



Cite this: *Phys. Chem. Chem. Phys.*,  
2025, **27**, 24734

# Prediction of Henry's law constants for CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid *via* different Monte Carlo approaches

Prasil Kapadiya  and Jhumpa Adhikari \*

Henry's law constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid are predicted at 313.15 K and 101.325 kPa. The Widom test particle insertion method (which is based on excess chemical potential computation) in the NPT ensemble *via* two different Monte Carlo approaches, *viz.*, CBMC and CFMC, is implemented. GEMC-NVT simulations are performed to calculate Henry's law constants from the slope of the line fitted to the plot of partial pressure of solute (CO<sub>2</sub> or CH<sub>4</sub>) in the vapour phase *versus* the mole fraction of solute in liquid levulinic acid. Henry's law constants estimated *via* all three Monte Carlo methods and for both the solutes show good agreement within statistical uncertainties. The negative value of the Gibbs free energy of transfer ( $\Delta G_{\text{trans}}$ ) of CO<sub>2</sub> from the vapour to the solvent (levulinic acid) phase suggests a preference of CO<sub>2</sub> for liquid levulinic acid over the vapour phase. The solvation number computed from solute–solvent centre-of-mass radial distribution functions (RDFs) shows a higher value for CH<sub>4</sub> than CO<sub>2</sub>, contrary to the trend in Henry's law constant. Therefore, the local environment around the solute molecule is investigated by computing the solute–solvent site–site intermolecular RDFs. The number integral calculated (up to the first minimum) from site–site intermolecular RDFs of the C(CO<sub>2</sub>)–methyl group of levulinic acid (CH<sub>3</sub>(1)) is more than that calculated from the CH<sub>4</sub>–CH<sub>3</sub>(1) RDF, confirming the higher solubility of CO<sub>2</sub> in levulinic acid than CH<sub>4</sub>.

Received 10th July 2025,  
Accepted 14th October 2025

DOI: 10.1039/d5cp02630j

rsc.li/pccp

## 1. Introduction

The Henry's law constant ( $H_{2,1}(T, p)$ ) is generally employed to provide a measure of the solubility of a gas (solute species 2) in a liquid solvent (species 1) and is defined as the ratio of the equilibrium partial pressure of the solute in the gas phase ( $p_2^{\text{gas}}$ ) to its mole fraction in the liquid solvent ( $x_2^{\text{liq}}$ ) at infinite dilution as described in eqn (1).<sup>1,2</sup>

$$H_{2,1}(T, p) = \lim_{x_2^{\text{liq}} \rightarrow 0} \left( \frac{p_2^{\text{gas}}}{x_2^{\text{liq}}} \right) \quad (1)$$

From eqn (1), it follows (as stated by Zhang and Siepmann)<sup>2</sup> that the value of the Henry's law constant depends on the type of solute, the solvent, the temperature and the pressure.<sup>2</sup> Knowledge of phase behaviour of mixtures, in general, and that of the solvation of gases in liquids in particular is crucial to the design, optimization and modelling of various industrial processes<sup>3</sup> such as the use of existing commercial solvents (namely, amines, Selexol and Rectisol) for precombustion CO<sub>2</sub> capture,<sup>4–6</sup> amine scrubbing process for CO<sub>2</sub> capture from flue gas generated in power plants producing electricity under post

combustion conditions,<sup>5</sup> and removal of acid gas impurities from synthesis gas or natural gas.<sup>5</sup> Determination of Henry's law constants is therefore important from an environmental perspective also. Challenges arise in measuring the Henry's law constant at elevated temperatures for technologically important mixtures, if the solvent is flammable and/or if the solute is an oxidizing agent, for example, when measuring the solubility of O<sub>2</sub> in ethanol.<sup>2</sup> Hence, *in silico* computation of Henry's law constants, using molecular simulation techniques, where the accuracy of the value obtained depends on the molecular models adopted and the sampling of microstates, is therefore necessary. Various research groups world-wide have developed algorithms<sup>2,3,7–11</sup> for predicting the Henry's law constant and have also tested the sensitivity of the results to the different potential models or force fields and combining rules employed to describe the solute–solute, solute–solvent and solvent–solvent interactions, with the infinite dilution properties having been found to be extremely sensitive to the approximations in the solute–solvent interaction models.

Determination of the Henry's law constant at a given temperature and pressure *via* molecular simulation techniques often includes the computation of the excess chemical potential of the solute at infinite dilution. The Widom test-particle insertion method (TPI)<sup>7</sup> and staged free-energy perturbation (FEP)

Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai – 400076, India. E-mail: adhikari@iitb.ac.in

technique<sup>9,10</sup> are usually employed for calculation of the excess chemical potential. The Widom TPI technique computes the excess chemical potential as the difference in the free energy of the  $(N + 1)$  particle system and the system containing  $N$  particles by inserting a test or ghost particle during the perturbation move.<sup>7</sup> The Widom TPI method is a simple and elegant technique; however, since the method involves insertion of the test particle, longer simulation runs may be required (for example, at low temperatures) or other more efficient methods may be needed when a solvent contains a large solute molecule and/or the solvent fluid is extremely dense.<sup>11</sup> In the staged FEP method, the interaction between the solute particle and solvent particles is gradually varied using a coupling parameter  $\lambda$  which varies from 0 to 1.<sup>9,10</sup> Corresponding to the sequence of increasing  $\lambda$  values, a set of independent isobaric isothermal (NPT) sub-ensembles are considered.<sup>9,10</sup> This method expresses the Boltzmann factor of the excess chemical potential at infinite dilution as the product of the contributions (which are functions of the difference in configurational energy between two consecutive sub-ensembles) from the sequence of sub-ensembles.<sup>9,10</sup> Here, the value of free energy difference depends upon the direction of the perturbation either from the reference system to the target system or *vice versa*.<sup>3,8</sup> The application of FEP methods requires careful consideration of the phase-space relationship between the reference and target systems, as highlighted by Kofke and Cummings.<sup>3,8</sup> Alternatively, Gibbs ensemble Monte Carlo (GEMC) simulations can be used for direct calculation of Henry's law constants, analogous to experiments, without the use of excess chemical potential.<sup>2</sup> This two-box method allows for the simultaneous determination of the vapour pressure and the mole fractions of solutes in both the vapour and liquid phases.<sup>2</sup> Using these values, the Henry's law constant can be calculated as the slope of the  $p_{\text{solute}}^{\text{gas}}$  vs.  $x_{\text{solute}}^{\text{liq}}$  plot<sup>2</sup> (as described in eqn (1)). The expanded ensemble approach<sup>12–14</sup> is another commonly employed approach for free energy calculations where two systems of interest are connected *via* a collection of sub-ensembles. These sub-ensembles are transformed gradually from one system of interest to another *via* Monte Carlo trials attempting to change the identity of the sub-ensemble and from the probability distribution obtained, the relative free energy of each sub-ensemble can be determined.<sup>12–14</sup> The expanded ensemble technique was coupled with the self-adaptive transition matrix Monte Carlo (TMMC) method by Cichowski *et al.*<sup>3</sup> for the efficient computation of Henry's law constants of a wide-range of solutes in numerous solvents. This combination provides a powerful tool for calculating the balancing factors and, hence, the relative free energies for each sub-ensemble, with high efficiency.<sup>3</sup>

The above discussed Monte Carlo methods can be applied in conjunction with configurational-bias Monte Carlo (CBMC)<sup>15</sup> and continuous fractional component Monte Carlo (CFCMC)<sup>16</sup> algorithms. The CBMC algorithm developed by Siepmann and Frenkel,<sup>15</sup> based on the Rosenbluth and Rosenbluth scheme,<sup>17</sup> includes a Monte Carlo move that explores the large-scale conformational changes of a molecule in a single trial move. Here, we first discuss some examples reported in the literature for the computation of Henry's law constants using CBMC.

Lisal *et al.*<sup>11</sup> estimated the Henry's constant of CO<sub>2</sub> in water over a wide range of temperatures, starting from 0 °C to the critical temperature of water *via* Monte Carlo simulations in the NPT ensemble. The effect of the number of interaction sites on the value of the Henry's constant was investigated by employing two three-site potentials for CO<sub>2</sub> (EPM2 and Errington and Panagiotopoulos) and various water models with three (SPC, SPC/E, MSPC/E, and Errington and Panagiotopoulos), four (TIP4P), and five (TIP5P) interaction sites.<sup>11</sup> The Widom TPI and staged FEP methods were used to compute the Henry's constant.<sup>11</sup> The Henry's constants predicted using Errington and Panagiotopoulos models were in good agreement with the experimental values.<sup>11</sup> While extremely long simulation runs at lower temperatures were required for accurate prediction of the Henry's law constant of CO<sub>2</sub> in water, the FEP method allows further analysis of Henry's law constants for both low and high temperature calculations.<sup>11</sup> CBMC simulations in the canonical (NVT) version of the Gibbs ensemble have been performed by Zhang and Siepmann<sup>2</sup> to compute the Henry's law constants and the Gibbs free energies of transfer for O<sub>2</sub>, N<sub>2</sub>, CH<sub>4</sub>, and CO<sub>2</sub> in ethanol at 323 and 373 K. The solvent, ethanol, was modelled using the TraPPE-UA (transferable potentials for phase equilibria-united atom) force field, CH<sub>4</sub> was modelled by TraPPE-UA and TraPPE-EH (transferable potentials for phase equilibria-explicit hydrogen) force fields, rigid models were used for CO<sub>2</sub> (three LJ and three partial charge sites) and N<sub>2</sub> (two LJ and three partial charge sites), and the model developed by Coon *et al.* was used for O<sub>2</sub>.<sup>2</sup> Most of their simulations involved systems containing 1000 solvent molecules and 1, 2, 4 or 8 solute molecules.<sup>2</sup> An excellent agreement between the GEMC data with predictions made utilizing the transition matrix Monte Carlo method was observed.<sup>2</sup> A mean unsigned error of 15% was observed for the data predicted using the TraPPE-UA force field with experiments.<sup>2</sup> Zhang and Siepmann<sup>2</sup> determined that employing the united-atom representation (TraPPE-UA) for methane results in more accurate predictions of the Henry's law constant in united-atom (TraPPE-UA) ethanol compared to those obtained using the explicit-hydrogen (TraPPE-EH) methane. On the other hand, predictions for O<sub>2</sub> showed higher deviations from the experimental data as compared to N<sub>2</sub> and CO<sub>2</sub>.<sup>2</sup> Cichowski *et al.*<sup>3</sup> estimated Henry's law constants of CH<sub>4</sub>, O<sub>2</sub>, N<sub>2</sub>, and CO<sub>2</sub> in ethanol at 323 and 373 K using an expanded NPT ensemble coupled with the self-adaptive TMMC method. Infinite-dilution excess chemical potential values for the solute were calculated.<sup>3</sup> Molecular interactions were described by the TraPPE force field for all molecules except O<sub>2</sub>, where the model developed by Coon *et al.* was employed.<sup>3</sup> For N<sub>2</sub> in addition, the force field developed by Galassi and Tildesley was also used to investigate the sensitivity of the results to the molecular models employed.<sup>3</sup> The Henry's law constant of O<sub>2</sub> in ethanol estimated using this approach showed the highest deviation from the experimental data.<sup>3</sup> Shah and Maginn<sup>18</sup> used the Widom TPI method and expanded ensemble Monte Carlo approach in the NPT ensemble to compute the Henry's law constants of H<sub>2</sub>O, CO<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, O<sub>2</sub>, and N<sub>2</sub> in the ionic liquid 1-*n*-butyl-3-methylimidazolium hexafluorophosphate ([bmim][PF<sub>6</sub>]).

The force field for the ionic liquid was based on a UA model, while H<sub>2</sub>O, CO<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, O<sub>2</sub>, and N<sub>2</sub> were modelled using SPC/E, TraPPE-small, TraPPE-UA, Miyano and TraPPE-small force fields, respectively.<sup>18</sup> The predicted values of solubilities of the solutes in the ionic liquid were higher than the experimental values.<sup>18</sup> Additionally, the temperature dependency of the Henry's law constants for CO<sub>2</sub>, O<sub>2</sub> and water was investigated using the Widom TPI method.<sup>18</sup> While the simulations using the Widom TPI method accurately predicted the trend where solubility of H<sub>2</sub>O and CO<sub>2</sub> decreases with increasing temperature, the method failed to show the increase in solubility of O<sub>2</sub> with temperature due to entropic effects governing the dissolution of O<sub>2</sub>.<sup>18</sup>

The CFCMC algorithm, introduced by Shi and Maginn,<sup>16</sup> overcomes the difficulties encountered with the insertion and deletion of molecules in simulations, especially for dense systems. The method achieves efficient insertion and deletion of molecules by employing a continuously changing coupling parameter and an adaptive bias potential.<sup>16</sup> This scaling factor ( $\lambda$ ) modulates the strength of intermolecular interactions between the fractional molecule and its neighbouring molecules within the system.<sup>16</sup>  $\lambda$  ranges from 0 to 1,  $\lambda = 0$  represents no interactions between the fractional molecule and surrounding whole molecules, and  $\lambda = 1$  implies that interactions are fully developed.<sup>16</sup> Examples of studies reported in the literature where CFCMC has been employed to compute the Henry's law constant are described here. The solubilities of CO<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, CO, H<sub>2</sub>, N<sub>2</sub>, N<sub>2</sub>O and H<sub>2</sub>S in various commercial solvents, like Rectisol, Selexol, Purisol and Fluor, were studied by Chen *et al.*<sup>5</sup> using CFCMC simulations in the osmotic ensemble. CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, CO<sub>2</sub> and N<sub>2</sub> were modelled by the standard TraPPE force fields,<sup>5</sup> and molecular models for CO, H<sub>2</sub>, N<sub>2</sub>O and H<sub>2</sub>S developed by Martin-Calvo *et al.*, Cracknell *et al.*, Lachet *et al.* and Guitierrez-Sevillano *et al.*, respectively, were used.<sup>5</sup> Overall, a good agreement between the experimental values and the estimated Henry's law constants for these gases was observed by Chen and co-workers.<sup>5</sup> Ramdin *et al.*<sup>6</sup> computed the solubilities and Henry's law constants of CO<sub>2</sub>, CH<sub>4</sub>, CO, H<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>S in the ionic liquid 1-butyl-3-methylimidazolium bis(trifluoromethylsulfonyl) imide ([bmim][TF<sub>2</sub>N]) at 333.15 K using CFCMC simulations in the osmotic ensemble. For the ionic liquid, the force field developed by Liu and Maginn was used.<sup>6</sup> As in the study by Chen *et al.*, the standard TraPPE force fields were adopted for CH<sub>4</sub>, CO<sub>2</sub> and N<sub>2</sub>, and CO and H<sub>2</sub> were represented using the Martin-Calvo *et al.* and Cracknell *et al.* models, respectively.<sup>6</sup> Ramadin and co-workers examined three different force fields for H<sub>2</sub>S, *viz.*, a three-site model developed by Kamath *et al.*, a four-site model by Kristof and Lizi and a five-site model by Guitierrez-Sevillano *et al.*<sup>6</sup> The predicted gas solubilities and Henry's law constants aligned well with the experimental data.<sup>6</sup> Solubilities of CO<sub>2</sub>, CH<sub>4</sub>, CO, H<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>S in green deep eutectic solvents, namely, choline chloride urea and choline chloride ethylene glycol, were estimated by Salehi *et al.*<sup>19</sup> using CFCMC simulations in the NPT ensemble *via* the open source Brick-CFCMC software package. Two force fields, namely, GAFF (general Amber force field) and OPLS

(optimized potentials for liquid simulations), were used to estimate the solubility at 328 K.<sup>19</sup> It was observed that the OPLS force field yielded higher average Henry's constants in the case of choline chloride with ethylene glycol compared to GAFF.<sup>19</sup> Furthermore, the solubility order for these solutes in choline chloride urea and choline chloride ethylene glycol deep eutectic solvents was observed to be in reasonable agreement with experimental values.<sup>19</sup>

Chemocatalytic conversion of cellulose (through glucose), as well as a one-pass catalytic process in the presence of a multi-functional catalyst, produces a wide range of molecules, from large molecules such as furfural, 5-hydroxy-2-methylfurfural and levulinic acid to small molecules with one carbon atom such as CO<sub>2</sub> and CH<sub>4</sub>.<sup>20</sup> To the best of our knowledge, there is no reported value for the Henry's constants of CO<sub>2</sub> or CH<sub>4</sub> in levulinic acid in the literature. Thus, we have performed molecular simulations to compute the Henry's law constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K and 101.325 kPa. We have compared the results obtained from CBMC and CFCMC simulations in the NPT ensemble *via* the Widom TPI method. Moreover, this provides us with an opportunity to compare the computational resources required by the two algorithms (namely, CBMC and CFCMC). Additionally, canonical version of Gibbs ensemble Monte Carlo (GEMC-NVT) simulations in conjunction with CBMC moves with 1000 levulinic acid molecules and 1, 2, 4 or 8 CO<sub>2</sub> or CH<sub>4</sub> molecules are performed for the estimation of Henry's law constant. All CBMC simulations are performed using the MCCCSTowhee software package,<sup>21</sup> whereas CFCMC simulations are performed using the Brick-CFCMC software.<sup>22</sup> Furthermore, we have performed Metropolis MC simulations in the canonical ensemble with 1 molecule of each CO<sub>2</sub> and CH<sub>4</sub> in 300 molecules of levulinic acid (such that solute-solute interactions are absent) at 313.15 K. Solvation of CH<sub>4</sub> and CO<sub>2</sub> molecules in levulinic acid, which is a keto-acid containing two functional groups (keto-group and carboxylic acid group), is studied *via* computing the solute-solvent centre-of-mass (COM) radial distribution functions (RDFs), site-site intermolecular RDFs and their corresponding number integrals (NIs).

This article is organized as follows: a detailed description of the force fields and simulation methodologies adopted here is presented in Sections 2 and 3, respectively. The results are analysed and discussed in Section 4. The conclusions are summarised in the last section.

## 2. Force fields

The selection of the appropriate force field is crucial for any molecular simulation method, as the accuracy of the simulation results depends on the force field and combining rules employed to describe the interactions present in the system of interest. The force fields adopted in this study have been described here.

### 2.1. TraPPE-UA force field

The interactions present in levulinic acid<sup>23-25</sup> (refer to Scheme 1) and CH<sub>4</sub><sup>26</sup> (refer to Scheme 2a) are described using the TraPPE-UA



Scheme 1 Structure of LA for the TraPPE-UA force field.

Scheme 2 Structure of (a) CH<sub>4</sub> for the TraPPE-UA force field and (b) CO<sub>2</sub> for the TraPPE-small force field.

force field. The nonbonded interactions are modelled by 12-6 Lennard-Jones (LJ) and Coulombic potentials as shown in eqn (2),

$$u(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2)$$

where  $r_{ij}$ ,  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the site–site separation distance, the LJ well depth and the LJ diameter, respectively, and  $q_i$  and  $q_j$  are the partial charges on UAs  $i$  and  $j$ , respectively. The parameters for LJ cross-interactions between unlike UAs are calculated by employing Lorentz–Berthelot combining rules as shown below:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad (3)$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (4)$$

All the bond lengths are kept rigid and modelled as

$$u_{\text{bond}} = \begin{cases} \infty, & \text{if } \left| \frac{r - r_0}{r_0} \right| > 0.01 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where  $r_0$  is the equilibrium bond length.

The bond bending is modelled by the harmonic potential specified as

$$u_{\text{bend}} = \frac{K_\theta}{2} (\theta - \theta_{\text{eq}})^2 \quad (6)$$

where,  $K_\theta$ ,  $\theta$  and  $\theta_{\text{eq}}$  are the bending force constant, the bond angle and the equilibrium bond angle, respectively.

The torsional interactions are modelled using eqn (7), depending on the four united atoms involved.

$$u_{\text{torsion}}(\phi) = c_0 + c_1[1 + \cos \phi] + c_2[1 + \cos 2\phi] + c_3[1 + \cos 3\phi] \quad (7)$$

Here,  $\phi$  is the dihedral angle and  $c_i$  represents the  $i^{\text{th}}$  coefficient.

The TraPPE-UA force field parameters for levulinic acid and CH<sub>4</sub> are reported in Tables S1 and S2 of the SI, respectively.

## 2.2. TraPPE-small force field

The TraPPE-small force field is used to model CO<sub>2</sub><sup>27,28</sup> (refer to Scheme 2b). The nonbonded interactions between atoms are represented using the 12-6 LJ potential given in eqn (2). The cross-interactions between unlike atoms are modelled using the Lorentz–Berthelot combining rules and defined in eqn (3) and (4). The bonded interactions (bond stretching and bond bending) in the CO<sub>2</sub> molecule are held rigid and are described by eqn (5) and (8), respectively.

$$u_{\text{angle}} = \begin{cases} \infty, & \text{if } |\theta - \theta_0| > 0.0001 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

Here,  $\theta_0$  is the equilibrium bond angle.

The nonbonded and bonded parameters for CO<sub>2</sub> modelled using the TraPPE-small force field are reported in Table S3 of the SI.

## 3. Methodology

At temperature  $T$  and pressure  $p$ , the Henry's constant for solute  $i$  in a solvent  $j$  is given as<sup>29</sup>

$$H_{j,i}(T, p) = \lim_{x_i \rightarrow 0} \left( \frac{f_i^L(T, p, x_i)}{x_i} \right) \quad (9)$$

where  $x_i$  is the solute mole fraction and  $\hat{f}_i^L$  is the fugacity of the solute,  $i$ , in the mixture. Based on standard thermodynamic relationships, the expression can be written as<sup>1,11</sup>

$$H_{j,i}(T, p) = \frac{\rho_j}{\beta} \exp(\beta \mu_i^{\text{ex},\infty}(T, V)) \quad (10)$$

where  $\beta = 1/k_B T$ ,  $k_B$  is the Boltzmann constant,  $\rho_j$  is the density of the solvent,  $\mu_i^{\text{ex},\infty}$  is the excess chemical potential of the solute at infinite dilution and  $V$  is the system volume. Hence, from eqn (10), in order to estimate the Henry's law constant at  $T$  and  $p$  using molecular simulations, it is necessary to determine the solvent density and the excess chemical potential of the solute. It should be noted that approximations in the force fields related to solute–solvent interactions will significantly impact the infinite dilution properties,<sup>11</sup> as discussed in Section 2.

### 3.1. Widom test-particle insertion (TPI) method

The Widom test particle insertion method is a notably simple and elegant approach for measuring the chemical potential of a species in a pure fluid or a mixture. Here, the infinite dilution

chemical potential is calculated using the Widom test-particle insertion (TPI) method.<sup>7,11</sup> In this method, a solute particle (also called the “test” particle) is randomly inserted into a pure solvent. This insertion of the “test” particle can be either *via* CBMC or CFCMC moves. The exponential term containing the excess chemical potential of the solute at infinite dilution in eqn (10) can be computed either using eqn (11) when the CBMC algorithm is employed or eqn (12) for the CFCMC approach.

In the NPT ensemble simulation in conjunction with the CBMC algorithm, the excess chemical potential of the solute at infinite dilution is computed using eqn (11).<sup>7,11,30</sup>

$$\exp[\beta\mu_i^{\text{ex},\infty}(T, \langle V \rangle_{\text{NPT}})] = \frac{\langle V \rangle_{\text{NPT}}}{\langle V \exp[-\beta U_i^{\text{test}}] \rangle_{\text{NPT}}} \quad (11)$$

where  $\langle V \rangle_{\text{NPT}}$  denotes the ensemble average of the volume obtained from MC simulations in the NPT ensemble and  $U_i^{\text{test}}$  is the configurational energy of the inserted test particle.

In the CFCMC algorithm, the excess chemical potential term can be calculated using the following expression:<sup>22,31</sup>

$$\mu^{\text{ex}} = -k_{\text{B}}T \ln \left[ \frac{p(\lambda = 1)}{p(\lambda = 0)} \right] \quad (12)$$

where  $p(\lambda)$  is the Boltzmann probability distribution of  $\lambda$ .

### 3.2. Gibbs ensemble Monte Carlo method in the canonical ensemble

Similar to experiments, the GEMC-NVT<sup>32</sup> simulation technique allows for the direct calculation of Henry's law constants.<sup>2</sup> A two-phase system facilitates the direct measurement of both the vapour pressure and the solute mole fractions in both the vapour and liquid phases. The values of  $x_{\text{solute}}^{\text{liq}}$  and  $p_{\text{solute}}^{\text{gas}}$  for computing the Henry's law constant using this method can be determined simply by computing the ensemble averages of these properties. Using these values, the Henry's law constant can be calculated from the slope of the line fitted to the plot of  $p_{\text{solute}}^{\text{gas}}$  vs.  $x_{\text{solute}}^{\text{liq}}$  (as described in eqn (1)).<sup>2</sup>

### 3.3. Simulation details

**3.3.1. NPT ensemble using the CBMC algorithm.** CBMC simulations in the NPT ensemble have been performed to estimate the Henry's law constant of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid using the Widom test-particle insertion method at 313.15 K and 101.325 kPa *via* the MCCC'S Towhee version 8.2.3 software package.<sup>21</sup> The initial configuration consists of 300 levulinic acid molecules placed in a cubic box. Periodic boundary conditions have been implemented in all three dimensions. The LJ potential has been used with a cutoff distance of 14 Å and analytical tail corrections have been implemented. Ewald summation is utilized to calculate Coulombic interactions, with an electrostatic cutoff of 14 Å and parameters  $\kappa \times L = 5.6$  and  $K_{\text{max}} = 5$ , where  $\kappa$  is the Ewald sum convergence parameter,  $L$  is the length of the cubic simulation box and  $K_{\text{max}}$  is the maximum number of reciprocal space vectors included in the sum.<sup>33</sup> The trial moves are volume change moves, aggregation volume biased moves, CBMC regrowth moves, centre of mass translation and rotation moves

in the frequency ratio 10 : 5 : 25 : 30 : 30. The system is allowed to equilibrate over at least 1 500 000 MC cycles with averages computed over 100 000 cycles of the production run.

**3.3.2. NPT ensemble using the CFCMC algorithm.** CFCMC simulations in the NPT ensemble have been used to compute the Henry's law constant of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid *via* the Widom test-particle insertion method. These simulations are performed at 313.15 K and 101.325 kPa with 300 levulinic acid molecules. Here, the MC simulations have been carried out using an open source software package, Brick-CFCMC.<sup>22</sup> One fractional molecule of each CO<sub>2</sub> and CH<sub>4</sub> is introduced into the system. A coupling parameter  $\lambda$  is used to compute interactions between a fractional molecule and other molecules with  $\lambda$  ranging from 0 to 1, where 0 represents no interaction (ideal gas) and 1 signifies full interaction (whole molecule). Ewald summation is utilized to calculate Coulombic interactions, with an electrostatic cutoff of 14 Å and parameters  $\kappa \times L = 5.6$  and  $K_{\text{max}} = 5$ . The following trial moves are performed: translation and rotation of the molecules, volume change of the simulation box, changing the internal conformation of the molecules and  $\lambda$ -move in the frequency ratio 30 : 30 : 10 : 10 : 20. The system is equilibrated over at least 2 000 000 MC cycles, followed by another 100 000 MC cycles for the production run.

**3.3.3. GEMC simulations using configurational bias moves.** The GEMC-NVT technique is used to compute the Henry's law constant of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K. Here, simulations have been performed using the MCCC'S Towhee version 8.2.3 software package.<sup>21</sup> In the GEMC-NVT technique, there are two separate simulation boxes, one for the liquid phase and the other for the vapour phase, and an interface is absent.<sup>32</sup> 1000 levulinic acid (solvent) molecules and 1, 2, 4 or 8 CO<sub>2</sub> or CH<sub>4</sub> (solute) molecules are used for the estimation of Henry's law constant. Periodic boundary conditions have been implemented in all three dimensions. The LJ potential (with a cutoff distance of 14 Å and analytical tail corrections) has been used to model the nonbonded van der Waals interactions. Ewald summation is utilized to calculate Coulombic interactions, with an electrostatic cutoff of 14 Å and parameters  $\kappa \times L = 5.6$  and  $K_{\text{max}} = 5$ .<sup>33</sup> The trial moves in GEMC-NVT are volume exchange moves, rotational-bias MC inter-box molecule transfer moves, CBMC inter-box molecule transfer moves, intra-box CBMC re-insertion moves, aggregation volume bias moves, CBMC regrowth moves, centre of mass translation and rotation moves in the frequency ratio 10 : 5 : 5 : 5 : 20 : 25 : 25. The total volume of the two simulation boxes is adjusted such that on average approximately half of the total number of solute molecules can be found in the vapour box and half in the liquid box. The system is allowed to equilibrate for at least 450 000 MC cycles with averages computed over 50 000 MC cycles of the production run.

## 4. Results and discussion

### 4.1. Liquid phase density

The liquid density of levulinic acid is computed at 313.15 K and 100 kPa using single box NPT MC simulations *via* CBMC and

**Table 1** Comparison of simulation generated densities of levulinic acid with the literature reported experimental value.<sup>34</sup> For the simulation results, the uncertainties in the averages are reported as the standard deviations in the underlined digit(s) which are provided in parentheses

Simulation (kg m <sup>-3</sup> )	Experiment <sup>34</sup> (kg m <sup>-3</sup> )	% Relative deviation
CBMC	1123	4.1%
CFCMC	1082(3)	3.6%

CFCMC approaches separately. The values of the liquid densities are reported in Table 1 and compared with the experimental value of 1123 kg m<sup>-3</sup> reported in the literature.<sup>34</sup> The density of levulinic acid predicted *via* NPT Metropolis MC simulations in conjunction with the CBMC approach and the CFCMC approach is in good agreement with the experimental data (within 4.5%), with the value from the CFCMC approach being in better agreement with the experimental value. However, the values from the two approaches do not overlap within the statistical uncertainties associated with each average value; the difference between the two average values is less than 0.7%. This suggests the suitability of the TraPPE-UA force field and its parameters to model the interactions present in levulinic acid in the liquid phase.

#### 4.2. Henry's law constants *via* the Widom test-particle insertion technique

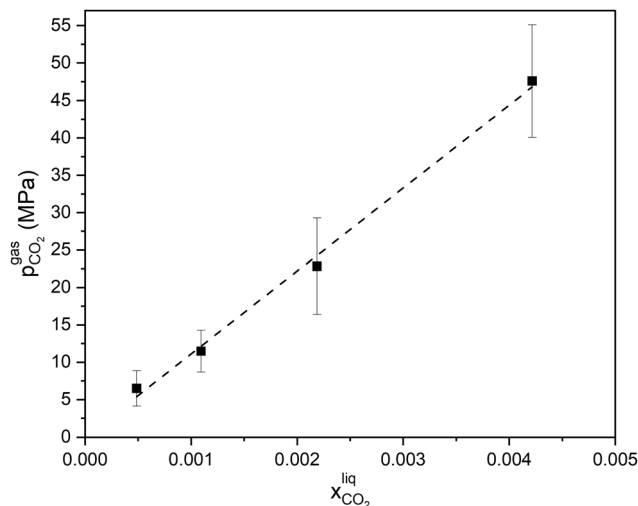
For the estimation of Henry's law constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K and 101.325 kPa, we have carried out NPT simulations using the Widom TPI technique *via* the use of both the CBMC and CFCMC algorithms. Table 2 shows the values of Henry's law constants predicted using CBMC and CFCMC approaches. The values of the Henry's law constant of CO<sub>2</sub> in levulinic acid using CBMC and CFCMC approaches are in excellent agreement. The Henry's constants for the same solvent (*i.e.*, levulinic acid) with CH<sub>4</sub> as the solute using both CBMC and CFCMC approaches are found to be close and within the error bars. We note here that there is no reported value for the Henry's constant of CO<sub>2</sub> or CH<sub>4</sub> in levulinic acid in the literature for comparison.

#### 4.3. Henry's law constants *via* GEMC-NVT simulations

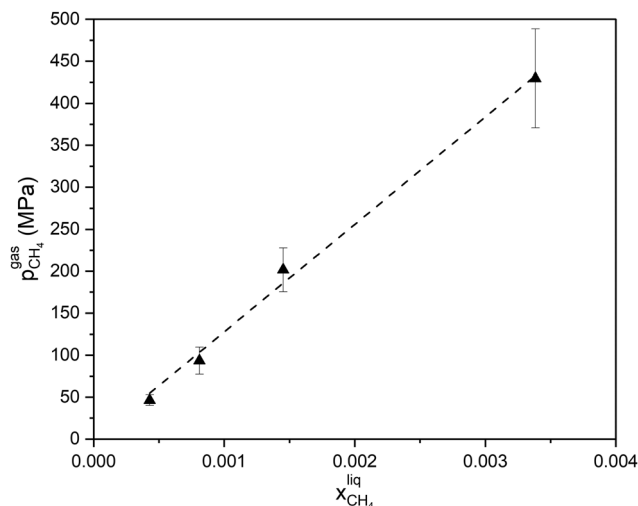
Additionally, we have carried out GEMC-NVT simulations using CBMC moves to calculate Henry's law constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid. Fig. 1 and 2 show the relationship between the partial pressure of CO<sub>2</sub> (modelled using the

**Table 2** Values of Henry's law constants (MPa) for CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K and 101.325 kPa using CBMC and CFCMC approaches in the NPT ensemble *via* the Widom TPI method. For the simulation results, the uncertainties in the averages are reported as the standard deviations in the underlined digit(s) which are provided in parentheses

Solute	CBMC	CFCMC
CO <sub>2</sub>	12(1)	12.1(4)
CH <sub>4</sub>	113(19)	118(20)



**Fig. 1** Partial pressure versus liquid-phase mole fraction for CO<sub>2</sub> in levulinic acid obtained *via* GEMC-NVT simulations.



**Fig. 2** Partial pressure versus liquid-phase mole fraction for CH<sub>4</sub> in levulinic acid obtained *via* GEMC-NVT simulations.

TraPPE-small force field) and CH<sub>4</sub> (modelled using the TraPPE-UA force field), and their corresponding liquid-phase mole fractions in levulinic acid (solvent) at 313.15 K. The four markers in each plot illustrate the simulation results for systems corresponding to the number of solute molecules varying as 1, 2, 4 and 8. The average values of Henry's law constants obtained from  $p_{\text{solute}}^{\text{gas}}$  and  $x_{\text{solute}}^{\text{liq}}$  for each solute have been determined from the GEMC-NVT simulations. Furthermore, a linear regression of  $p_{\text{solute}}^{\text{gas}}$  vs.  $x_{\text{solute}}^{\text{liq}}$  that is constrained to pass through the origin has been carried out and the slope of this fitted straight line provides the value of the Henry's law constant. We note here that the results for all four concentrations are accurately represented by a linear fit for both the solutes, *viz.*, CO<sub>2</sub> (refer to Fig. 1) and CH<sub>4</sub> (refer to Fig. 2). The values of Henry's law constants derived from the GEMC-NVT simulations are presented in Table 3. In the case of CO<sub>2</sub> as a

**Table 3** Values of Henry's law constant (MPa) for CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K using GEMC-NVT simulations via the CBMC approach. The uncertainties in the averages are reported as the standard deviations in the underlined digit(s) which are provided in parentheses

Solute	Henry's law constant (MPa)
CO <sub>2</sub>	11(2)
CH <sub>4</sub>	128(18)

solute, the Henry's law constant is determined to be 11 MPa, whereas the same for CH<sub>4</sub> as a solute is 128 MPa. These are similar to the corresponding values obtained from CBMC and CFCMC approaches via the Widom TPI method.

The value of the Henry's law constant for CH<sub>4</sub> is higher than that for CO<sub>2</sub> in levulinic acid as obtained from all the methods discussed above (*i.e.*, CBMC and CFCMC simulations in the NPT ensemble via the Widom TPI method and GEMC-NVT simulations using CBMC moves). This suggests less solubility of CH<sub>4</sub> compared to CO<sub>2</sub> in levulinic acid. We again note here that there are no experimental data nor any simulation studies available in the literature for the Henry's constant of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid for comparison here.

1000 MC cycles performed using the CBMC and CFCMC approaches in the NPT ensemble simulations (in the case of CO<sub>2</sub> in levulinic acid) require approximately 5304 and 2856 s of CPU time, respectively, and GEMC simulations with the CBMC approach in the case of 1000 levulinic acid molecules with 8 CO<sub>2</sub> molecules require about 44 076 s of CPU time on an Intel<sup>®</sup> Xenon<sup>®</sup> Silver processor at 2.40 GHz. Based on the total computational time requirement, the Widom TPI method in the NPT ensemble in conjunction with the CFCMC algorithm is recommended to compute the Henry's law constant via molecular simulations. We note here that the GEMC simulations allow us to additionally compute the Gibbs free energy of transfer of the solute.

#### 4.4. Gibbs free energy of transfer of the solute

We have calculated the Gibbs free energy of transfer ( $\Delta G_{\text{trans}}$ ) of the solute (CO<sub>2</sub> or CH<sub>4</sub>) from the vapour box (initially containing 1, 2, 4 or 8 solute molecules) to the liquid box (containing 1000 levulinic acid molecules) using the results obtained from the GEMC-NVT simulations. This can be expressed in terms of the number densities of the solute in the two phases as given in eqn (13).<sup>35</sup>

$$\Delta G_{\text{trans}} = -RT \ln \frac{\rho_{\text{liquid}}(X)}{\rho_{\text{vapour}}(X)} \quad (13)$$

where  $R$  is the universal gas constant,  $T$  is the absolute temperature, and  $\rho_{\text{liquid}}$  and  $\rho_{\text{vapour}}$  are the number densities of the solute ( $X$ ) in the liquid and vapour boxes, respectively.

Table 4 reports the values of  $\Delta G_{\text{trans}}$  of solute from the vapour phase to the liquid phase for various number ratios of levulinic acid to solute in the case of each solute. The  $\Delta G_{\text{trans}}$  of CO<sub>2</sub> from the vapour phase to the liquid levulinic acid phase is negative indicating a preference for the liquid phase in contrast to the positive values noted for CH<sub>4</sub>, confirming our previously stated results for the Henry's law constant values.

**Table 4**  $\Delta G_{\text{trans}}$  for CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K from GEMC-NVT simulations

Levulinic acid : solute (in terms of number of molecules)	$\Delta G_{\text{trans}}$ (kJ mol <sup>-1</sup> )	
	CO <sub>2</sub>	CH <sub>4</sub>
1000 : 1	-4(1)	3.9(8)
1000 : 2	-3.1(4)	4(1)
1000 : 4	-3(1)	4.6(9)
1000 : 8	-2.2(7)	5(1)

#### 4.5. Radial distribution functions

Molecular simulations also allow us to gain molecular-level insights into the underlying reasons for the relative solubilities of the solutes, *viz.*, CO<sub>2</sub> and CH<sub>4</sub>, in the solvent, levulinic acid. We have performed Metropolis MC simulations in the NVT ensemble to investigate the solvation of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K with a system consisting of 300 levulinic acid molecules and 1 molecule of CO<sub>2</sub> or CH<sub>4</sub> (to maintain the dilute nature of the system such that solute-solute interactions are absent). Prior to the NVT ensemble simulations, we have performed NPT ensemble MC simulations at 313.15 K and 101.325 kPa for at least 250 000 cycles to determine the average box volume (which is then input to the NVT MC simulations). The densities are found to be 1076(2) kg m<sup>-3</sup> for the mixtures of either CO<sub>2</sub> or CH<sub>4</sub> as solutes in levulinic acid. We have first plotted solute-solvent COM RDFs and the corresponding NIs for levulinic acid-solute interactions. The proximity of CH<sub>4</sub> and CO<sub>2</sub> molecules to the individual groups of the levulinic acid molecule (a ketoacid containing two functional groups: keto-group and carboxylic acid group, see Scheme 1) is assessed via site-site intermolecular RDFs and the corresponding NIs. We discuss each of these types of RDFs in separate sub-sections.

##### 4.5.1. COM RDFs for levulinic acid-solute interactions.

The levulinic acid COM-solute COM RDFs and their corresponding NIs for both the solutes (*i.e.*, CO<sub>2</sub> and CH<sub>4</sub>) are presented in Fig. 3(a) and (b), respectively. The structural features of the RDFs for both the solutes are similar. There is a distinct first peak, with peak heights of  $g(r) = 1.54$  and  $g(r) = 1.67$ , for CO<sub>2</sub> and CH<sub>4</sub>, respectively. The solvation of CO<sub>2</sub> or CH<sub>4</sub> in levulinic acid molecules is described by the solvation number which is computed by integrating the RDF up to the separation distance at which the first minimum occurs (7.15 Å and 7.65 Å for CO<sub>2</sub> and CH<sub>4</sub>, respectively). The solvation numbers for CO<sub>2</sub> and CH<sub>4</sub> are 8.3 and 9.9, respectively. The solvation number is higher for CH<sub>4</sub> than CO<sub>2</sub> (noting that the separation distance is also higher for CH<sub>4</sub>) which is opposite to the trends in  $\Delta G_{\text{trans}}$  of the solute from the vapour phase to the liquid phase. To investigate this further, we have analysed the site-site intermolecular RDFs for these systems.

##### 4.5.2. Site-site intermolecular RDFs for CO<sub>2</sub>-levulinic acid interactions

**4.5.2.1. C(CO<sub>2</sub>)-levulinic acid interactions.** The site-site intermolecular RDFs for the carbon of CO<sub>2</sub> represented as C(CO<sub>2</sub>) and the various united atoms of the carbon containing groups

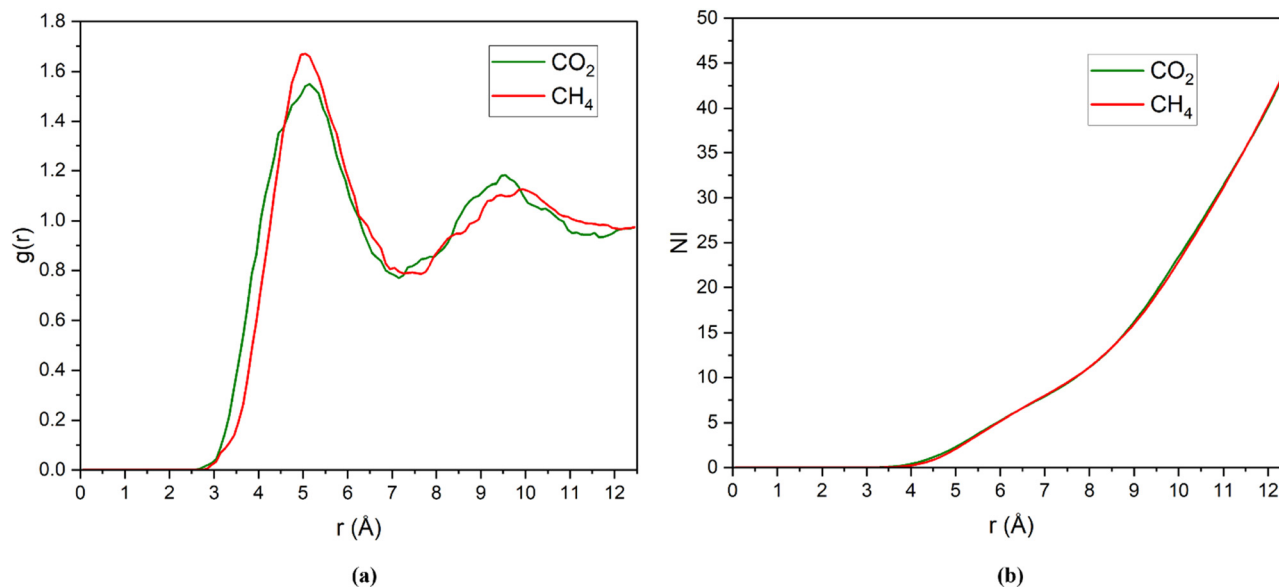


Fig. 3 (a) Levulinic acid COM–solute (CO<sub>2</sub> or CH<sub>4</sub>) COM RDFs and (b) their corresponding NIs.

of levulinic acid (*i.e.*, CH<sub>3</sub>(1), C2, CH<sub>2</sub>(3), CH<sub>2</sub>(5) and C6 groups as shown in Scheme 1) are plotted in Fig. 4(a). Fig. 4(b) shows the RDFs for the different O atoms of levulinic acid (*i.e.*, O4, O7, and O8 in Scheme 1) and the H9 atom (refer to Scheme 1) with C(CO<sub>2</sub>). We find that the first peak heights are above unity for all the site–site intermolecular RDFs in Fig. 4(a) and (b). The first peak heights from site–site RDFs arranged in descending order are as follows: CH<sub>3</sub>(1) > C2 > CH<sub>2</sub>(5) > C6 > O4 > CH<sub>2</sub>(3) > O7 > H9 > O8. The highest peak height is observed in the case of C(CO<sub>2</sub>)–CH<sub>3</sub>(1) (at the separation distance of 4.543 Å). The highest peak height for the C(CO<sub>2</sub>)–CH<sub>3</sub>(1) site–site intermolecular RDF corresponds to the highest value of

well depth (of the 12-6 LJ potential which describes the van der Waals interactions) for C(CO<sub>2</sub>)–CH<sub>3</sub>(1) interactions obtained using the Lorentz–Berthelot combining rule (eqn (4)). The value of NI calculated up to a separation distance of 6.08 Å (distance corresponding to the first minimum) is 5.42 (refer to Fig. 5). We have also shown the snapshots from our simulation for two different configurations of the local environment around the CO<sub>2</sub> molecule in Fig. 5. We see that a CO<sub>2</sub> molecule is surrounded by six levulinic acid molecules at a separation distance of 6.08 Å.

Fig. 4(b) shows that the heights of the first peak in the C(CO<sub>2</sub>)–O4, C(CO<sub>2</sub>)–O7, and C(CO<sub>2</sub>)–O8 RDFs are in descending



Fig. 4 Site–site intermolecular RDFs for the C(CO<sub>2</sub>)–levulinic acid system: (a) CH<sub>3</sub>(1), C2, CH<sub>2</sub>(3), CH<sub>2</sub>(5) and C6 groups of levulinic acid and (b) O4, O7, O8 and H9 groups of levulinic acid.



Fig. 5 VMD snapshots (a) and (b) from our simulation show the levulinic acid molecules in the local environment around  $\text{CO}_2$  [levulinic acid molecules are graphically represented as lines and the atoms in the  $\text{CO}_2$  molecule are shown as spheres. Blue denotes carbon or carbon containing united atoms, red represents oxygen, and hydrogen is shown in white].

order and the partial charges (in the Coulomb potential modelling the electrostatic interactions listed in Table S1 of the SI) of O4, O7 and O8 atoms also exhibit a decreasing trend. Moreover, the first peaks in the RDFs of the  $\text{C}(\text{CO}_2)$  with O4, O7, O8 and H9 (Fig. 4(b)) appear before the first peaks in the RDFs of the  $\text{C}(\text{CO}_2)$  with  $\text{CH}_3(1)$ , C2,  $\text{CH}_2(3)$ ,  $\text{CH}_2(5)$  and C6 united atoms (Fig. 4(a)) suggesting that the  $\text{C}(\text{CO}_2)$  atom is near the oxygen atoms of the carboxylic acid and the keto groups of the levulinic acid molecule. We note here that the minimum separation distances as calculated from the Lorentz–Berthelot combining rule (eqn (3)) are less in the case of van der Waals interactions (represented by the 12-6 LJ potential) of  $\text{C}(\text{CO}_2)$  with O4, O7, O8 and H9 than with carbon containing united atoms of levulinic acid. Levulinic acid molecules in the liquid

phase adopt different conformations, some of which are illustrated in Fig. 5 and we observe from Fig. 5(b) that the  $\text{CO}_2$  molecule is in close proximity to the carboxylic acid group of levulinic acid. We note the presence of two peaks in the  $\text{C}(\text{CO}_2)$ –O8 and the  $\text{C}(\text{CO}_2)$ –H9 RDFs with the first peak in both cases appearing at a separation distance of 3.17 Å. In the  $\text{C}(\text{CO}_2)$ –O8 RDF, the two peak heights (with  $g(r) = 1.237$  and  $g(r) = 1.244$ ) are almost the same, implying that there are two preferred locations of the  $\text{CO}_2$  molecule with respect to O8. A third peak with  $g(r) = 1.173$  at 5.53 Å is also observed in the  $\text{C}(\text{CO}_2)$ –O8 RDF.

**4.5.2.1.  $\text{O}(\text{CO}_2)$ –levulinic acid interactions.** Additionally, we have computed the site–site intermolecular RDFs for the oxygen of  $\text{CO}_2$  represented as  $\text{O}(\text{CO}_2)$  and the different united atoms

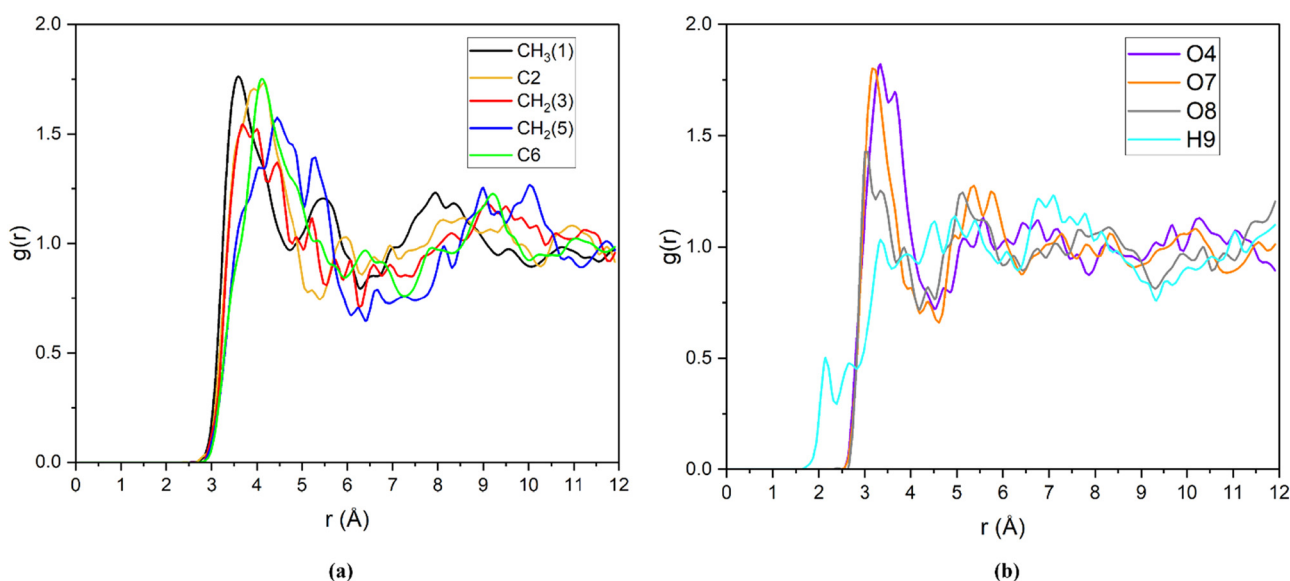


Fig. 6 Site–site intermolecular RDFs for the  $\text{O}(\text{CO}_2)$ –levulinic acid system: (a)  $\text{CH}_3(1)$ , C2,  $\text{CH}_2(3)$ ,  $\text{CH}_2(5)$  and C6 groups of levulinic acid and (b) O4, O7, O8 and H9 groups of levulinic acid.

representing the carbon containing groups of levulinic acid (Scheme 1) and these RDFs are presented in Fig. 6(a). The plots of the RDFs for O4, O7, O8 and H9 of levulinic acid (Scheme 1) with O(CO<sub>2</sub>) are shown in Fig. 6(b). The central C(CO<sub>2</sub>) atom is bonded to the two O(CO<sub>2</sub>) atoms in the CO<sub>2</sub> molecule and any association of C(CO<sub>2</sub>) with the sites on the levulinic acid molecule will involve associations with the two O(CO<sub>2</sub>) atoms also. Here, the heights of the first peaks, except for the O(CO<sub>2</sub>)–H9 RDF, are observed to be greater than one. The value of NI calculated from the O(CO<sub>2</sub>)–CH<sub>3</sub>(1) RDF up to the first coordination shell (with a separation distance of 4.71 Å) is 2.07. As in Fig. 4(b), we find that Fig. 6(b) also shows that the heights of the first peak in the O(CO<sub>2</sub>)–O4, O(CO<sub>2</sub>)–O7, and O(CO<sub>2</sub>)–O8 RDFs in descending order are in correspondence with the

decreasing trend in the partial charges (in the Coulomb potential modelling the electrostatic interactions given in Table S1 of the SI) of O4, O7 and O8 atoms.

The O(CO<sub>2</sub>)–O8 RDF exhibits features distinct from the C(CO<sub>2</sub>)–O8 RDF, with the first peak (with  $g(r) = 1.428$  which is slightly higher than the first two peaks of the C(CO<sub>2</sub>)–O8 RDF) displaying one pronounced peak at a separation distance ( $r$ ) of 3.00 Å, a first shoulder with  $g(r) > 1.0$  at  $r = 3.34$  Å and a second shoulder which appears with  $g(r) \approx 1.0$  at  $r = 3.86$  Å. From Fig. 4(b), we note that the third peak of the C(CO<sub>2</sub>)–O8 RDF (with peak height  $g(r) = 1.173$ ) is at  $r = 5.23$  Å, and this peak corresponds to the second peak (with a slightly higher peak height of  $g(r) = 1.234$ ) of the O(CO<sub>2</sub>)–O8 RDF at  $r = 5.06$  Å in Fig. 6(b).



Fig. 7 Site–site intermolecular RDFs for the CH<sub>4</sub> and levulinic acid system: (a) CH<sub>4</sub> with CH<sub>3</sub>(1), C2, CH<sub>2</sub>(3), CH<sub>2</sub>(5) and C6 groups of levulinic acid and (b) CH<sub>4</sub> with O4, O7, O8 and H9 groups of levulinic acid.



Fig. 8 VMD snapshots ((a) and (b)) from our simulation show the levulinic acid molecules in the local environment around CH<sub>4</sub> [levulinic acid molecules are graphically represented as lines and the CH<sub>4</sub> united atom is shown as an orange sphere. Blue denotes carbon or carbon containing united atoms, red represents oxygen, and hydrogen is shown in white].

**4.5.3. Site-site intermolecular RDFs for CH<sub>4</sub>–levulinic acid interactions.** Fig. 7(a) illustrates the site-site intermolecular RDFs for CH<sub>4</sub> and the various united atoms of the carbon containing groups in levulinic acid (*i.e.*, CH<sub>3</sub>(1), C2, CH<sub>2</sub>(3), CH<sub>2</sub>(5) and C6 groups in Scheme 1). The RDFs for CH<sub>4</sub> with the O4, O7, O8 and H9 atoms of levulinic acid (refer to Scheme 1) are plotted in Fig. 7(b). The heights of the first peak in all the CH<sub>4</sub>–levulinic acid united atom RDFs are above unity. The descending order in the peak heights from site-site RDFs shown in Fig. 7 is CH<sub>3</sub>(1) > C2 > C6 > CH<sub>2</sub>(5) > CH<sub>2</sub>(3) > O7 > O4 > O8 > H9. The CH<sub>4</sub>–CH<sub>3</sub>(1) intermolecular RDF has one distinct and dominant peak at  $r = 4.03 \text{ \AA}$  with the peak height ( $g(r) = 2.59$ ) being much higher than those of all the other peaks in Fig. 7. This highest first peak height of the CH<sub>4</sub>–CH<sub>3</sub>(1) RDF also corresponds to the highest value of the 12-6 LJ well depth computed *via* the Lorentz–Berthelot combining rule using eqn (4). As described in Section 2, the van der Waals interactions are modelled using the 12-6 LJ potential. There are also no partial charges present on either CH<sub>4</sub> or the CH<sub>3</sub>(1) united atom, and hence, the nonbonded interactions between these two united atoms are only *via* the van der Waals interactions. The value of NI calculated up to a separation distance of 5.23 Å (where the first minimum of the CH<sub>4</sub>–CH<sub>3</sub>(1) RDF occurs) is 3.95. Fig. 8 presents two different simulation snapshots illustrating the local environment around the CH<sub>4</sub> molecule, where the CH<sub>4</sub> molecule is surrounded by four levulinic acid molecules within a separation distance of 5.23 Å.

## 5. Conclusions

In this study, the Henry's law constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid at 313.15 K and 101.325 kPa have been estimated *via* three different Monte Carlo approaches. The interactions present in levulinic acid and CH<sub>4</sub> are modelled using the TraPPE-UA force field, whereas the TraPPE-small force field is used to describe interactions present in CO<sub>2</sub>. First, the suitability of the force field is checked by computing the liquid density of levulinic acid at 313.15 K and 100 kPa using NPT MC simulations *via* both CBMC and CFCMC approaches. We find that the values of the liquid densities estimated using CBMC and CFCMC approaches are in good agreement with the experimentally reported value (within 4.5%) with the prediction *via* the CFCMC approach being closer to the experimental value. First, we have implemented the Widom TPI method (which is based on the calculation of excess chemical potential) in the NPT ensemble using two Monte Carlo algorithms, *viz.*, CBMC and CFCMC. We find that the values of Henry's law constant *via* both CBMC and CFCMC are in excellent agreement for each solute. Additionally, we have performed GEMC-NVT simulations including the CBMC moves to calculate Henry's law constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid, because this method does not require calculation of excess chemical potential. From the slope of the line fitted to the plot of  $p_{\text{solute}}^{\text{gas}}$  vs.  $x_{\text{solute}}^{\text{liq}}$ , the values of the Henry's law constant in the case of

each solute are calculated. The values obtained *via* GEMC-NVT simulations are similar to the values obtained from CBMC and CFCMC approaches *via* the Widom TPI method. Moreover, from GEMC-NVT simulations,  $\Delta G_{\text{trans}}$  of solutes (*viz.*, CO<sub>2</sub> and CH<sub>4</sub>) can also be estimated. The negative value of  $\Delta G_{\text{trans}}$  of CO<sub>2</sub> from the vapour phase to the liquid (levulinic acid) phase suggests an obvious preference of CO<sub>2</sub> for the liquid levulinic acid phase over the vapour phase. We note that there are no experimental data nor any prior simulation studies available in the literature for Henry's constants of CO<sub>2</sub> and CH<sub>4</sub> in levulinic acid for comparison here. This knowledge of Henry's law constants is useful in the various industrial processes for design, optimization and modelling of separation equipment.

We have also studied the solvation of CH<sub>4</sub> and CO<sub>2</sub> molecules in liquid levulinic acid by performing single box NVT simulations at 313.15 K using one solute molecule (*viz.*, CO<sub>2</sub> or CH<sub>4</sub>) in 300 levulinic acid molecules such that solute–solute interactions are absent. The site-site intermolecular RDFs for C(CO<sub>2</sub>), O(CO<sub>2</sub>) and CH<sub>4</sub> with the various united atoms representing the carbon containing groups of levulinic acid and those with O4, O7, O8 and H9 atoms of levulinic acid (Scheme 1) are computed to investigate the local environment around the solute molecule. From the solute–solvent site-site RDFs, we can conclude that CO<sub>2</sub> shows a preference for the carboxylic acid and keto groups of levulinic acid, whereas CH<sub>4</sub> shows a tendency to reside in the proximity of the CH<sub>3</sub>(1) group of levulinic acid. The NIs calculated from the integration of the site-site intermolecular RDFs up to the first minima of the C(CO<sub>2</sub>)–CH<sub>3</sub>(1) and CH<sub>4</sub>–CH<sub>3</sub>(1) RDFs show that the solvation number of CO<sub>2</sub> in levulinic acid is more than that of CH<sub>4</sub>. This observation (which is in contrast to the respective solvation numbers computed from the solute–solvent COM RDFs) aligns well with the lower Henry's law constant of CO<sub>2</sub> in levulinic acid as compared to CH<sub>4</sub> at 313.15 K and atmospheric pressure. This is also consistent with the negative value of  $\Delta G_{\text{trans}}$  of CO<sub>2</sub> from the vapour phase to the liquid phase, confirming a preference for the dissolution of CO<sub>2</sub> in levulinic acid over CH<sub>4</sub>.

## Author contributions

Prasil Kapadiya: conceptualization (equal); data curation (lead); formal analysis (lead); funding acquisition; investigation (equal); methodology (equal); software (lead); validation; visualization (lead); roles/writing – original draft (lead); and writing – review & editing (equal). Jhumpa Adhikari: conceptualization (equal); data curation (supporting); formal analysis (supporting); funding acquisition; investigation (equal); methodology (equal); project administration; resources (lead); software (supporting); supervision (lead); validation; visualization; roles/writing – original draft (supporting); and writing – review & editing (equal).

## Conflicts of interest

There are no conflicts of interest to declare.

## Data availability

Data will be made available on request.

Supplementary information (SI): nonbonded and bonded interaction parameters for the TraPPE-UA force field for levulinic acid and CH<sub>4</sub> are provided in Tables S1 and S2, respectively. The parameters for the TraPPE-small force field in the case of CO<sub>2</sub> are reported in Table S3. Site–site intermolecular radial distribution functions for CO<sub>2</sub>–levulinic acid and CH<sub>4</sub>–levulinic acid systems are presented in Fig. S1–S3. See DOI: <https://doi.org/10.1039/d5cp02630j>.

## Acknowledgements

P. K. acknowledges funding support from the Prime Minister's Research Fellowship (PMRF), Ministry of Education, Government of India. Computational resources and funding for this project have been provided by the Science and Engineering Research Board (SERB), Department of Science and Technology, India (Sanction Order #: CRG/2022/000148 of Anusandhan National Research Foundation (ANRF)).

## References

- 1 S. I. Sandler, *Chemical, Biochemical, and Engineering Thermodynamics*, 5th edn, John Wiley & Sons, 2017.
- 2 L. Zhang and J. I. Siepmann, Direct calculation of Henry's law constants from Gibbs ensemble Monte Carlo simulations: Nitrogen, oxygen, carbon dioxide and methane in ethanol, *Theor. Chem. Acc.*, 2006, **115**(5), 391–397, DOI: [10.1007/s00214-005-0073-1](https://doi.org/10.1007/s00214-005-0073-1).
- 3 E. C. Cichowski, T. R. Schmidt and J. R. Errington, Determination of Henry's law constants through transition matrix Monte Carlo simulation, *Fluid Phase Equilib.*, 2005, **236**(1–2), 58–65, DOI: [10.1016/j.fluid.2005.05.001](https://doi.org/10.1016/j.fluid.2005.05.001).
- 4 K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T. H. Bae and J. R. Long, Carbon dioxide capture in metal–organic frameworks, *Chem. Rev.*, 2012, **112**(2), 724–781, DOI: [10.1021/cr2003272](https://doi.org/10.1021/cr2003272).
- 5 Q. Chen, M. Ramdin and T. J. Vlugt, Solubilities of CO<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, CO, H<sub>2</sub>, N<sub>2</sub>, N<sub>2</sub>O, and H<sub>2</sub>S in commercial physical solvents from Monte Carlo simulations, *Mol. Simul.*, 2023, **49**(13–14), 1341–1349, DOI: [10.1080/08927022.2023.2228918](https://doi.org/10.1080/08927022.2023.2228918).
- 6 M. Ramdin, S. P. Balaji, J. M. Vicent-Luna, J. J. Gutierrez-Sevillano, S. Calero, T. W. de Loos and T. J. Vlugt, Solubility of the precombustion gases CO<sub>2</sub>, CH<sub>4</sub>, CO, H<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>S in the ionic liquid [bmim][Tf<sub>2</sub>N] from Monte Carlo simulations, *J. Phys. Chem. C*, 2014, **118**(41), 23599–23604, DOI: [10.1021/jp5080434](https://doi.org/10.1021/jp5080434).
- 7 B. Widom, Some topics in the theory of fluids, *J. Chem. Phys.*, 1963, **39**(11), 2808–2812, DOI: [10.1063/1.1734110](https://doi.org/10.1063/1.1734110).
- 8 D. A. Kofke and P. T. Cummings, Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation, *Mol. Phys.*, 1997, **92**(6), 973–996, DOI: [10.1080/002689797169600](https://doi.org/10.1080/002689797169600).
- 9 N. Lu, J. K. Singh and D. A. Kofke, Appropriate methods to combine forward and reverse free-energy perturbation averages, *J. Chem. Phys.*, 2003, **118**(7), 2977–2984, DOI: [10.1063/1.1537241](https://doi.org/10.1063/1.1537241).
- 10 N. Lu and D. A. Kofke, Optimal intermediates in staged free energy calculations, *J. Chem. Phys.*, 1999, **111**(10), 4414–4423, DOI: [10.1063/1.479206](https://doi.org/10.1063/1.479206).
- 11 M. Lísal, W. R. Smith and K. Aim, Analysis of Henry's constant for carbon dioxide in water via Monte Carlo simulation, *Fluid Phase Equilib.*, 2005, **228**, 345–356, DOI: [10.1016/j.fluid.2005.03.005](https://doi.org/10.1016/j.fluid.2005.03.005).
- 12 A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov and P. N. Vorontsov-Velyaminov, New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles, *J. Chem. Phys.*, 1992, **96**(3), 1776–1783, DOI: [10.1063/1.462133](https://doi.org/10.1063/1.462133).
- 13 A. P. Lyubartsev, A. Laaksonen and P. N. Vorontsov-Velyaminov, Determination of free energy from chemical potentials: Application of the expanded ensemble method, *Mol. Simul.*, 1996, **18**(1–2), 43–58, DOI: [10.1080/08927029608022353](https://doi.org/10.1080/08927029608022353).
- 14 A. P. Lyubartsev, O. K. Førrisdahl and A. Laaksonen, Solvation free energies of methane and alkali halide ion pairs: An expanded ensemble molecular dynamics simulation study, *J. Chem. Phys.*, 1998, **108**(1), 227–233, DOI: [10.1063/1.475374](https://doi.org/10.1063/1.475374).
- 15 J. I. Siepmann and D. Frenkel, Configurational bias Monte Carlo: A new sampling scheme for flexible chains, *Mol. Phys.*, 1992, **75**(1), 59–70, DOI: [10.1080/00268979200100061](https://doi.org/10.1080/00268979200100061).
- 16 W. Shi and E. J. Maginn, Continuous fractional component Monte Carlo: An adaptive biasing method for open system atomistic simulations, *J. Chem. Theory Comput.*, 2007, **3**(4), 1451–1463, DOI: [10.1021/ct7000039](https://doi.org/10.1021/ct7000039).
- 17 M. N. Rosenbluth and A. W. Rosenbluth, Monte Carlo calculation of the average extension of molecular chains, *J. Chem. Phys.*, 1955, **23**(2), 356–359, DOI: [10.1063/1.1741967](https://doi.org/10.1063/1.1741967).
- 18 J. K. Shah and E. J. Maginn, Monte Carlo simulations of gas solubility in the ionic liquid 1-*n*-butyl-3-methylimidazolium hexafluorophosphate, *J. Phys. Chem. B*, 2005, **109**(20), 10395–10405, DOI: [10.1063/1.1741967](https://doi.org/10.1063/1.1741967).
- 19 H. S. Salehi, R. Hens, O. A. Moulto and T. J. H. Vlugt, Computation of gas solubilities in choline chloride urea and choline chloride ethylene glycol deep eutectic solvents using Monte Carlo simulations, *J. Mol. Liq.*, 2020, **316**, 113729, DOI: [10.1016/j.molliq.2020.113729](https://doi.org/10.1016/j.molliq.2020.113729).
- 20 C. H. Zhou, X. Xia, C. X. Lin, D. S. Tong and J. Beltramini, Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels, *Chem. Soc. Rev.*, 2011, **40**(11), 5588–5617.
- 21 M. G. Martin, MCCCSTowhee: a tool for Monte Carlo molecular simulation, *Mol. Simul.*, 2013, **39**(14–15), 1212–1222, DOI: [10.1080/08927022.2013.828208](https://doi.org/10.1080/08927022.2013.828208).
- 22 R. Hens, A. Rahbari, S. Caro-Ortiz, N. Dawass, M. Erdos, A. Poursaeidesfahani, H. S. Salehi, A. T. Celebi, M. Ramdin, O. A. Moulto and D. Dubbeldam, Brick-CFCMC: Open source software for Monte Carlo simulations of phase and

- reaction equilibria using the continuous fractional component method, *J. Chem. Inf. Model.*, 2020, **60**(6), 2678–2682, DOI: [10.1021/acs.jcim.0c00334](https://doi.org/10.1021/acs.jcim.0c00334).
- 23 J. M. Stubbs, J. J. Potoff and J. I. Siepmann, Transferable potentials for phase equilibria. 6. United-atom description for ethers, glycols, ketones, and aldehydes, *J. Phys. Chem. B*, 2004, **108**(45), 17596–17605, DOI: [10.1021/jp049459w](https://doi.org/10.1021/jp049459w).
- 24 G. Kamath, F. Cao and J. J. Potoff, An improved force field for the prediction of the vapor-liquid equilibria for carboxylic acids, *J. Phys. Chem. B*, 2004, **108**(37), 14130–14136, DOI: [10.1021/jp048581s](https://doi.org/10.1021/jp048581s).
- 25 T. Chakraborti, A. Desouza and J. Adhikari, Prediction of Thermodynamic Properties of Levulinic Acid via Molecular Simulation Techniques, *ACS Omega*, 2018, **3**(12), 18877–18884, DOI: [10.1021/acsomega.8b02793](https://doi.org/10.1021/acsomega.8b02793).
- 26 M. G. Martin and J. I. Siepmann, Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes, *J. Phys. Chem. B*, 1998, **102**(14), 2569–2577, DOI: [10.1021/jp972543](https://doi.org/10.1021/jp972543).
- 27 J. J. Potoff and J. I. Siepmann, Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen, *AIChE J.*, 2001, **47**(7), 1676–1682, DOI: [10.1002/aic.690470719](https://doi.org/10.1002/aic.690470719).
- 28 B. L. Eggimann, Y. Sun, R. F. DeJaco, R. Singh, M. Ahsan, T. R. Josephson and J. I. Siepmann, Assessing the quality of molecular simulations for vapor-liquid equilibria: An analysis of the TraPPE database, *J. Chem. Eng. Data*, 2019, **65**(3), 1330–1344, DOI: [10.1021/acs.jced.9b00756](https://doi.org/10.1021/acs.jced.9b00756).
- 29 J. M. Smith, H. C. Van Ness and M. M. Abbott, *Introduction to Chemical Engineering Thermodynamics*, 7th edn, McGraw-Hill, New York, NY, 2005.
- 30 D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd edn, Academic Press, San Diego, 2001.
- 31 A. Rahbari, R. Hens, M. Ramdin, O. A. Moulto, D. Dubbeldam and T. J. H. Vlugt, Recent advances in the continuous fractional component Monte Carlo methodology, *Mol. Simul.*, 2021, **47**(10–11), 804–823, DOI: [10.1080/08927022.2020.1828585](https://doi.org/10.1080/08927022.2020.1828585).
- 32 A. Z. Panagiotopoulos, Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble, *Mol. Phys.*, 1987, **61**(4), 813–826, DOI: [10.1080/00268978700101491](https://doi.org/10.1080/00268978700101491).
- 33 MCCCSTowhee. <https://towhee.sourceforge.net/>.
- 34 H. Guerrero, C. Lafuente, F. Royo, L. Lomba and B. Giner, PpT behavior of several chemicals from biomass, *Energy Fuels*, 2011, **25**(7), 3009–3013, DOI: [10.1021/ef200653s](https://doi.org/10.1021/ef200653s).
- 35 S. J. Keasler, J. L. Lewin, J. I. Siepmann, N. M. Gryska, R. B. Ross, N. E. Schultz and M. Nakamura, Molecular insights for the optimization of solvent-based selective extraction of ethanol from fermentation broths, *AIChE J.*, 2013, **59**(8), 3065–3070, DOI: [10.1002/aic.14055](https://doi.org/10.1002/aic.14055).