



# Technical note: Estimating aqueous solubilities and activity coefficients of mono- and $\alpha,\omega$ -dicarboxylic acids using COSMO-RS-DARE

Noora Hyttinen<sup>1,2</sup>, Reyhaneh Heshmatnezhad<sup>1</sup>, Jonas Elm<sup>3</sup>, Theo Kurtén<sup>2</sup>, and Nønne L. Prisle<sup>1</sup>

<sup>1</sup>Nano and Molecular Systems Research Unit, University of Oulu, P.O. Box 3000, FI-90014 Oulu, Finland

<sup>2</sup>Department of Chemistry and Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, P.O. Box 55, FI-00014 Helsinki, Finland

<sup>3</sup>Department of Chemistry and iClimate, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark

**Correspondence:** Noora Hyttinen (noora.hyttinen@oulu.fi), Nønne L. Prisle (nonne.prisle@oulu.fi)

**Abstract.** We have used the COSMO $therm$  program to estimate activity coefficients and solubilities of mono- and  $\alpha,\omega$ -dicarboxylic acids, and water in binary acid–water systems. The deviation from ideality was found to be larger in the systems containing larger acids than in the systems containing smaller acids. We found a better agreement between estimated and experimental activity coefficients of monocarboxylic acids when the water clustering with a carboxylic acid and itself was taken into account using the dimerization, aggregation and reaction extension (COSMO-RS-DARE) of COSMO $therm$ . Based on effective equilibrium constants of different clustering reactions in the binary solutions, acid dimer formation is more dominant in systems containing larger dicarboxylic acids (C5–C8), while for monocarboxylic acids (C1–C6) and smaller dicarboxylic acids (C2–C4), hydrate formation is more favorable, especially in dilute solutions.

## 1 Introduction

Mono- and dicarboxylic acids ( $\text{CH}_3(\text{CH}_2)_{n-2}\text{COOH}$  and  $\text{COOH}(\text{CH}_2)_{m-2}\text{COOH}$ , respectively) are common atmospheric compounds that have been detected in both the gas (Kawamura et al., 2000; Fisseha et al., 2006) and aerosol phase (Fisseha et al., 2006; Verma et al., 2017; Guo et al., 2014; Hyder et al., 2012). Carboxylic acids have been detected in high abundance in various environments, such as urban (Zhao et al., 2018; Kawamura et al., 2000; Fisseha et al., 2006; Jung et al., 2010; Guo et al., 2014), semi-urban (Verma et al., 2017), marine (Kawamura and Sakaguchi, 1999; Mochida et al., 2003) and Antarctic (Kawamura et al., 1996) measurement sites. In general, small carboxylic acids ( $n \leq 3$  and  $m \leq 4$ ) are more abundant than large acids ( $n > 3$  and  $m > 4$ ) (Jung et al., 2010; Fisseha et al., 2006; Tsai and Kuo, 2013; Zhao et al., 2018; Guo et al., 2014; Kawamura et al., 2000). For example, Tsai and Kuo (2013) found that 77.2% of all carboxylic acids in fine particulate matter (PM<sub>2.5</sub>) were small carboxylic acids (formic, acetic and oxalic acid) in a broad-leaved forest in central Taiwan.

Accurate description of the different aerosol phases is important for determining parameters used in aerosol modeling, such as gas-to-particle partitioning, in particular water uptake and chemical reactivity. A large number of reactions in the aqueous aerosol phase are strongly pH dependent (Pye et al., 2020; Weber et al., 2016), but accurate predictions of aerosol acidity are highly challenging. One element to resolve is the nature and amount of acidic material dissolved in the aqueous aerosol phase.



The aqueous bulk solubility of mono- and dicarboxylic acids have been measured in multiple studies (Saxena and Hildemann, 1996; Apelblat and Manzurola, 1987, 1989, 1990; Cornils and Lappe, 2000; Song et al., 2012; Romero and Suárez, 2009; Omar and Ulrich, 2006; Brooks et al., 2002). However, the activity data of carboxylic acid–water systems is much more scarce. Jones and Bury (1927) derived the activity coefficients of formic ( $n = 1$ ), acetic ( $n = 2$ ), propanoic ( $n = 3$ ) and butanoic ( $n = 4$ ) acids in aqueous solutions at the freezing points of the binary solutions using freezing point depression measurements. Using freezing point depression measurements, activity coefficients are calculated using Lewis and Randall’s equation for non-electrolytes. Hansen et al. (1955) derived activity coefficients of acetic, propanoic and butanoic acid in water and the activity coefficients of water in acetic, propanoic, butanoic, pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acids, at 298.15 K, using partial pressure measurements. In addition, the Gibbs-Duhem equation was fitted to the experimental activity coefficients of the acetic, propanoic and butanoic acid–water systems. Most of the activity data of aqueous dicarboxylic acid systems in the literature (Davies and Thomas, 1956; Maffia and Meirelles, 2001; Peng et al., 2001; Choi and Chan, 2002; Wise et al., 2003) is only obtained from measurements on dilute solutions due to the low solubility of the acids in water.

The solubilities and activity coefficients of carboxylic acids have previously been estimated using the COSMO-RS theory (Klamt, 1995; Klamt et al., 1998; Eckert and Klamt, 2002) implemented in the COSMO $therm$  program (COSMO $therm$ , 2019). Schröder et al. (2010) estimated the aqueous solubilities of various polycarboxylic acids using the TZVP parametrization of COSMO $therm$  and found that COSMO $therm$  was able to predict the temperature dependence of the solubilities of dicarboxylic acids ( $m = 2–8$ ) well, while the absolute solubility estimates were not in a good agreement with experiments. Michailoudi et al. (2020) estimated the activity coefficients of monocarboxylic acids with even number of carbon atoms ( $n = 2, 4, 6, 8, 10, 12$ ) in the infinite dilution state. In addition, they estimated the solubility of the same acids in pure water and different aqueous electrolyte solutions. They found a good agreement between experimental and estimated aqueous solubilities of the acids with the exception of butanoic acid, which in experiments has been seen to be fully soluble (Saxena and Hildemann, 1996), while COSMO $therm$  predicted a finite solubility.

Recent work has shown that the absolute COSMO $therm$  solubility and activity coefficient estimates can be improved by excluding conformers containing intramolecular hydrogen bonds from the COSMO $therm$  calculation (Hyttinen and Prisle, 2020). However, based on the hydrogen bonding definition of COSMO $therm$ , monocarboxylic acids are not able to form intramolecular hydrogen bonds. Therefore, other methods are needed to improve COSMO $therm$  estimates of monocarboxylic acids. On the other hand, carboxylic acids are able to form hydrogen bonded dimers where two molecules are bound by two simultaneous intermolecular hydrogen bonds. These concerted multiple contacts, such as is seen in carboxylic acid dimer formation, are not captured by COSMO-RS. A dimerization, aggregation and reaction extension to the COSMO-RS theory (COSMO-RS-DARE) was developed to account for these interactions (Sachsenhauser et al., 2014). For example, Cysewski (2019) was able to improve the agreement between experimental and estimated solubilities of ethenzamide in various organic solvents using COSMO-RS-DARE.

Here we use the newly developed COSMO-RS-DARE method to estimate activity coefficients of monocarboxylic acids ( $n = 1–6$ ) and  $\alpha,\omega$ -dicarboxylic acids ( $m = 2–8$ ), and water, in binary acid–water mixtures. In addition, we estimate aqueous solubilities of the acids using both the COSMO-RS-DARE and COSMO-RS methods.



## 2 COSMOtherm calculations

We use the COSMOtherm software (release 19 and parametrization BP\_TZVPD\_FINE\_19) (COSMOtherm, 2019) to estimate  
60 the solubilities and activity coefficient of linear mono- and dicarboxylic acids in binary aqueous solutions. In addition, we compute the effective equilibrium constants of water and acid dimerization (formation of a hydrogen bonded cluster containing two water molecules or two acid molecules, respectively), and acid hydration (formation of a hydrogen bonded cluster containing one acid and one water molecule).

### 2.1 Activity coefficients

65 COSMOtherm calculates the activity coefficient ( $\gamma$ ) of compound  $i$  with mole fraction  $x_i$  using the pseudo-chemical potentials at composition  $\{x_i\}$  ( $\mu_i^*(x_i)$ ) and at the reference state ( $\mu_i^{*\circ}(x^\circ, T, P)$ ). By default, the reference state used in COSMOtherm is the pure compound (labeled as convention I (Levine, 2009)):

$$\ln \gamma_i^I(x_i) = \frac{\mu_i^*(x_i) - \mu_i^{*\circ, I}(x^\circ, T, P)}{RT} \quad (1)$$

The pseudo-chemical potential is defined using the standard chemical potential at the reference state  $\mu^\circ$ :

70 
$$\mu_i^*(x_i) = \mu_i^\circ(x^\circ, T, P) + RT \ln \gamma_i \quad (2)$$

at  $P = 10^5$  Pa reference pressure.  $T$  is the temperature (K) and  $R$  the gas constant ( $\text{kcal K}^{-1} \text{mol}^{-1}$ ).

### 2.2 Solubility

Solubilities are calculated by finding the liquid-liquid (LLE) or the solid-liquid equilibrium (SLE) of the binary liquid–water and solid–water systems, respectively. The LLE is found by solving the liquid-liquid equilibrium condition between the solvent  
75 rich phase ( $\alpha$ ) and the solute rich phase ( $\beta$ ):

$$a_i^\alpha = a_i^\beta \quad (3)$$

At the LLE, the activity ( $a_i^I(x_i) = \gamma_i^I(x_i)x_i$ ) of both the solvent and the solute is equal in the two phases. For example, for the acid solute:

$$a_i^{I, \alpha}(x_{\text{SOL, acid}}) = a_i^{I, \beta}(1 - x_{\text{SOL, w}}) \quad (4)$$

80 The SLE is solved from the solid-liquid equilibrium condition:

$$\log_{10}(x_{\text{SOL, i}}) = \frac{\mu_i^{*\circ, I}(x^\circ, T, P) - \mu_i^*(x_{\text{SOL, i}}) - \Delta G_{\text{fus}}(T)}{RT \ln(10)} \quad (5)$$

The free energy of fusion of the solute ( $\Delta G_{\text{fus}}(T)$ ) is calculated from the experimentally determined heat of fusion ( $\Delta H_{\text{fus}}$ ) and melting temperature ( $T_{\text{melt}}$ ) using the Schröder-van Laar equation (Prigogine and Defay, 1954):

$$\Delta G_{\text{fus}}(T) = \Delta H_{\text{fus}} \left(1 - \frac{T}{T_{\text{melt}}}\right) - \Delta C_{p, \text{fus}}(T_{\text{melt}} - T) + \Delta C_{p, \text{fus}} T \ln \frac{T_{\text{melt}}}{T} \quad (6)$$



85 Here the heat capacity of fusion ( $\Delta C_{p,\text{fus}}$ ) is estimated from the melting point and the heat of fusion:

$$\Delta C_{p,\text{fus}} = \frac{\Delta H_{\text{fus}}}{T_{\text{melt}}} \quad (7)$$

Table 1 shows the literature values of melting points and heats of fusion of the dicarboxylic acids of this study. Melting points and heats of fusion of the monocarboxylic acids are not used, since all of the monocarboxylic acids studied here are in liquid phase at 298.15 K.

### 90 2.3 Equilibrium constants

COSMOtherm estimates equilibrium constants of condensed-phase reactions from the free energy of the reaction ( $\Delta G_r$ ):

$$K_r = e^{-\frac{\Delta G_r}{RT}} \quad (8)$$

The reaction free energy is calculated from the free energies of the reactants ( $G_{\text{react}}$ ) and products ( $G_{\text{prod}}$ ) in the solution:

$$\Delta G_r = \Sigma G_{\text{prod}} - \Sigma G_{\text{react}} \quad (9)$$

95 The free energy of compound  $i$  (reactant or product) is the sum of the energy of the solvated compound ( $E_{\text{COSMO}}$ ), the averaged correction for the dielectric energy ( $dE$ ) and the chemical potential of the compound in the solution:

$$G_i = E_{\text{COSMO},i} + dE_i + \mu_i^*(x_i) \quad (10)$$

The solvent-independent effective equilibrium constant can be calculated by multiplying the reaction equilibrium constant with the ratio of the activity coefficients of the reactants ( $\gamma_{\text{react}}$ ) and products ( $\gamma_{\text{prod}}$ ):

$$100 \quad K_{\text{eff}} = \frac{\Pi \gamma_{\text{prod}}}{\Pi \gamma_{\text{react}}} K_r \quad (11)$$

Combining Eqs 8 and 11 gives:

$$\ln K_{\text{eff}} = \Sigma(\ln \gamma_{\text{prod}}) - \Sigma(\ln \gamma_{\text{react}}) - \frac{\Sigma G_{\text{prod}} - \Sigma G_{\text{react}}}{RT} \quad (12)$$

where the activity coefficients can be replaced by Eq. 1 and the free energies by Eq. 10:

$$\begin{aligned} RT \ln K_{\text{eff}} &= \Sigma(\mu_{\text{prod}}^*(x_{\text{prod}}) - \mu_{\text{prod}}^{*\circ,\text{I}}(x^\circ, T, P)) - \Sigma(\mu_{\text{react}}^*(x_{\text{react}}) - \mu_{\text{react}}^{*\circ,\text{I}}(x^\circ, T, P)) \\ &- \Sigma(E_{\text{COSMO},\text{prod}} + dE_{\text{prod}} + \mu_{\text{prod}}^*(x_{\text{prod}})) + \Sigma(E_{\text{COSMO},\text{react}} + dE_{\text{react}} + \mu_{\text{react}}^*(x_{\text{react}})) \\ &= -\Sigma(E_{\text{COSMO},\text{prod}} + dE_{\text{prod}} + \mu_{\text{prod}}^{*\circ,\text{I}}(x^\circ, T, P)) + \Sigma(E_{\text{COSMO},\text{react}} + dE_{\text{react}} + \mu_{\text{react}}^{*\circ,\text{I}}(x^\circ, T, P)) \\ 110 &= -\Sigma G_{\text{prod}}^{\circ,\text{I}} + \Sigma G_{\text{react}}^{\circ,\text{I}} \end{aligned} \quad (13)$$

This gives us the final equation for calculating the effective equilibrium constant of a reaction:

$$K_{\text{eff}} = e^{-\frac{\Delta G_r^{\circ,\text{I}}}{RT}}, \quad (14)$$

where the equilibrium constant depends on only the free energies of the pure compounds ( $G_i^{\circ,\text{I}}$ ). COSMOtherm calculates the equilibrium constants on a mole fraction basis, which leads to dimensionless equilibrium constants.



## 115 2.4 Concentration dependent reactions

In COSMO-RS, the surface of a molecule is divided into surface segments that represent the surface charges of the molecule. Interactions between molecules are described through the interaction between surface segments of the molecules. Concerted multiple contacts, such as carboxylic acid dimer formation, are not captured by COSMO-RS. COSMO*therm* is able to consider these hydrogen bonded clusters using the dimerization, aggregation, and reaction extension (COSMO-RS-DARE, Sachsenhauser et al. (2014)). We use the COSMO-RS-DARE method in our activity coefficient and solubility calculations. In our equilibrium constant calculations, the clusters in the system are included as the product of the clustering reactions. A clustering reaction between molecules A and B can be described by the equilibrium:



In acid–water systems, A and B can be either a carboxylic acid or a water molecule. In COSMO-RS-DARE, the product clusters are included in property calculations by using the screening charge density surface ( $\sigma$ -surface) of molecule A in the cluster, omitting the part of the  $\sigma$ -surface that is assigned to the molecule clustered with A (i.e., molecule B).

The formation of hydrogen bonds (in hydrates or dimers) is taken into account using the interaction energy of the two reacting compounds. The free energy between the interaction sites  $G(A, A \cdot B)$  is calculated using the energy difference between the monomer compound (A) and the same compound in the cluster (A·B),  $\Delta^0$ :

$$130 \quad G(A, A \cdot B) = -2\Delta^0 + c_H - c_S T, \quad (15)$$

where fitting parameters  $c_H$  and  $c_S$  are the enthalpic and entropic contributions to the interaction energy, respectively. Because we are only calculating condensed-phase properties at one temperature ( $T = 298.15$  K), we are only fitting the enthalpic parameter  $c_H$ , keeping the entropic parameter  $c_S$  at zero.

COSMO-RS-DARE was originally developed for systems containing carboxylic acids in non-polar solvents (Sachsenhauser et al., 2014). In a carboxylic acid–water system, both the carboxylic acid and water are able to form strongly bound clusters. In addition, hydrated acids can be formed. We are thus including the interactions of the clustering reactions for both A and B, even when  $A = B$ .

## 2.5 Input file generation

The .cosmo files of water and the monocarboxylic acids with a low number of conformers ( $<10$ ) are taken from the COSMO*base* (COSMO*base*, 2011) database. For the dicarboxylic acids, acid and water dimers, and the hydrates of pimelic ( $m=7$ ) and suberic ( $m=8$ ) acids, we use the following systematic conformer search approach detailed by Kurtén et al. (2018) as it has been shown to give more consistent results than other conformer sampling approaches. The conformers are found using the systematic conformer search in the Spartan program (Wavefunction Inc., 2014, 2016). The conformer set is then used as input to the COSMO*conf* program (COSMO*conf*, 2013) (using the TURBOMOLE program (TURBOMOLE, 2010)), which runs initial single-point COSMO calculations at the BP/def-SV(P) level of theory to compare the chemical potentials of the



conformers and remove similar structures. Initial geometry optimizations are calculated at the BP/def-SV(P) level of theory, duplicate structures are removed by comparing the new geometries and chemical potentials. Final geometries are optimized at the BP/def-TZVP level of theory and after a second duplicate removal step, final single-point energies are calculated at the BP/def2-TZVPD-FINE level of theory.

150 For acid dimers, we use the lowest gas-phase energy structures found by Elm et al. (2019) as a starting structure for systematic conformer search. For hydrated monocarboxylic acids and smaller dicarboxylic acids ( $m \leq 6$ ), the clusters are built by adding a water molecule to each conformer of the free acids. For monocarboxylic acids, the water molecule is placed on the carboxylic acid group forming two intermolecular hydrogen bonds between the molecules. For dicarboxylic acids, a water molecule is added to either end of the acid, forming two hydrate conformers from a single acid conformer. For the dicarboxylic  
155 acid conformers with the two acid groups close to each other, additional conformers are created for cases where the water molecule is interacting with both acid groups. Due to the large number of conformers of non-hydrated pimelic ( $m = 7$ ) and suberic ( $m = 8$ ) acid (75 and 132, respectively), the monohydrate conformers of those two acids are sampled separately using Spartan.

We use only clusters of two molecules in our calculations. In carboxylic acid dimers, the hydrogen bond donors and acceptors  
160 are saturated, which means that carboxylic acids are unlikely to form larger clusters than dimers (Vawdrey et al., 2004; Elm et al., 2014, 2019). Computational studies (Aloisio et al., 2002; Weber et al., 2012; Kildgaard et al., 2018) have shown that, in the gas phase, the energetically most favorable dihydrate is formed by two water molecules attaching to the same carboxylic acid group. Therefore, adding a second water molecule to the cluster does not significantly change the probability distribution of the screening charge density ( $\sigma$ -profile) of the acid in the cluster compared to the acid in a monohydrate or dimer.

165 Conformers containing no intramolecular hydrogen bonds (Kurtén et al., 2018; Hyttinen and Prisle, 2020) are used in the COSMO-RS solubility and activity coefficient calculations. Due to the intermolecular hydrogen bonding in the hydrate and dimer clusters, all conformers (up to 40 conformers) of monomers and clusters are used in the effective equilibrium constant calculations. In COSMO-RS-DARE calculations, we use all conformers of the monomers and only the lowest solvated energy conformers of the clusters.

## 170 3 Results and discussion

### 3.1 Effective equilibrium constants of clustering reactions

We estimated the effective equilibrium constants of the different clustering reactions (i.e., hydration and dimerization) of the binary acid–water systems. A comparison between the hydration and acid dimerization equilibrium constants in the aqueous phase is given in Fig. 1 and Table 2. The equilibrium constants for both the dimerization and hydration reactions are similar  
175 between all of the monocarboxylic acids, and are not labeled in Fig. 1. For the dicarboxylic acid, we can see larger variation in both the hydration and dimerization reactions. Note that the COSMO-RS-DARE method is not used in the effective equilibrium constant calculations because the clusters are already included in the calculation as products.



For all of the acids, the effective equilibrium constant of dimerization is higher than that of the hydrate formation of the corresponding acid, meaning that acid dimer formation is energetically more favorable than hydrate formation. However, in dilute conditions, water is more abundant, shifting the equilibrium from acid dimerization to hydration. The dimerization:hydration ratio is the lowest for oxalic ( $m = 2$ ) and malonic ( $m = 3$ ) acids, while monocarboxylic acids and succinic acid ( $m = 4$ ) have similar (intermediate) ratios, and the larger dicarboxylic acids ( $m = 5-8$ ) have higher ratios. This means that, in dilute solutions, oxalic, malonic and succinic acid will most likely interact with water instead of other acid molecules.

Vawdrey et al. (2004) calculated the dimerization enthalpies (at the B3LYP/6-31++G(2d,p) level of theory) of monocarboxylic acids ( $n = 2-6$ ) and found a notable even-odd variation (dimerization of the acids with odd number of carbon atoms is more favorable than of the even carbon number acids). The same is seen here in the condensed phase, where the effective equilibrium constant of butanoic and hexanoic acids are lower than of propanoic and pentanoic acids, respectively. Otherwise there is a slightly increasing trend in the effective equilibrium constants with the increasing number of carbon atoms in the monocarboxylic acids. For larger dicarboxylic acids ( $m \geq 4$ ), Elm et al. (2019) found an even-odd alternation in the dimer:monomer ratio in the gas phase, calculated at the DLPNO-CCSD(T)/aug-cc-pVTZ// $\omega$ B97X-D/6-31++G(d,p) level of theory. We observe a similar increasing effective equilibrium constant with the increasing carbon chain length in the smaller dicarboxylic acids ( $m = 2-5$ ) and an even-odd alternation in the larger dicarboxylic acids ( $m = 4-8$ ) in any condensed phase.

## 3.2 Activity coefficients

### 3.2.1 Monocarboxylic acids

We calculated the activity coefficient of the carboxylic acids and water in the binary acid-water mixtures using the COSMO-RS-DARE method. Hansen et al. (1955) derived the activity coefficients of acetic ( $n = 2$ ), propanoic ( $n = 3$ ) and butanoic ( $n = 4$ ) acids in mixtures with water from partial pressure measurements. In addition, they determined activity coefficients of water in aqueous acetic, propanoic, butanoic, pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acid mixtures. We used these experimental activity coefficients to fit the enthalpic parameters ( $c_H$ ) for each of the acids in the COSMO-RS-DARE calculations. Figure 2 shows a comparison between the estimated and experimentally determined activity coefficients of these monocarboxylic acids, and formic acid ( $n = 1$ ), for which no experimental activity coefficient data is available.

The reactions included in the calculations are water dimer ( $\text{H}_2\text{O} \cdot \text{H}_2\text{O}$ ) and acid hydrate ( $\text{RCOOH} \cdot \text{H}_2\text{O}$ ) formation. A comparison between COSMO-RS activity coefficients and the COSMO-RS-DARE method with different clusters included in the calculation, are shown for acetic acid in Fig. S1 of the Supplement. For acetic acid, we found the best fit between experimental and estimated activity coefficients using  $c_H = 0 \text{ kcal mol}^{-1}$  for both the water dimerization and acid hydration reactions. Decreasing the  $c_H$  of either clustering reaction leads to stronger deviation from ideality, which in our case leads to a worse fit for water activity coefficient, and positive parameter values cannot be used to lower the interaction enthalpy. The effective equilibrium constant for water dimer formation ( $5.71 \times 10^5$ ) is below that of acetic acid hydration ( $4.36 \times 10^6$ ), which explains why the fitting parameter of the water dimer hydration should be higher (or equal, since positive values are not possible) than the parameter for acid hydrate formation.



For the other monocarboxylic acids, we used the same  $c_H$  value for water dimerization that was found for the acetic acid–water system, and fitted the  $c_H$  of acid hydrate formation to the experimental activity coefficients of water and the acids in the corresponding acid–water systems (Hansen et al., 1955). The enthalpic parameter values of acid hydration used to estimate the activity coefficients shown in Fig. 2 are 0.0, 0.0, -2.5, -3.5, -2.2 and -2.0 kcal mol<sup>-1</sup>, for formic ( $n = 1$ ), acetic ( $n = 2$ ),  
215 propanoic ( $n = 3$ ), butanoic ( $n = 4$ ), pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acid, respectively. For formic acid, we used the same  $c_H$  parameter as for acetic acid due to lack of experimental activity coefficients. Sachsenhauser et al. (2014) used the COSMO-RS-DARE method for binary systems containing either acetic ( $n = 2$ ) or propanoic ( $n = 3$ ) acid and a non-polar organic solvent. Their calculations show that the dimerization parameter (equivalent to  $c_H$  in our calculations) is higher for propanoic acid than for acetic acid. This is opposite to what we observed for the hydration parameters, where  $c_H$  was found to  
220 be higher for acetic acid than for propanoic acid. This indicates that the fitting parameters of one clustering reaction cannot be used to estimate the corresponding fitting parameters of another clustering reactions of the same compound.

The increasing length of the acid carbon backbone leads to larger deviation from ideality ( $\gamma = 1$ ) for both the acid and water. In convention I, this means the acid and water activity coefficient values are higher in mixtures containing the longer acids than the shorter acids. We observe that COSMO-RS-DARE estimated activity coefficients agree well with the experiments once the  
225  $c_H$  parameter is fitted. However, when the hydrate and water dimer reactions are included, COSMO-RS-DARE is not able to predict realistic activity coefficients for water at high mole fractions of the acids. This is likely due to the low concentration of water in the binary solution, leading to errors in the description of the interactions between water molecules. The COSMO-RS-DARE estimates agree well with the experiments at least up to 0.9 mole fraction of the monocarboxylic acids. For higher acid mole fractions, COSMO-RS-DARE predicts several orders of magnitude higher activity coefficients than what was seen  
230 in experiments.

### 3.2.2 Dicarboxylic acids

Multiple experimental studies have reported activity (coefficient) data for either dicarboxylic acids or water in binary acid–water systems (Davies and Thomas, 1956; Soonsin et al., 2010; Maffia and Meirelles, 2001; Peng et al., 2001; Choi and Chan, 2002; Wise et al., 2003). However, due to the low solubility of the dicarboxylic acids in water (see Fig. 5 for the experimental  
235 solubilities), fitting of the  $c_H$  parameter is not feasible for all of the acids. The concentration of the dicarboxylic acid at the reference state, with respect to which the experimental activity coefficients of dicarboxylic acids are determined, is either 0.5 or 1 mol kg<sup>-1</sup>, and the activity coefficients have only been measured in states close to the reference state (Davies and Thomas, 1956). In addition, Soonsin et al. (2010) reported activity coefficients for malonic, succinic and glutaric acid in convention I at various temperatures in the particle phase. However, Hyttinen and Prisle (2020) have previously found a good agreement  
240 between experimental particle phase activity coefficients and COSMO<sup>therm</sup> estimates of citric, tartaric, malic and malonic acid.

We tested the effect of including different clusters in the activity coefficient calculation of malonic ( $m = 3$ ) and glutaric ( $m = 5$ ) acid. Both of these acids have relatively high aqueous solubilities ( $x_{\text{SOL,acid}} \approx 0.5$ ) and multiple experimental studies that have measured water activities in binary acid–water solutions. For the other dicarboxylic acids in this study, we were





245 only able to find acid activity coefficients for succinic acid ( $m = 4$ ), at 298.15 K and the values were only measured in a  
small acid mole fraction range ( $0.007 < x_{\text{acid}} < 0.01$ ) with respect to  $x_{\text{acid}} = 0.009$ . A comparison between the experimentally  
determined activity coefficients and COSMO $therm$  estimates of malonic and glutaric acid are shown in Figs S2 and S3 of the  
Supplement, respectively. The agreement between experimental and estimated activity coefficients is not as good as for the  
monocarboxylic acid but the agreement is reasonable for both acids using the COSMO-RS method with conformers containing  
250 no intramolecular hydrogen bonds. For glutaric acid, using COSMO-RS-DARE with acid dimerization gives similar results  
as COSMO-RS (see Fig. S3 of the Supplement). For succinic acid, there is a large variation in the water activity coefficients  
between the different experiments and the COSMO-RS and COSMO-RS-DARE methods both agree reasonably well with the  
activity coefficients measured by Maffia and Meirelles (2001) and Peng et al. (2001). However, succinic acid activity coefficient  
estimates agree better with the particle phase activity coefficients of Soonsin et al. (2010) using COSMO-RS as opposed to  
255 COSMO-RS-DARE (Fig. S4 of the Supplement).

From Fig. 1, we can see that the larger dicarboxylic ( $m > 4$ ) acids have higher dimerization:hydration ratios than the  
monocarboxylic acids. Since glutaric acid ( $m = 5$ ) is better described by including the acid dimerization reaction to the system  
than by including water dimerization and/or acid hydration, we used the acid dimerization ( $c_H = 0 \text{ kcal mol}^{-1}$ ) in the acid  
description of the other larger dicarboxylic acids ( $m = 6-8$ ). With relatively high effective hydration equilibrium constant, the  
260 hydrated oxalic acid is likely the most appropriate clustering reaction to be added in the COSMO-RS-DARE calculation of  
oxalic acid. However, without experimental activity coefficient data of oxalic acid–water systems, we are not able to fit the  $c_H$   
parameter needed for reliable estimates. In addition, COSMO-RS is likely sufficient in describing the acid–water interactions  
of oxalic acid, similar to the other dicarboxylic acids.

The estimated activity coefficients of the dicarboxylic acids are shown in Fig. 3. We can see that the activity coefficients of  
265 the smaller dicarboxylic acids are lower than for the larger dicarboxylic acids. Water activity coefficients in all of the binary  
acid–water systems are close to 1 with  $x_{\text{acid}} < 0.1$ . For malonic ( $m = 3$ ) and glutaric ( $m = 5$ ) acid (with higher aqueous  
solubility), the activity coefficients deviate from ideality ( $\gamma = 1$ ) at higher acid mole fractions. There is no significant difference  
in the COSMO-RS and COSMO-RS-DARE activity coefficients of the dicarboxylic acid–water systems, the largest difference  
is seen in suberic and pimelic acids.

### 270 3.3 Aqueous solubility

We estimated the aqueous solubility of the carboxylic acids (monocarboxylic acids with  $n = 1-6$  and dicarboxylic acids  
with  $m = 5-8$ ) using the COSMO-RS-DARE method. Since activity coefficients are used in the equilibrium conditions of  
the LLE and SLE calculations, we used the same  $c_H$  parameters that were fitted in the activity coefficient calculations, to  
determine whether the same parameter value can be used in LLE and SLE calculations. As a comparison, we computed the  
275 same solubilities using COSMO-RS and only conformers containing no intramolecular hydrogen bonds (dicarboxylic acids  
only) (Kurtén et al., 2018; Hyttinen and Prisle, 2020). Based on previous COSMO $therm$  calculations, Michailoudi et al. (2020)  
found a good agreement with experimental aqueous solubilities of fatty acids with even number of carbon atoms ( $n = 2-12$ ). In  
addition, Schröder et al. (2010) compared COSMO $therm$  estimated aqueous solubilities of dicarboxylic acids ( $m = 2-9$ ) with



experiments and found that COSMO $therm$  is able to predict the temperature dependence quite well. A comparison between  
280 experimentally determined aqueous solubilities and the COSMO $therm$  estimates of mono- and dicarboxylic acids are shown  
in Figs 4 and 5, respectively.

We see that using COSMO-RS-DARE, COSMO $therm$  is able to predict the miscibility of the smaller monocarboxylic acids  
( $n = 1-4$ ) but the experimental solubilities of pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acids are overestimated to a greater  
degree than when using COSMO-RS. On the other hand, COSMO-RS underestimates the experimental solubility of butanoic  
285 acid by a factor of 18, while COSMO-RS-DARE overestimates the experimental solubilities (upper limit) of pentanoic and  
hexanoic acids only by factors of 3.4 and 4.1, respectively.

Different experimental enthalpy of fusion and melting temperature values have been reported for the dicarboxylic acids. We  
calculated the lower and upper limit free energies of fusion by combining the different experimental values and estimated the  
solubilities using the two different free energy of fusion values. Using COSMO-RS, the higher  $\Delta G_{\text{fus}}$  estimate gives a lower  
290 aqueous solubility. COSMO $therm$  overestimates the experimental aqueous solubility of oxalic acid ( $m = 2$ ) by a factor of 35  
when the experimental melting point and heat of fusion by Booth et al. (2010) are used to estimate the free energy of fusion of  
oxalic acid. Cornils and Lappe (2000) and Omar and Ulrich (2006) also measured the melting point of oxalic acid and found  
temperatures almost 100 K higher than Booth et al. (2010). In addition, the heat of fusion measured by Omar and Ulrich (2006)  
is more than one order of magnitude higher than what was measured by Booth et al. (2010). This explains the differences in the  
295 estimated aqueous solubility of oxalic acid depending on whether the lower or upper limit estimate of the free energy of fusion  
is used in the COSMO $therm$  calculation. For the other acids, the variability in the measured melting points is much smaller (at  
most 11 K). For the SLE calculation using the COSMO-RS-DARE method, we used the upper limit  $\Delta G_{\text{fus}}$  estimate.

Adding the acid dimerization in the system ( $m = 5-8$ ) has a smaller impact on the solubility estimate than the addition of the  
hydrated acid and water dimer ( $n = 4-6$ ). The COSMO-RS solubility estimates of most of the dicarboxylic acids ( $m = 3-7$ )  
300 are within the experimentally determined solubility range. Using all lowest energy conformers (up to 40 conformers), instead  
of only conformers containing no intramolecular hydrogen bonds, lowers the solubility estimates of all acids by a factor of 1.2  
on average. The same effect of including conformers containing intramolecular hydrogen bonds has been previously seen in  
aqueous solubilities of citric, tartaric, malic and maleic acid, and multifunctional organosulfates (Hyttinen and Prisle, 2020).

#### 4 Conclusions

305 Using COSMO-RS-DARE, we were able to significantly improve the activity coefficient estimates of water in the binary  
monocarboxylic acid–water system. The COSMO-RS estimates of monocarboxylic acid activity coefficient in aqueous solu-  
tions agree with the experiments quite well, and are further improved by COSMO-RS-DARE. In addition, COSMO-RS-DARE  
is able to predict the miscibility of butanoic acid in water, while COSMO-RS predicts a finite solubility. However, in aqueous  
solubility calculations of pentanoic and hexanoic acid, COSMO-RS leads to a better agreement between the experiments and  
310 COSMO $therm$  estimates than COSMO-RS-DARE.



COSMO $therm$  is able to reproduce the same even-odd behavior of the dicarboxylic acid properties that have previously been seen experimentally in vapor pressures (Bilde et al., 2003) and solubilities (Zhang et al., 2013), and computationally in gas-phase dimer formation (Elm et al., 2019). The calculated even-odd behavior observed here in aqueous solubilities is likely partially due to the even-odd variation of the melting points and heats of fusion. However, even-odd variation is seen in the effective equilibrium constants of dimerization of the larger dicarboxylic acids, which do not rely on experimental properties.

Cysewski (2019) found that COSMO-RS overestimates solubilities of ethenzamide in solvents such as ethyl acetate and methanol by up to a factor of 6 and a better estimates was found by fitting the enthalpic parameter in COSMO-RS-DARE to experiments. For the mono- and dicarboxylic acids, COSMO-RS over-/underestimates the experimental solubility by up to a factor of 3, with the exception of the solubility of butanoic ( $n = 4$ ), which is underestimated by a factor of 18. For each of the larger dicarboxylic acids ( $m = 5-8$ ), the solubility estimate changes by less than a factor of 1.6 between the COSMO-RS and COSMO-RS-DARE. Overall, the change in the COSMO-RS-DARE solubility compared to COSMO-RS solubility of mono- and dicarboxylic acids is not as large as was seen by Cysewski (2019), using the enthalpic parameter fitted to experimental activity coefficients.

Mono- and dicarboxylic acids are very common in the atmosphere and often used as model compounds for oxygenated functionalities in a range of applications from vapor pressure, condensation-evaporation, cloud condensation nuclei activity and hygroscopicity, but also aerosol phase and heterogeneous reactivity (Prenni et al., 2001; McNeill et al., 2008; Schwier et al., 2012; Rossignol et al., 2016). Solubilities and activity coefficients of these SOA constituents are needed to accurately predict their activities, to determine central properties such as composition, phase state, and chemical reactivity. Accurate computational tools are critical to provide this information for systems where experimental data are not readily accessible in literature or by experimental design.

*Data availability.* The research data have been deposited in a reliable public data repository (the CERN Zenodo service) and can be accessed at <https://doi.org/10.5281/zenodo.3842593> (Hytinen et al., 2020).

*Author contributions.* NH ran the COSMO $therm$  calculations and analyzed the results. NLP conceived, planned, supervised and secured funding for the project. NH, RH and JE generated input files for the COSMO $therm$  calculations. NH wrote the paper with contributions from all co-authors.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We thank CSC - IT Center for Science, Finland, for computational resources.



*Financial support.* This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Project SURFACE (Grant Agreement No. 717022). Nønne L. Prisle, Noora Hyttinen, Reyhaneh Heshmatnezhad and Theo Kurtén also gratefully acknowledge the financial contribution from the Academy of Finland, including Grant No. 308238, 314175 and 315600. Jonas Elm is grateful for financial support from the Swedish Research Council Formas project number 2018-01745-COBACCA.



## References

- Aloisio, S., Hintze, P. E., and Vaida, V.: The hydration of formic acid, *J. Phys. Chem. A*, 106, 363–370, <https://doi.org/10.1021/jp012190l>, 2002.
- 345 Apelblat, A. and Manzurola, E.: Solubility of oxalic, malonic, succinic, adipic, maleic, malic, citric, and tartaric acids in water from 278.15 to 338.15 K, *J. Chem. Thermodyn.*, 19, 317–320, [https://doi.org/10.1016/0021-9614\(87\)90139-X](https://doi.org/10.1016/0021-9614(87)90139-X), 1987.
- Apelblat, A. and Manzurola, E.: Solubility of ascorbic, 2-furancarboxylic, glutaric, pimelic, salicylic, and o-phthalic acids in water from 279.15 to 342.15 K, and apparent molar volumes of ascorbic, glutaric, and pimelic acids in water at 298.15 K, *J. Chem. Thermodyn.*, 21, 1005–1008, [https://doi.org/10.1016/0021-9614\(89\)90161-4](https://doi.org/10.1016/0021-9614(89)90161-4), 1989.
- 350 Apelblat, A. and Manzurola, E.: Solubility of suberic, azelaic, levulinic, glycolic, and diglycolic acids in water from 278.25 K to 361.35 K, *J. Chem. Thermodyn.*, 22, 289–292, [https://doi.org/10.1016/0021-9614\(90\)90201-Z](https://doi.org/10.1016/0021-9614(90)90201-Z), 1990.
- Bilde, M., Svenningsson, B., Mønster, J., and Rosenørn, T.: Even-odd alternation of evaporation rates and vapor pressures of C3–C9 dicarboxylic acid aerosols, *Environ. Sci. Technol.*, 37, 1371–1378, <https://doi.org/10.1021/es0201810>, 2003.
- 355 Booth, A. M., Barley, M. H., Topping, D. O., McFiggans, G., Garforth, A., and Percival, C. J.: Solid state and sub-cooled liquid vapour pressures of substituted dicarboxylic acids using Knudsen Effusion Mass Spectrometry (KEMS) and Differential Scanning Calorimetry, *Atmos. Chem. Phys.*, 10, 4879–4892, <https://doi.org/10.5194/acp-10-4879-2010>, 2010.
- Brooks, S. D., Wise, M. E., Cushing, M., and Tolbert, M. A.: Deliquescence behavior of organic/ammonium sulfate aerosol, *Geophys. Res. Lett.*, 29, 23–1, <https://doi.org/10.1029/2002GL014733>, 2002.
- 360 Choi, M. Y. and Chan, C. K.: Continuous measurements of the water activities of aqueous droplets of water-soluble organic compounds, *J. Phys. Chem. A*, 106, 4566–4572, <https://doi.org/10.1021/jp013875o>, 2002.
- Cornils, B. and Lappe, P.: Dicarboxylic acids, aliphatic, *Ullmann's Encyclopedia of Industrial Chemistry*, pp. 1–18, 2000.
- COSMObase: version 17.01, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2011.
- COSMOconf: version 4.3, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2013.
- 365 COSMOtherm: version C3.0, Release 19, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2019.
- Cysewski, P.: Prediction of ethenzamide solubility in organic solvents by explicit inclusions of intermolecular interactions within the framework of COSMO-RS-DARE, *J. Mol. Liq.*, p. 111163, <https://doi.org/10.1016/j.molliq.2019.111163>, 2019.
- Davies, M. and Thomas, D. K.: Isopiestic Studies of Aqueous. Dicarboxylic Acid Solutions, *J. Phys. Chem.*, 60, 41–44, <https://doi.org/10.1021/j150535a011>, 1956.
- 370 Eckert, F. and Klamt, A.: Fast solvent screening via quantum chemistry: COSMO-RS approach, *AIChE J.*, 48, 369–385, <https://doi.org/10.1002/aic.690480220>, 2002.
- Elm, J., Kurtén, T., Bilde, M., and Mikkelsen, K. V.: Molecular interaction of pinic acid with sulfuric acid: Exploring the thermodynamic landscape of cluster growth, *J. Phys. Chem. A*, 118, 7892–7900, <https://doi.org/10.1021/jp503736s>, 2014.
- Elm, J., Hyttinen, N., Lin, J. J., Kurtén, T., and Prisle, N. L.: Strong Even/Odd Pattern in the Computed Gas-Phase Stability of Dicarboxylic Acid Dimers: Implications for Condensation Thermodynamics, *J. Phys. Chem. A*, 123, 9594–9599, <https://doi.org/10.1021/acs.jpca.9b08020>, 2019.
- Fisseha, R., Dommen, J., Gaeggeler, K., Weingartner, E., Samburova, V., Kalberer, M., and Baltensperger, U.: Online gas and aerosol measurement of water soluble carboxylic acids in Zurich, *J. Geophys. Res.: Atmos.*, 111, <https://doi.org/10.1029/2005JD006782>, 2006.



- 380 Guo, H., Zhou, J., Wang, L., Zhou, Y., Yuan, J., Zhao, R., et al.: Seasonal variations and sources of carboxylic acids in PM<sub>2.5</sub> in Wuhan, China, *Aerosol Air Qual. Res.*, 15, 517–528, <https://doi.org/10.4209/aaqr.2014.02.0040>, 2014.
- Hansen, R. S., Miller, F. A., and Christian, S. D.: Activity Coefficients of Components in the Systems Water–Acetic Acid, Water–Propionic Acid and Water–n-Butyric Acid at 25°, *J. Phys. Chem.*, 59, 391–395, <https://doi.org/10.1021/j150527a003>, 1955.
- Hyder, M., Genberg, J., Sandahl, M., Swietlicki, E., and Jönsson, J. Å.: Yearly trend of dicarboxylic acids in organic aerosols from south of Sweden and source attribution, *Atmos. Environ.*, 57, 197–204, <https://doi.org/10.1016/j.atmosenv.2012.04.027>, 2012.
- 385 Hyttinen, N. and Prisle, N. L.: Improving solubility and activity estimates of multifunctional atmospheric organics by selecting conformers in COSMO<sub>therm</sub>, *J. Phys. Chem. A*, <https://doi.org/10.1021/acs.jpca.0c04285>, 2020.
- Hyttinen, N., Heshmatnezhad, R., Elm, J., Kurtén, T., and Prisle, N. L.: Supplementary data for the manuscript "Technical note: Estimating aqueous solubilities and activity coefficients of mono- and  $\alpha,\omega$ -dicarboxylic acids using COSMO-RS-DARE" (Version 1) [Data set], Zenodo, <https://doi.org/10.5281/zenodo.3842593>, 2020.
- 390 Jones, E. R. and Bury, C. R.: LXXVII. The freezing-points of concentrated solutions.–Part II. Solutions of formic, acetic, propionic, and butyric acids, *London, Edinburgh Dublin Philos. Mag. J. Sci.*, 4, 841–848, <https://doi.org/10.1080/14786441008564388>, 1927.
- Jung, J., Tsatsral, B., Kim, Y. J., and Kawamura, K.: Organic and inorganic aerosol compositions in Ulaanbaatar, Mongolia, during the cold winter of 2007 to 2008: dicarboxylic acids, ketocarboxylic acids, and  $\alpha$ -dicarbonyls, *J. Geophys. Res.: Atmos.*, 115, <https://doi.org/10.1029/2010JD014339>, 2010.
- 395 Kawamura, K. and Sakaguchi, F.: Molecular distributions of water soluble dicarboxylic acids in marine aerosols over the Pacific Ocean including tropics, *J. Geophys. Res.: Atmos.*, 104, 3501–3509, <https://doi.org/10.1029/1998JD100041>, 1999.
- Kawamura, K., Seméré, R., Imai, Y., Fujii, Y., and Hayashi, M.: Water soluble dicarboxylic acids and related compounds in Antarctic aerosols, *J. Geophys. Res.: Atmos.*, 101, 18 721–18 728, <https://doi.org/10.1029/96JD01541>, 1996.
- Kawamura, K., Steinberg, S., and Kaplan, I. R.: Homologous series of C1–C10 monocarboxylic acids and C1–C6 carbonyls in Los Angeles  
400 air and motor vehicle exhausts, *Atmos. Environ.*, 34, 4175–4191, [https://doi.org/10.1016/S1352-2310\(00\)00212-0](https://doi.org/10.1016/S1352-2310(00)00212-0), 2000.
- Kildgaard, J. V., Mikkelsen, K. V., Bilde, M., and Elm, J.: Hydration of Atmospheric Molecular Clusters II: Organic Acid–Water Clusters, *J. Phys. Chem. A*, 122, 8549–8556, <https://doi.org/10.1021/acs.jpca.8b07713>, 2018.
- Klamt, A.: Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.*, 99, 2224–2235, <https://doi.org/10.1021/j100007a062>, 1995.
- 405 Klamt, A., Jonas, V., Bürger, T., and Lohrenz, J. C. W.: Refinement and parametrization of COSMO-RS, *J. Phys. Chem. A*, 102, 5074–5085, <https://doi.org/10.1021/jp980017s>, 1998.
- Kurtén, T., Hyttinen, N., D’Ambro, E. L., Thornton, J., and Prisle, N. L.: Estimating the saturation vapor pressures of isoprene oxidation products C<sub>5</sub>H<sub>12</sub>O<sub>6</sub> and C<sub>5</sub>H<sub>10</sub>O<sub>6</sub> using COSMO-RS, *Atmos. Chem. Phys.*, 18, 17 589–17 600, <https://doi.org/10.5194/acp-18-17589-2018>, 2018.
- 410 Levine, I. N.: *Physical Chemistry*, McGraw-Hill, New York, USA, 6th edn., 2009.
- Maffia, M. C. and Meirelles, A. J. A.: Water activity and pH in aqueous polycarboxylic acid systems, *J. Chem. Eng. Data*, 46, 582–587, <https://doi.org/10.1021/je0002890>, 2001.
- McNeill, V. F., Yatavelli, R. L. N., Thornton, J. A., Stipe, C. B., and Landgrebe, O.: Heterogeneous OH oxidation of palmitic acid in single component and internally mixed aerosol particles: vaporization and the role of particle phase, *Atmos. Chem. Phys.*, 8,  
415 <https://doi.org/10.5194/acp-8-5465-2008>, 2008.



- Michailoudi, G., Hyttinen, N., Kurtén, T., and Prisle, N. L.: Solubility and Activity Coefficients of Atmospheric Surfactants in Aqueous Solution Evaluated using COSMO<sub>therm</sub>, *J. Phys. Chem. A*, 124, 430–443, <https://doi.org/10.1021/acs.jpca.9b09780>, 2020.
- Mochida, M., Kawamura, K., Umemoto, N., Kobayashi, M., Matsunaga, S., Lim, H.-J., Turpin, B. J., Bates, T. S., and Simoneit, B. R.: Spatial distributions of oxygenated organic compounds (dicarboxylic acids, fatty acids, and levoglucosan) in marine aerosols over the western Pacific and off the coast of East Asia: Continental outflow of organic aerosols during the ACE-Asia campaign, *J. Geophys. Res.: Atmos.*, 108, <https://doi.org/10.1029/2002JD003249>, 2003.
- Omar, W. and Ulrich, J.: Solid liquid equilibrium, metastable zone, and nucleation parameters of the oxalic acid- water system, *Cryst. Growth Des.*, 6, 1927–1930, <https://doi.org/10.1021/cg060112n>, 2006.
- O’Neil, M. J.: *The Merck index: an encyclopedia of chemicals, drugs, and biologicals*, RSC Publishing, 2013.
- Peng, C., Chan, M. N., and Chan, C. K.: The hygroscopic properties of dicarboxylic and multifunctional acids: Measurements and UNIFAC predictions, *Environ. Sci. Technol.*, 35, 4495–4501, <https://doi.org/10.1021/es0107531>, 2001.
- Prenni, A. J., DeMott, P. J., Kreidenweis, S. M., Sherman, D. E., Russell, L. M., and Ming, Y.: The effects of low molecular weight dicarboxylic acids on cloud formation, *J. Phys. Chem. A*, 105, 11 240–11 248, <https://doi.org/10.1021/jp012427d>, 2001.
- Prigogine, I. and Defay, R.: *Chemical Thermodynamics*, Longmans Green, 1st edition edn., 1954.
- Pye, H. O. T., Nenes, A., Alexander, B., Ault, A. P., Barth, M. C., Clegg, S. L., Collett Jr., J. L., Fahey, K. M., Hennigan, C. J., Herrmann, H., Kanakidou, M., Kelly, J. T., Ku, I.-T., McNeill, V. F., Riemer, N., Schaefer, T., Shi, G., Tilgner, A., Walker, J. T., Wang, T., Weber, R., Xing, J., Zaveri, R. A., and Zuend, A.: The acidity of atmospheric particles and clouds, *Atmos. Chem. Phys.*, 20, 4809–4888, <https://doi.org/10.5194/acp-20-4809-2020>, 2020.
- Romero, C. M. and Suárez, F.: Effect of temperature on the solubility of short-chain carboxylic acids in water, *J. Solution Chem.*, 38, 315–320, <https://doi.org/10.1007/s10953-009-9375-6>, 2009.
- Rossignol, S., Tinel, L., Bianco, A., Passananti, M., Brigante, M., Donaldson, D. J., and George, C.: Atmospheric photochemistry at a fatty acid-coated air-water interface, *Science*, 353, 699–702, <https://doi.org/10.1126/science.aaf3617>, 2016.
- Roux, M. V., Temprado, M., and Chickos, J. S.: Vaporization, fusion and sublimation enthalpies of the dicarboxylic acids from C<sub>4</sub> to C<sub>14</sub> and C<sub>16</sub>, *J. Chem. Thermodyn.*, 37, 941–953, <https://doi.org/10.1016/j.jct.2004.12.011>, 2005.
- Sachsenhauser, T., Rehfeldt, S., Klamt, A., Eckert, F., and Klein, H.: Consideration of dimerization for property prediction with COSMO-RS-DARE, *Fluid Phase Equilib.*, 382, 89–99, <https://doi.org/10.1016/j.fluid.2014.08.030>, 2014.
- Saxena, P. and Hildemann, L. M.: Water-soluble organics in atmospheric particles: A critical review of the literature and application of thermodynamics to identify candidate compounds, *J. Atmos. Chem.*, 24, 57–109, <https://doi.org/10.1007/BF00053823>, 1996.
- Schröder, B., Santos, L. M. N. B. F., Marrucho, I. M., and Coutinho, J. A. P.: Prediction of aqueous solubilities of solid carboxylic acids with COSMO-RS, *Fluid Phase Equilib.*, 289, 140–147, <https://doi.org/10.1016/j.fluid.2009.11.018>, 2010.
- Schwier, A., Mitroo, D., and McNeill, V. F.: Surface tension depression by low-solubility organic material in aqueous aerosol mimics, *Atmos. Environ.*, 54, 490–495, <https://doi.org/10.1016/j.atmosenv.2012.02.032>, 2012.
- Song, M., Marcolli, C., Krieger, U. K., Zuend, A., and Peter, T.: Liquid-liquid phase separation and morphology of internally mixed dicarboxylic acids/ammonium sulfate/water particles, *Atmos. Chem. Phys.*, 12, 2691–2712, <https://doi.org/10.5194/acp-12-2691-2012>, 2012.
- Soonsin, V., Zardini, A. A., Marcolli, C., Zuend, A., and Krieger, U. K.: The vapor pressures and activities of dicarboxylic acids reconsidered: the impact of the physical state of the aerosol, *Atmos. Chem. Phys.*, 10, 11 753–11 767, <https://doi.org/10.5194/acp-10-11753-2010>, 2010.
- Tsai, Y. I. and Kuo, S.-C.: Contributions of low molecular weight carboxylic acids to aerosols and wet deposition in a natural subtropical broad-leaved forest environment, *Atmos. Environ.*, 81, 270–279, <https://doi.org/10.1016/j.atmosenv.2013.08.061>, 2013.



- TURBOMOLE: Version 7.11, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, TURBOMOLE GmbH,  
455 2010.
- Vawdrey, A. C., Oscarson, J. L., Rowley, R. L., and Wilding, W. V.: Vapor-phase association of n-aliphatic carboxylic acids, *Fluid Phase Equilib.*, 222, 239–245, <https://doi.org/10.1016/j.fluid.2004.06.043>, 2004.
- Verma, N., Satsangi, A., Lakhani, A., and Kumari, K. M.: Low molecular weight monocarboxylic acids in PM<sub>2.5</sub> and PM<sub>10</sub>: Quantification, seasonal variation and source apportionment, *Aerosol Air Qual. Res.*, 17, 485–498, <https://doi.org/10.4209/aaqr.2016.05.0183>, 2017.
- 460 Wavefunction Inc.: Spartan'14, Irvine, CA, 2014.
- Wavefunction Inc.: Spartan'16, Irvine, CA, 2016.
- Weber, K. H., Morales, F. J., and Tao, F.-M.: Theoretical study on the structure and stabilities of molecular clusters of oxalic acid with water, *J. Phys. Chem. A*, 116, 11 601–11 617, <https://doi.org/10.1021/jp308499f>, 2012.
- Weber, R. J., Guo, H., Russell, A. G., and Nenes, A.: High aerosol acidity despite declining atmospheric sulfate concentrations over the past  
465 15 years, *Nat. Geosci.*, 9, 282–285, <https://doi.org/10.1038/ngeo2665>, 2016.
- Wise, M. E., Surratt, J. D., Curtis, D. B., Shilling, J. E., and Tolbert, M. A.: Hygroscopic growth of ammonium sulfate/dicarboxylic acids, *J. Geophys. Res.: Atmos.*, 108, 4638–4642, <https://doi.org/10.1029/2003JD003775>, 2003.
- Zhang, H., Xie, C., Liu, Z., Gong, J., Bao, Y., Zhang, M., Hao, H., Hou, B., and Yin, Q.-x.: Identification and molecular understanding of  
470 the odd–even effect of dicarboxylic acids aqueous solubility, *Ind. Eng. Chem. Res.*, 52, 18 458–18 465, <https://doi.org/10.1021/ie4030837>,  
2013.
- Zhao, W., Kawamura, K., Yue, S., Wei, L., Ren, H., Yan, Y., Kang, M., Li, L., Ren, L., Lai, S., Li, Y., Wang, Z., and Fu, P.: Molecular distribution and compound-specific stable carbon isotopic composition of dicarboxylic acids, oxocarboxylic acids and  $\alpha$ -dicarbonyls in PM<sub>2.5</sub> from Beijing, China, *Atmos. Chem. Phys.*, 18, 2749–2767, <https://doi.org/10.5194/acp-18-2749-2018>, 2018.

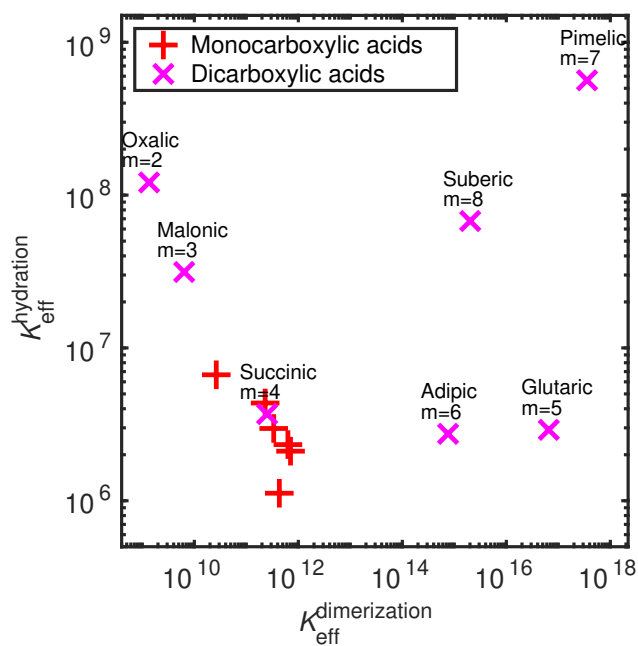




**Table 1.** List of the studied  $\alpha,\omega$ -dicarboxylic acids and their literature values of melting point and heat of fusion.

| Acid     | Chemical formula                    | Melting point (K)                 | Heat of fusion ( $\text{kJ mol}^{-1}$ ) |
|----------|-------------------------------------|-----------------------------------|---|
| Oxalic   | $\text{C}_2\text{H}_2\text{O}_4$    | $370^a$ ; $465.26^b$ ; $462.65^c$ | $3.424^a$ ; $58.158^b$                  |
| Malonic  | $\text{C}_3\text{H}_4\text{O}_4$    | $406^a$ ; $408.15^c$              | $18.739^a$                              |
| Succinic | $\text{C}_4\text{H}_6\text{O}_4$    | $458^a$ ; $461.15^c$              | $31.259^a$                              |
| Glutaric | $\text{C}_5\text{H}_8\text{O}_4$    | $369^a$ ; $372.15^c$              | $22.043^a$                              |
| Adipic   | $\text{C}_6\text{H}_{10}\text{O}_4$ | $423^a$ ; $426.15^c$              | $35.891^a$                              |
| Pimelic  | $\text{C}_7\text{H}_{12}\text{O}_4$ | $368^d$ ; $379.15^c$              | $28.9^d$                                |
| Suberic  | $\text{C}_8\text{H}_{14}\text{O}_4$ | $413^d$ ; $417.15^c$              | $38.2^d$                                |

<sup>a</sup>Booth et al. (2010); <sup>b</sup>Omar and Ulrich (2006); <sup>c</sup>Cornils and Lappe (2000); <sup>d</sup>Roux et al. (2005)

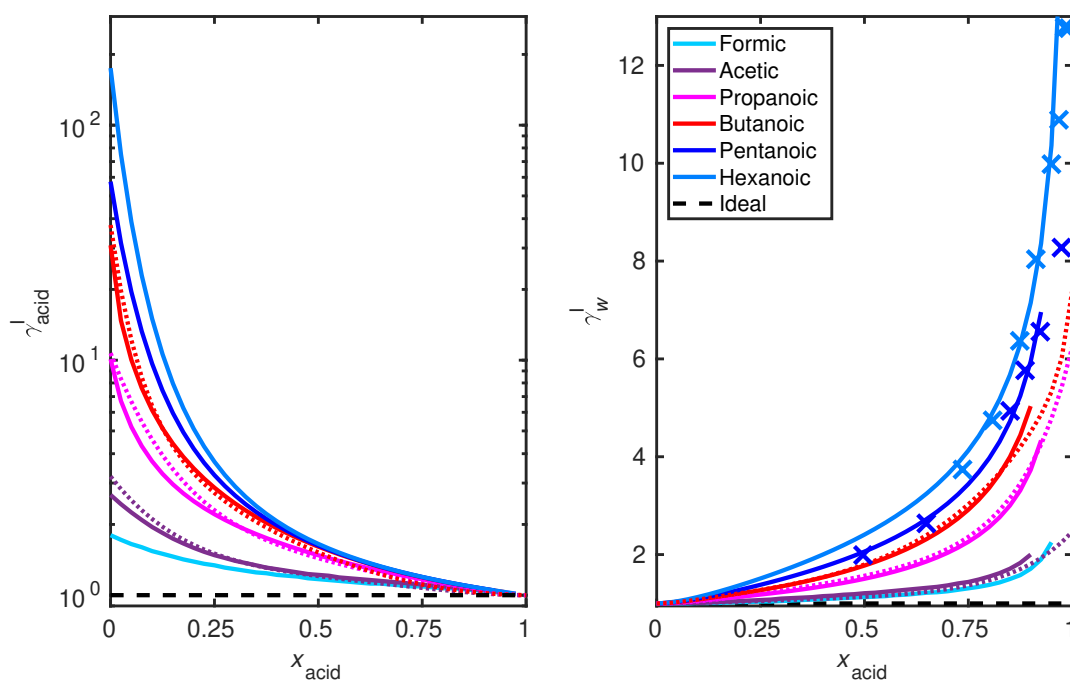


**Figure 1.** Dimensionless effective equilibrium constants of acid dimer and hydrate formation ( $K_{\text{eff}}^{\text{dimerization}}$  and  $K_{\text{eff}}^{\text{hydration}}$ , respectively, in mole fraction basis), at 298 K. See Table 2 for the values.

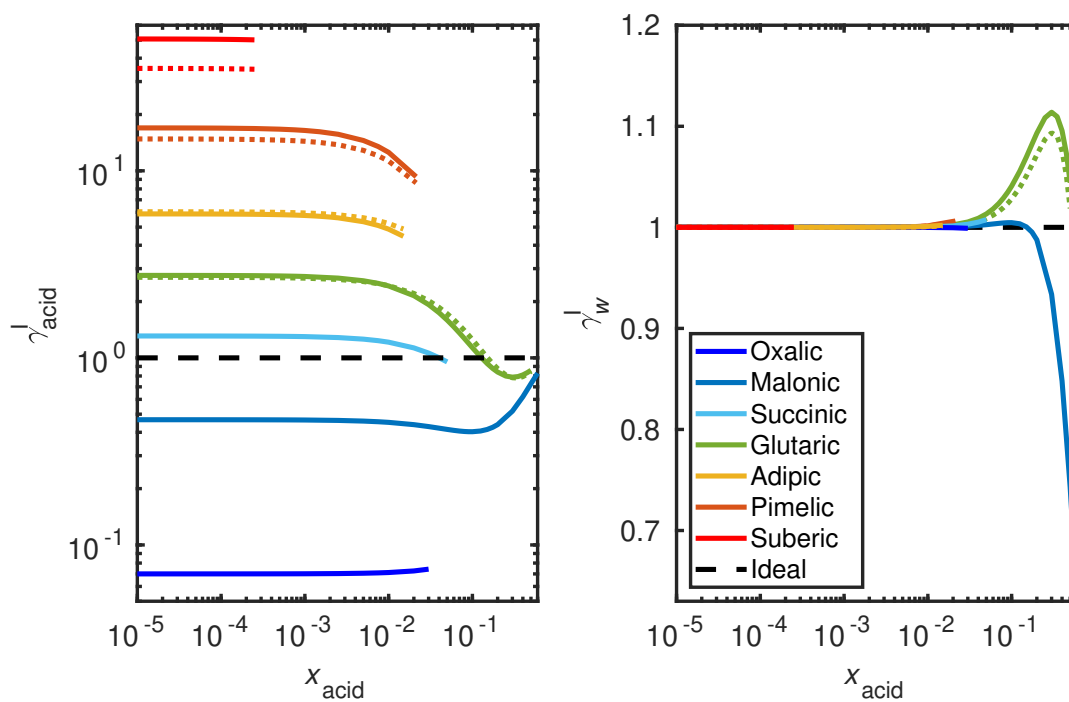


**Table 2.** Dimensionless effective equilibrium constants ( $K_{\text{eff}}$  in mole fraction basis) of cluster formation.

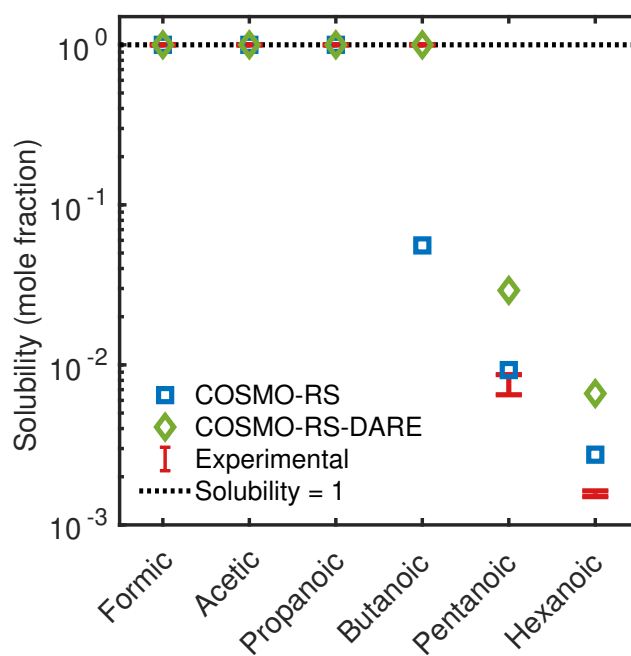
| Acid      | Carbon # | $K_{\text{eff}}^{\text{hydration}}$ | $K_{\text{eff}}^{\text{dimerization}}$ |
|-----------|----------|-------------------------------------|--|
| Formic    | 1        | $6.67 \times 10^6$                  | $2.64 \times 10^{10}$                  |
| Acetic    | 2        | $4.36 \times 10^6$                  | $2.30 \times 10^{11}$                  |
| Propanoic | 3        | $1.12 \times 10^6$                  | $4.31 \times 10^{11}$                  |
| Butanoic  | 4        | $2.97 \times 10^6$                  | $3.36 \times 10^{11}$                  |
| Pentanoic | 5        | $2.11 \times 10^6$                  | $7.08 \times 10^{11}$                  |
| Hexanoic  | 6        | $2.33 \times 10^6$                  | $6.30 \times 10^{11}$                  |
| Oxalic    | 2        | $1.21 \times 10^8$                  | $1.35 \times 10^9$                     |
| Malonic   | 3        | $3.14 \times 10^7$                  | $6.38 \times 10^9$                     |
| Succinic  | 4        | $3.69 \times 10^6$                  | $2.51 \times 10^{11}$                  |
| Glutaric  | 5        | $2.91 \times 10^6$                  | $6.56 \times 10^{16}$                  |
| Adipic    | 6        | $2.74 \times 10^6$                  | $7.58 \times 10^{14}$                  |
| Pimelic   | 7        | $5.64 \times 10^8$                  | $3.57 \times 10^{17}$                  |
| Suberic   | 8        | $6.77 \times 10^7$                  | $2.02 \times 10^{15}$                  |
| Water     | -        | -                                   | $5.71 \times 10^5$                     |



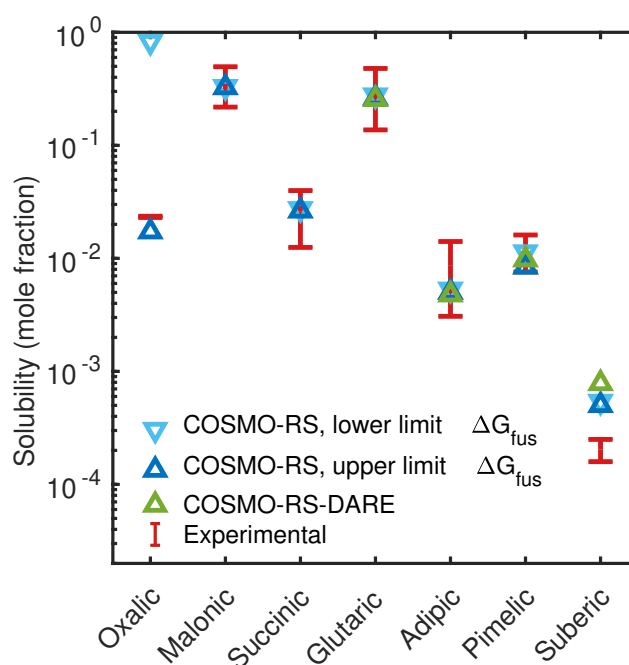
**Figure 2.** Activity coefficients ( $\gamma^1$ ) of the monocarboxylic acids (left panel) and water (right panel) in all mixing states of the binary aqueous solutions, at 298.15 K. The solid lines represent activity coefficient estimates using COSMO-RS-DARE, dotted lines are calculated from the equations fitted to experiments by Hansen et al. (1955), and the markers are the experimental points from the same study. The water activity coefficients at high  $x_{\text{acid}}$  are not shown in the figure, because COSMO-RS-DARE overpredicts the experiments by several orders of magnitude. All activity coefficient values are given in Tables S3 and S4 of the Supplement.



**Figure 3.** COSMOtherm estimated activity coefficients ( $\gamma^I$ ) of dicarboxylic acids (left panel) and water (right panel) in the binary acid–water mixtures at 298.15 K. Solid lines represent COSMO-RS estimates and dotted lines COSMO-RS-DARE estimates with the acid dimers included in the calculation. The activity coefficients were estimated up to the experimental solubility limit of the acids ( $x_{\text{acid}} \leq x_{\text{SOL,acid}}$ ). All activity coefficient values are given in Tables S5 and S6 of the Supplement.



**Figure 4.** COSMOtherm estimated aqueous solubility of monocarboxylic acids calculated using COSMO-RS and COSMO-RS-DARE, at 298.15 K. Experimental solubilities by Saxena and Hildemann (1996) ( $n = 1-6$ ) and Romero and Suárez (2009) ( $n = 5-6$ ).



**Figure 5.** Estimated solubilities of the dicarboxylic acids compared to their experimental solubilities (oxalic (Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Omar and Ulrich, 2006; Zhang et al., 2013), malonic (Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Zhang et al., 2013), succinic (Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Brooks et al., 2002; Zhang et al., 2013), glutaric (Apelblat and Manzurola, 1989; Saxena and Hildemann, 1996; Song et al., 2012; O’Neil, 2013; Zhang et al., 2013), adipic (Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Zhang et al., 2013), pimelic (Apelblat and Manzurola, 1990; Zhang et al., 2013) and suberic (Apelblat and Manzurola, 1990; Zhang et al., 2013)), at 298.15 K.