
$a_{\text{eff}}$ (Å <sup>2</sup> )	7.5	7.25	7.25
$f_{\text{decay}}$ (—)	3.57	3.57	3.57
$r$ (Å <sup>2</sup> )	66.69	66.69	66.69
$q$ (Å <sup>3</sup> )	79.53	79.53	79.53



the  $K_{ow}$  data were taken from the CRC handbook.<sup>51</sup> The  $K_{ow}$  is determined from the IDAC in a water-rich phase (which is nearly pure water) and an octanol-rich phase (containing approximately 0.725 mole fraction of octanol), respectively,<sup>15</sup>

$$\log K_{ow,i} = \log \left( \frac{8.37\gamma_i^{W,\infty}}{55.5\gamma_i^{O,\infty}} \right) \quad (21)$$

Our choice of comparing the prediction results with experimental phase equilibrium data (IDAC and VLE) and not the solvation free energy deserves some explanation. The solvation free energy consists of a repulsive contribution ( $\Delta G^{cav}$ ) in addition to the attractive charging contribution (sum of  $\Delta G^{is}$ ,  $\Delta G^{res}$ , and  $\Delta G^{dsp}$ ). Since the dispersion and cavity terms are often estimated by semi-empirical models that contain fitting parameters, direct comparison would be inconclusive because of the additional empirical terms. To reduce the uncertainty from the empirical models, we have made a comparison with various experiments based on the activity coefficient. For associating fluids, the activity coefficient is dominated by the restoring free energy (eqn (5)) and thus the error from the Staverman–Guggenheim combinatorial term is minimized.

Furthermore, the computational cost of the COSMO-SAC(DHB) model is similar to that of the COSMO-SAC 2002 and much less than that of the COSMO-SAC 2010. For example, the times for calculating activity coefficients for 100 binary mixtures are 3.84, 7.70 and 4.16 seconds using COSMO-SAC 2002, COSMO-SAC 2010 and COSMO-SAC(DHB) models, respectively.

## Results and discussions

In this work, a comprehensive data set containing VLE, LLE, IDAC and  $K_{ow}$  was used to validate the COSMO-SAC(DHB) model. For VLE, the collected data include 586 binary mixtures (7583 data points, temperature range from 263.15 K to 548.15 K); for LLE, 190 binary mixtures; IDAC, 431 data points (283 K to 373 K); and  $K_{ow}$ , 291 species (at 298.15 K). The prediction accuracy of each phase behavior is analysed and compared for COSMO-SAC 2002, COSMO-SAC 2010 and COSMO-SAC(DHB) in Table 3.

### Vapor–liquid equilibrium and liquid–liquid equilibrium

The VLE is useful for estimating the general performance of different models over the whole concentration range. The accuracy of VLE is evaluated using the average absolute relative deviation in

vapor pressure (AARD-P) and average absolute deviation in vapor phase mole fraction (AAD-y):

$$\text{AARD-P} = \sum_{i=1}^N \frac{1}{N} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} \left| \frac{P^{\text{calc}} - P^{\text{expt}}}{P^{\text{expt}}} \right| \right] \times 100\% \quad (22)$$

$$\text{AAD-y} = \sum_{i=1}^N \frac{1}{N} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} (y_i^{\text{calc}} - y_i^{\text{expt}}) \right] \times 100\% \quad (23)$$

where  $N$  is the number of vapor–liquid equilibrium mixtures;  $M_i$  is the number of data points in the  $i$ th mixture. The overall AARD-P from COSMO-SAC(DHB) is improved by 27% and 11% compared to COSMO-SAC 2002 and COSMO-SAC 2010, respectively. Furthermore, the AAD-y from COSMO-SAC(DHB) is reduced by 26% and 9% compared to COSMO-SAC 2002 and COSMO-SAC 2010, respectively. Fig. 4 illustrates the VLE diagrams of four example binary mixtures. Fig. 4(a)–(c) demonstrate three different types of hydrogen bond mixtures: (a) self-association, (b) cross-association with one hb-donor and (c) cross-association with two hb-donors. In all three cases, COSMO-SAC 2002 underestimates the non-ideality of the mixtures. COSMO-SAC 2010 shows an improvement in cases (a) and (c) but overestimates the non-ideality in case (b). With consideration of the directional character of hydrogen bonds, the COSMO-SAC(DHB) model is found to describe all these three types of mixtures well. (The detailed deviations based on these three types of mixtures are given in Tables S1 and S2, ESI.†) COSMO-SAC(DHB) has the lowest mean deviation and standard deviation for all three types of mixture. It again validates the proposed method in identifying hb segments. It is noteworthy that COSMO-SAC(DHB), similar to COSMO-SAC 2002, overestimates the non-ideality in the mixture of carbon tetrachloride and aniline, as shown in Fig. 4(d). The slightly better prediction from COSMO-SAC 2010 in this case supports the differentiation of hydrogen bond interactions based on the type of donor atoms.

The LLE is useful to assess the contrast in describing two coexisting fluids in the richness of each of the two components. The performance in LLE was evaluated using the root-mean square error in liquid composition:

$$\text{RMS} - x_i = \frac{1}{N} \sum_{i=1}^{N_{\text{sys}}} \left[ \frac{1}{N_i} \sum_{j=1}^{N_i} (x_j^{\text{calc}} - x_j^{\text{expt}})^2 \right]^{1/2} \quad (24)$$

where  $N$  is the number of liquid–liquid equilibrium mixtures. For LLE, the COSMO-SAC 2010 model is the most accurate, possibly due to the recognition of different hb interaction strengths for different donor–acceptor pairs, and also the fact

**Table 3** Prediction accuracy of COSMO-SAC models in vapor–liquid equilibrium (VLE), liquid–liquid equilibrium (LLE), infinite dilution activity coefficients (IDAC) and octanol–water partition coefficient ( $K_{ow}$ )

Method	Model parameters	VLE		LLE		IDAC	$K_{ow}$
		AARD-P (%)	AAD-y (%)	RMS- $x_i$	$N_{\text{sys}}^a$	RMS- $\ln \gamma_i$	RMS- $\log K_{ow}$
COSMO-SAC 2002	7	9.04	3.60	0.1061	147	1.816	0.614
COSMO-SAC 2010	10	7.37	2.91	0.0843	168	1.424	0.464
COSMO-SAC(DHB)	9	6.58	2.65	0.1021	167	1.336	0.468
Number of systems		586		190		431	291

<sup>a</sup>  $N_{\text{sys}}$  represents the number of systems (out of a total of 190) that a converged LLE calculation is attained.

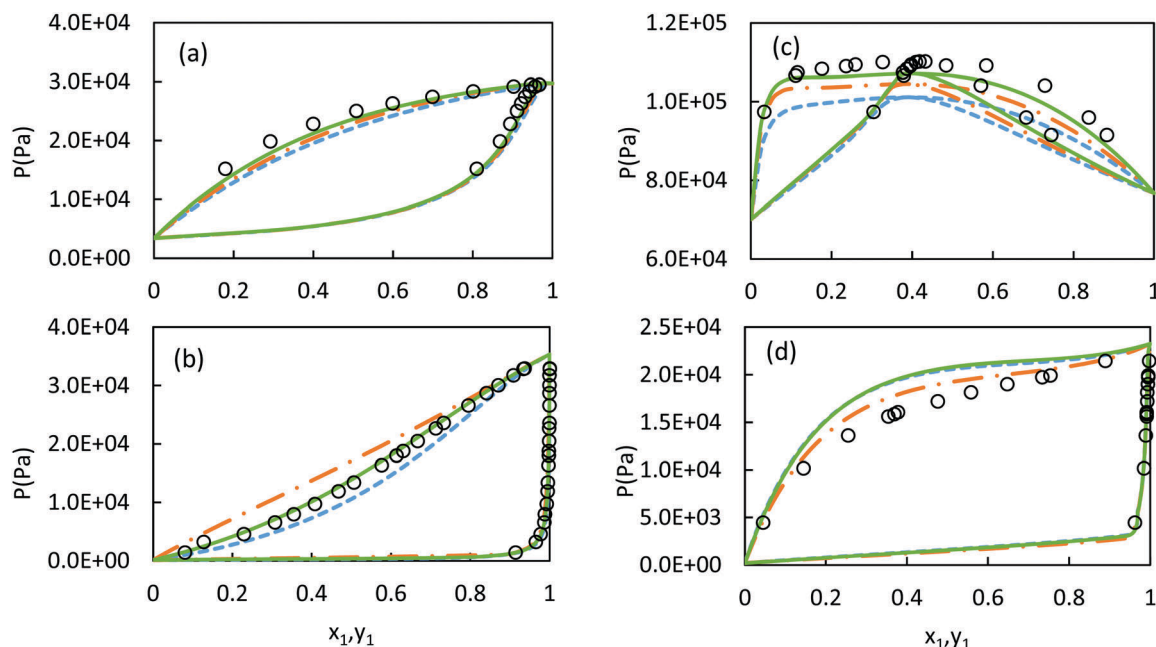


Fig. 4 Comparison of vapor–liquid equilibria prediction from COSMO-SAC 2002(—), COSMO-SAC 2010(---) and COSMO-SAC(DHB) (—) to experimental data (O): (a) benzene (1) + 1-butanol (2), (b) methanol (1) + *N*-methyl-2-pyrrolidone (2), (c) 1-propanol (1) + water (2) and (d) carbon tetrachloride (1) + aniline (2).

that its parameters were determined based solely on LLE data.<sup>14</sup> The performance of the new COSMO-SAC(DHB) is slightly better than the COSMO-SAC 2002 model, indicating the effectiveness of inclusion of spatial constraints for hydrogen bonds. It is worth mentioning that the RMS alone may not always provide a correct assessment of the quality of LLE prediction. For example, only COSMO-SAC(DHB) can correctly identify the phase separation in the mixture of ethylene diamine and benzene (see Fig. 5(a)). As a second example, both COSMO-SAC 2010 and COSMO-SAC(DHB) well predict the phase separation in the mixture of methanol and cyclohexane (Fig. 5(b)) but the slight difference in predicting the upper critical solution temperature results in a higher RMS in COSMO-SAC(DHB) compared to COSMO-SAC 2010 (0.113 compared to 0.081). In general, the capability in describing LLE from COSMO-SAC(DHB) is similar to that from COSMO-SAC 2010.

#### Infinite dilution activity coefficient and octanol–water partition coefficient

The IDAC and  $K_{ow}$  are useful to assess the solvation of a solute by another solvent. The performance in IDAC and  $K_{ow}$  was evaluated using the root-mean square error as follows:

$$\text{RMS} - \ln \gamma_i = \left[ \frac{1}{N} \sum_{i=1}^N (\ln \gamma_i^{\text{calc}} - \ln \gamma_i^{\text{expt}})^2 \right]^{1/2} \quad (25)$$

$$\text{RMS} - \log K_{ow} = \left[ \frac{1}{N} \sum_{i=1}^N (\log K_{ow}^{\text{calc}} - \log K_{ow}^{\text{expt}})^2 \right]^{1/2} \quad (26)$$

where  $N$  is the number of pieces of experimental data. For IDAC, COSMO-SAC(DHB) has the lowest deviation among all the methods. The scatterplot of experimental data *versus* predicted

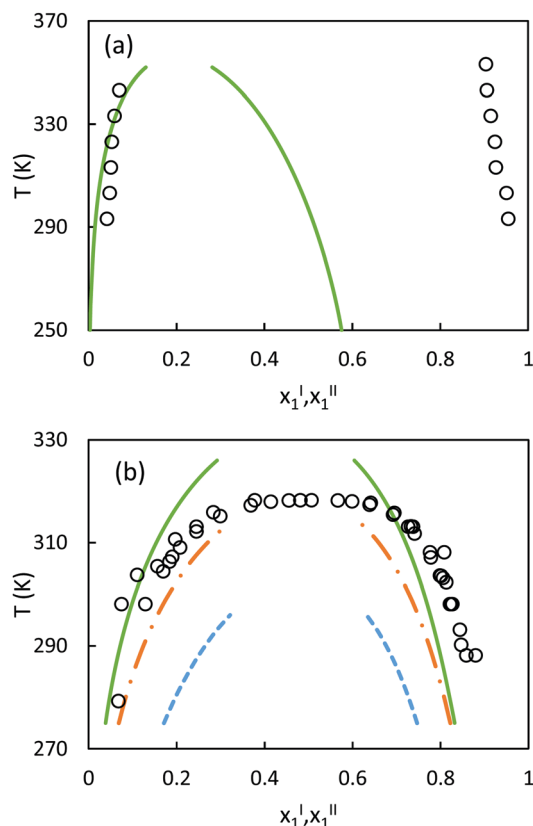


Fig. 5 Comparison of the liquid–liquid equilibria prediction from COSMO-SAC 2002(—), COSMO-SAC 2010(---) and COSMO-SAC(DHB) (—) to experimental data (O): (a) ethylenediamine (1) + benzene (2), (b) methanol (1) + cyclohexane (2).

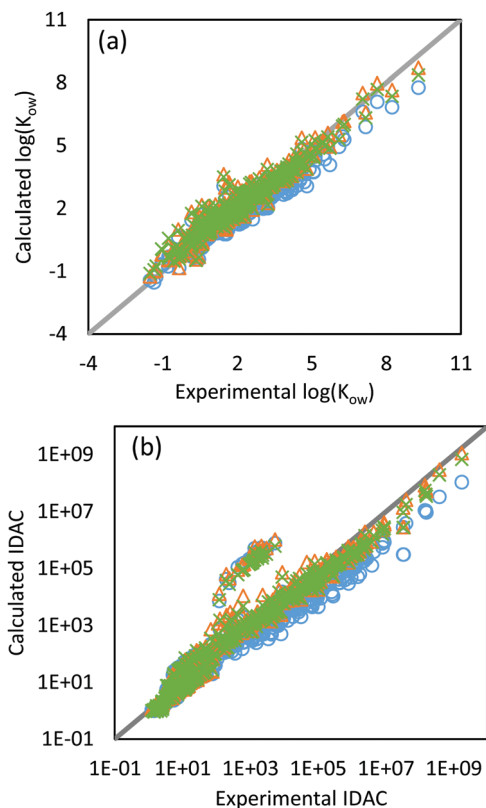


Fig. 6 Comparison of the prediction of (a) infinite dilution activity coefficient (IDAC) and (b) octanol–water partition coefficient ( $K_{ow}$ ) from COSMO-SAC 2002 (○), COSMO-SAC 2010 (△) and COSMO-SAC(DHB) (×). The horizontal axis shows the experimental value and the vertical axis shows the calculated value from different models.

value, Fig. 6(a), shows that all COSMO-based methods are in good agreement with the experimental data but have some obvious outliers. The outliers come from overestimating the IDAC of water in alkane, a weakness of such methods, as reported previously.<sup>52</sup> For  $K_{ow}$ , the COSMO-SAC(DHB) shows a similar accuracy when compared to COSMO-SAC 2010 (0.468 versus 0.464 in RMS). On the other hand, COSMO-SAC 2002 shows less accuracy in predicting the  $K_{ow}$ . The scatterplot of experimental data versus predicted value is provided in Fig. 6(b).

### Statistical analysis on VLE, IDAC and $K_{ow}$

In addition to the comparison of overall accuracy, the standard deviation and maximum deviation over 586, 431 and 291

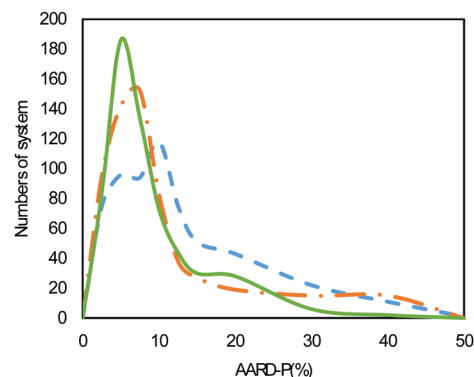


Fig. 7 Comparison of the distribution of AARD-P(%) in the vapor–liquid equilibrium prediction from COSMO-SAC 2002(–), COSMO-SAC 2010(–) and COSMO-SAC(DHB) (—). The vertical axis shows the number of systems.

mixtures in VLE, IDAC and  $K_{ow}$  are evaluated. Table 4 lists the statistics on deviation in VLE using AARD-P and AAD-y; the deviation in IDAC using  $\Delta \ln \gamma_i = \ln \gamma_i^{\text{calc}} - \ln \gamma_i^{\text{expt}}$ ; and  $K_{ow}$  using  $\Delta \log K_{ow} = \log K_{ow}^{\text{calc}} - \log K_{ow}^{\text{expt}}$ . For VLE, the COSMO-SAC(DHB) model indeed has a significant improvement over the COSMO-SAC 2010 model where the standard deviation is reduced by 25% and 15% in AARD-P and AAD-y, respectively. The narrower distribution in AARD-P from COSMO-SAC(DHB) shown in Fig. 7 suggests that this approach is more reliable than the other two methods. In addition to VLE, for IDAC and  $K_{ow}$  the performance of COSMO-SAC(DHB) is the best among the four models. The improvement in standard deviation and maximum deviation again supports the inclusion of directional constraints on hb segments in COSMO-based methods. Compared to the COSMO-SAC 2010 model, the proposed COSMO-SAC(DHB) uses fewer universal parameters (9 vs. 10, see Table 2) but still shows improved accuracy and reliability.

### Comparison with molecular dynamics simulations

To show that the electrostatic component of the charging energy ( $\Delta G^{\text{el}}$ ) from the proposed method is physically significant, we compare in Table 5 the hydration energy of 4 ions ( $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$ ) determined from the COSMO-SAC model and the molecular dynamic (MD) simulations of Reif and Hünenberger.<sup>53</sup>

The  $\Delta G^{\text{el}}$  from MD depends on the solvation energy used for protons. Sets L, M, and H in the final 3 columns of Table 5 correspond to the use of  $-1100$ ,  $-1075$  and  $-1050 \text{ kJ mol}^{-1}$  for the hydration energy of  $\text{H}^+$ . It can be seen that  $\Delta G^{\text{el}}$  from the

Table 4 Comparison of standard deviation and maximum deviation in vapor–liquid equilibrium (VLE), infinite dilution activity coefficient (IDAC) and octanol–water partition coefficient ( $K_{ow}$ ) among different COSMO-SAC models

Method	Standard deviation ( $\sigma$ )			Maximum deviation		
	VLE	IDAC	$K_{ow}$	VLE	IDAC	$K_{ow}$
	AARD-P/AAD-y	$\Delta \ln \gamma_i$	$\Delta \log K_{ow}$	AARD-P/AAD-y	$\Delta \ln \gamma_i$	$\Delta \log K_{ow}$
COSMO-SAC 2002	6.49/3.15	1.229	0.366	45.15/20.85	5.931	1.749
COSMO-SAC 2010	6.30/2.68	1.142	0.318	38.91/19.24	6.084	2.173
COSMO-SAC(DHB)	4.72/2.29	1.028	0.312	35.52/19.52	5.591	2.043
Number of systems	586	431	291	586	431	291



**Table 5** Comparison of the electrostatic component of the solvation free energy (in kJ mol<sup>−1</sup>) for ions in water from the Born, COSMO-SAC models, and molecular dynamics simulations

	Radius (Å)	Born <sup>a</sup>	COSMO-SAC 2002 <sup>b</sup>	COSMO-SAC 2010 <sup>b</sup>	COSMO-SAC(DHB) <sup>b</sup>	MD set L <sup>c</sup>	MD set M <sup>c</sup>	MD set H <sup>c</sup>
Li <sup>+</sup>	1.39	−494.0	−714.7	−568.7	−573.3	−542.3	−518.4	−494.0
Na <sup>+</sup>	1.69	−406.3	−601.4	−484.8	−498.5	−440.9	−418.4	−395.9
K <sup>+</sup>	2.00	−343.3	−475.2	−409.1	−419.2	−370.5	−349.2	−329.8
Rb <sup>+</sup>	2.15	−319.4	−415.5	−374.3	−382.2	−347.0	−328.6	−309.4

<sup>a</sup> The Born model ( $\Delta G^{\text{el}} = -(1/2)(1 - 1/\epsilon)q^2/R$ ) with  $\epsilon = 80$  (water). <sup>b</sup>  $\Delta G^{\text{el}} = \Delta G^{\text{is}} + \Delta G^{\text{res}}$ . <sup>c</sup>  $\Delta G^{\text{el}} = \Delta G^{\text{hyd}} - \Delta G^{\text{cav}} - \Delta G^{\text{std}}$  from Reif and Hünenberger.

COSMO-SAC 2002 model is too high (by about 150 kJ mol<sup>−1</sup>); however, the 2010 and DHB versions provide values close to the set L results (within 50 kJ mol<sup>−1</sup>). The Born model results, without explicit consideration of the hydrogen bonding interactions, show a much lower (in terms of magnitude) charging energy that is consistent with the set H results. These results indicate that the COSMO-SAC(DHB) model proposed in this work does provide a physically meaningful hydration energy that is consistent with MD simulation results.

## Conclusions

A novel approach is developed to include the directional character of hydrogen bonds for predicting the phase equilibrium of mixtures using the COSMO-SAC model. In this new method, denoted as COSMO-SAC(DHB), the hb segments are recognized within the confined area of the donor-acceptor pair. The spatial locations for the allowed hb segments are identified based on the VSEPR theory and the molecular geometry. The new approach requires only 3 universal parameters to describe the size of the contacting surface area of the donor-acceptor pairs, the strength of the hb interaction, and the cutoff charge density for the formation of an hb bond. Based on an examination of a large data base of VLE, LLE, IDAC and  $K_{\text{ow}}$ , we conclude that the inclusion of the directional constraints is very effective in improving both the prediction accuracy and the reliability of COSMO-based methods.

## Acknowledgements

This research was partially supported by the Ministry of Science and Technology of Taiwan (MOST 104-2221-E-002-186-MY3) and National Taiwan University (NTU-CDP-105R7876). The computation resources from the National Center for High-Performance Computing of Taiwan and the Computing and Information Networking Center of the National Taiwan University are acknowledged.

## Notes and references

- 1 A. Ben-Naim, *Solvation Thermodynamics*, Plenum Press, New York, 1st edn, 1987.
- 2 S. T. Lin, C. M. Hsieh and M. T. Lee, *J. Chin. Inst. Chem. Eng.*, 2007, **38**, 467–476.
- 3 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
- 4 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
- 5 A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
- 6 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.
- 7 J. Tomasi and M. Persico, *Chem. Rev.*, 1994, **94**, 2027–2094.
- 8 C. J. F. Bötcher, *Theory of electric polarisation*, Elsevier, Amsterdam, 1st edn, 1952.
- 9 J. Tomasi, B. Mennucci and E. Cancès, *J. Mol. Struct.*, 1999, **464**, 211–226.
- 10 S. T. Lin and C. M. Hsieh, *J. Chem. Phys.*, 2006, 125.
- 11 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
- 12 S. T. Lin and S. I. Sandler, *Ind. Eng. Chem. Res.*, 2002, **41**, 899–913.
- 13 S. Wang, S. I. Sandler and C.-C. Chen, *Ind. Eng. Chem. Res.*, 2007, **46**, 7275–7288.
- 14 C.-M. Hsieh, S. I. Sandler and S.-T. Lin, *Fluid Phase Equilib.*, 2010, **297**, 90–97.
- 15 C.-M. Hsieh, S. Wang, S.-T. Lin and S. I. Sandler, *J. Chem. Eng. Data*, 2011, **56**, 936–945.
- 16 C. C. Shu and S. T. Lin, *Ind. Eng. Chem. Res.*, 2011, **50**, 142–147.
- 17 W.-L. Chen, C.-M. Hsieh, L. Yang, C.-C. Hsu and S.-T. Lin, *Ind. Eng. Chem. Res.*, 2016, **55**, 9312–9322.
- 18 H. Grensemann and J. Gmehling, *Ind. Eng. Chem. Res.*, 2005, **44**, 1610–1624.
- 19 S. Wang, Y. Song and C.-C. Chen, *Ind. Eng. Chem. Res.*, 2011, **50**, 176–187.
- 20 T. Ingram, T. Gerlach, T. Mehling and I. Smirnova, *Fluid Phase Equilib.*, 2012, **314**, 29–37.
- 21 T. Husch and M. Korth, *Phys. Chem. Chem. Phys.*, 2015, **17**, 22596–22603.
- 22 T. Husch, N. D. Yilmazer, A. Balducci and M. Korth, *Phys. Chem. Chem. Phys.*, 2015, **17**, 3394–3401.
- 23 C. Schütter, T. Husch, V. Viswanathan, S. Passerini, A. Balducci and M. Korth, *J. Power Sources*, 2016, **326**, 541–548.
- 24 B. S. Lee and S. T. Lin, *Fluid Phase Equilib.*, 2013, **356**, 309–320.
- 25 B. S. Lee and S. T. Lin, *Chem. Eng. Sci.*, 2015, **121**, 157–168.
- 26 B. S. Lee and S. T. Lin, *Ind. Eng. Chem. Res.*, 2015, **54**, 9005–9012.
- 27 M. Diedenhofen and A. Klamt, *Fluid Phase Equilib.*, 2010, **294**, 31–38.
- 28 Y. C. Kuo, C. C. Hsu and S. T. Lin, *Ind. Eng. Chem. Res.*, 2013, **52**, 13505–13515.

- 29 L. Yang, X. Xu, C. Peng, H. Liu and Y. Hu, *AIChE J.*, 2010, **56**, 2687–2698.
- 30 C. Loschen and A. Klamt, *Ind. Eng. Chem. Res.*, 2014, **53**, 11478–11487.
- 31 A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz, *J. Phys. Chem. A*, 1998, **102**, 5074–5085.
- 32 S. T. Lin, J. Chang, S. Wang, W. A. Goddard and S. I. Sandler, *J. Phys. Chem. A*, 2004, **108**, 7429–7439.
- 33 A. Klamt and F. Eckert, *Fluid Phase Equilib.*, 2000, **172**, 43–72.
- 34 J. Kroon, J. A. Kanters, J. g. Vanduijnneveltdt, F. B. Vanduijnneveltdt and J. A. Vliegthart, *J. Mol. Struct.*, 1975, **24**, 109–129.
- 35 R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1983, **105**, 5761–5766.
- 36 W. L. Jorgensen, *J. Am. Chem. Soc.*, 1980, **102**, 543–549.
- 37 A. Luzar and D. Chandler, *J. Chem. Phys.*, 1993, **98**, 8160–8173.
- 38 M. Korth, *ChemPhysChem*, 2011, **12**, 3131–3142.
- 39 N. D. Yilmazer and M. Korth, *Comput. Struct. Biotechnol. J.*, 2015, **13**, 169–175.
- 40 B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnald, D. Sitkoff and B. Honig, *J. Phys. Chem.*, 1996, **100**, 11775–11788.
- 41 E. Gallicchio, K. Paris and R. M. Levy, *J. Chem. Theory Comput.*, 2009, **5**, 2544–2564.
- 42 A. J. Staverman, *Recl. Trav. Chim. Pays-Bas*, 1950, **69**, 163–174.
- 43 S. T. Lin, M. K. Hsieh, C. M. Hsieh and C. C. Hsu, *J. Chem. Thermodyn.*, 2009, **41**, 1145–1153.
- 44 S. T. Lin and S. I. Sandler, *Ind. Eng. Chem. Res.*, 1999, **38**, 4081–4091.
- 45 G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, Freeman, San Francisco, 1960.
- 46 R. H. Petrucci, *General chemistry: principles and modern applications*, Learning Solutions, New York, 2011.
- 47 M. Korth, M. Pitoňák, J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2010, **6**, 344–352.
- 48 J. r. Gmehling, U. Onken and W. Arlt, *Vapor–liquid equilibrium data collection*, Dechema, Frankfurt, 1977.
- 49 J. M. Sørensen and W. Arlt, *Liquid–liquid equilibrium data collection*, Dechema, Frankfurt, 1979.
- 50 D. Tiegs, *Activity coefficients at infinite dilution*, Dechema, Frankfurt, 1986.
- 51 W. M. Haynes, D. R. Lide and T. J. Bruno, *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data*, 2015.
- 52 C. S. Chen and S. T. Lin, *Ind. Eng. Chem. Res.*, 2016, **55**, 9284–9294.
- 53 M. M. Reif and P. H. Hunenberger, *J. Chem. Phys.*, 2011, **134**, 25.