Article

Improving Solubility and Activity Estimates of Multifunctional Atmospheric Organics by Selecting Conformers in COSMOtherm

Noora Hyttinen* and Nønne L. Prisle*

Cite This: *J. Phys. Chem. A* 2020, 124, 4801−4812

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆂🅸 Supporting Information

ABSTRACT: We estimated aqueous solubilities and activity coefficients of atmospherically relevant highly oxidized multifunctional organic compounds in binary mixtures with water at temperatures between 278.15 and 338.15 K, using the COSMOtherm program. Physicochemical properties of organic aerosol constituents are needed in the modeling of atmospheric aerosol processes. As experimental data are often impossible to obtain, reliable estimates from theoretical approaches are a promising path to fill this gap. We investigated the effect of intramolecular hydrogen bonds on the estimation of these condensed-phase properties, attempting to improve the agreement between experimental and estimated values. Citric, tartaric, malic, and maleic acids, which are often used in atmospheric models as representatives of oxidized compounds, were selected to benchmark our calculations. In addition, we estimated aqueous solubilities and activity coefficients of α-pinene-derived organosulfates and highly oxidized isoprene-derived organic compounds, for which no experimental data are available. Our results indicate that the absolute aqueous solubility and activity coefficient estimates of citric, tartaric, malic, and maleic acids, and likely other multifunctional organics, can be improved significantly by selecting conformers on the basis of their intramolecular hydrogen bonding in COSMOtherm calculations.
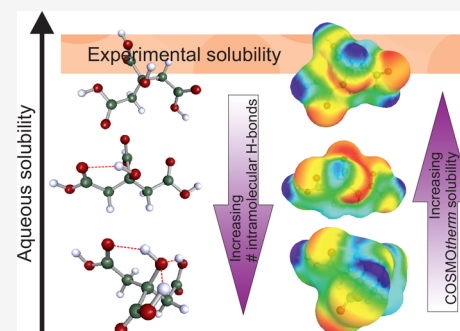
INTRODUCTION

Aerosol particles can have either warming (direct) or cooling (direct or indirect, through cloud droplets) effects in the atmosphere.[1] Fundamentally, physicochemical properties of aerosol components and all their relevant mixtures are needed to describe the formation and transformation of atmospheric aerosol particles and their climate effects. For example, properties (e.g., solubility, activity) describing interactions of aerosol particle constituents with water are needed for modeling the hygroscopic growth and cloud formation of atmospheric aerosol according to Köhler theory.[2]

The experimental determination of the relevant liquid-phase properties of atmospheric trace gases is challenged by the vast amount of different components identified in atmospheric aerosols and the exotic nature of especially many organic compounds. Even if compounds may be detected, in many cases it is not possible either to extract sufficient amounts of pure compound or to synthesize the reference compound for available analytical methods. Computational models are therefore a promising pathway to estimate these properties of compounds that are not possible to measure with sufficient accuracy using current experimental methods. One such method is the conductor-like screening model for real solvents (COSMO-RS[3−5]) theory, which uses quantum chemical data to estimate thermophysical properties of liquids.

COSMO-RS is implemented, for instance, in the COSMO-therm program. COSMOtherm has been used in several studies to estimate the solubilities, activity coefficients, and partitioning coefficients of atmospherically relevant molecules.[6−12] Estimated absolute solubilities in these previous works do not agree well with experimental solubilities, whereas the relative temperature dependence of the solubilities is described better.[7] In addition, solubility estimates using COSMOtherm at different temperatures can be improved if the solubility is already known at some temperature(s). The description of the compound at the reference temperature is then used, as opposed to using the experimental heat of fusion and melting temperature, to estimate the free energy of fusion. For instance, Reinisch et al.[13] were able to predict liquid−liquid equilibria (LLE) of polyether water mixtures at different temperatures using the experimental solubility at 298.15 K as a reference in the COSMOtherm calculation. Other methods have also been proposed to further improve the COSMOtherm estimates. Reinisch et al.[13] also achieved better agreement between experimental and estimated solubilities by rescaling the molecular surface areas in the COSMOtherm calculation.

The advantage of COSMOtherm, compared to, for example, group contribution methods, is the ability to take into account

the position of the functional groups in the molecule. The possibility of intramolecular H-bond formation is included by using multiple conformers with different hydrogen bonding patterns in the calculation. In different solutions, the conformer distributions of multifunctional compounds vary largely on the basis of free energies of each conformer in the solution. For multifunctional compounds, all important hydrogen bonding patterns are not taken into account if only a small set of conformers (normally 10 conformers) is included in COSMO-*therm* calculations. Recently, Kurtén et al.[14] showed that COSMO*therm* estimates of saturation vapor pressures are highly dependent on the conformational sampling of multifunctional compounds containing hydroperoxide groups and proposed a more comprehensive conformer sampling and increasing the number of conformers used in the COSMO*therm* calculations. In addition, Kurtén et al. investigated the effect of the treatment of intramolecular hydrogen bonding on the estimated saturation vapor pressures and found that selecting conformers containing fewer intramolecular H-bonds gave a better agreement between the estimates and experiments because of the tendency of the computational level of theory to overestimate the favorability of intramolecular H-bonds.[14]

Here, we continue to investigate the effect of the treatment of intramolecular hydrogen bonding in the conformers chosen for COSMO*therm* calculations, by focusing on estimations of condensed-phase properties for multifunctional organic compounds. While there are no experimental data on highly oxygenated organic molecules (HOM[15]), we here use the relatively well-constrained and atmospherically relevant citric acid ($C_6H_8O_7$, CA) to benchmark our calculations. As a representative model for highly oxidized soluble multifunctional secondary organic aerosol (SOA) compounds, citric acid is used to gauge which types of conformers give the best agreement between experiments and COSMO*therm* predictions. Citric acid has four hydrogen bond donating functional groups, which allows for the investigation of the effect of selecting conformers with different number of intramolecular H-bonds in the COSMO*therm* calculations. In addition, many studies[16−29] are available on different physicochemical properties of citric acid, which allows for the comparison between the COSMO*therm* estimates and experimental values. As additional benchmark compounds for aqueous solubility and water activity, we use tartaric ($COOH(CHOH)_2COOH$), malic ($HOOCCH_2CHOHCOOH$), and maleic ($HOOCCH=CHCOOH$) acids, which have some experimental data available for a comparison.

As other examples of SOA components, we also include isoprene-derived dihydroperoxy hydroxy aldehyde ($C_5H_{10}O_6$, all isomers) and dihydroxy dihydroperoxide ($C_5H_{12}O_6$, iso1),[14] and α-pinene-derived organosulfates (R-OSO$_3$H, α-pinene-OS[30]). Although experimental data are not available for these compounds, they may contribute significant SOA mass in the atmosphere and have previously been included in COSMO-*therm* calculations. We therefore utilize previously generated cosmo-files from studies that used the same systematic conformer sampling scheme. The chemical structures of the studied isoprene- and α-pinene-derived compounds are shown in Figures S1 and S2 of the Supporting Information, respectively.

■ COMPUTATIONAL METHODS

Input File Generation. We use the COSMO*therm*[31] program release 19 to calculate various condensed-phase properties of the studied compounds. The cosmo-files used as input in the COSMO*therm* calculations are generated using the following systematic method used in our previous studies.[14,30] A systematic conformer search is performed using MMFF force fields in Spartan'14.[32] This method, similar to the conformer sampling algorithms in COSMO*conf*, finds the conformers in a vacuum. These conformers are then used as input in COSMO*conf* (version 4.2).[33] COSMO*conf* is used to optimize condensed- and gas-phase geometries and calculate screening charge densities of the conformers and to discard duplicate conformers and conformers with similar chemical potentials in a set of solvent compounds. Our template in the COSMO*conf* calculations follows the default BP-TZVPD-FINE-COSMO template without the initial conformer sampling, and setting the maximum number of conformers at the final single-point energy calculation to 500. The final cosmo-files are calculated at the BP/def2-TZVPD-FINE//BP/def-TZVP level of theory and the most recent parametrization (BP_TZVPD_FINE 19) is used in the COSMO*therm* calculations. Cosmo-files for the isoprene-derived compounds are taken from Kurtén et al.[34] and for the organosulfates from Hyttinen et al.[35]

We test the effect of selecting conformers on COSMO*therm*-estimated condensed-phase properties solubility, activity coefficient, and pK_a. Conformers are grouped on the basis of their intramolecular hydrogen bonding. A similar comparison was done for the gas phase by Kurtén et al.[14] in estimation of saturation vapor pressures. COSMO*therm* is able to indicate which of the hydrogen bond donating functional groups have their hydrogen bonding area partially (partial H-bond) or completely free (no H-bond). On the basis of this information, we form groups of conformers according to the maximum number of full intramolecular H-bonds (see Figure 1), so that "0 H-bonds" ⊆ "1 H-bond" ⊆ "2 H-bonds" ⊆···⊆ "all conformers".
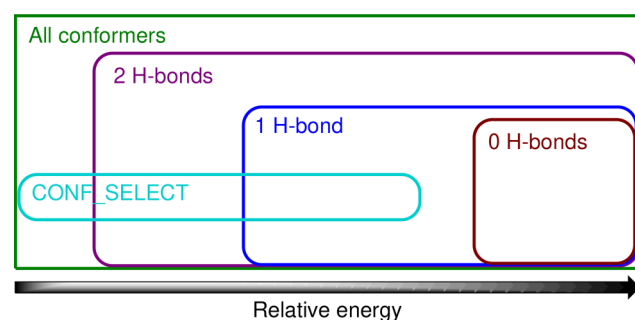
Figure 1. Different conformer sets used in the COSMO*therm* calculations. The 2, 1, and 0 H-bond conformer sets are formed on the basis of the maximum number of intramolecular H-bonds. Conformers in the CONF_SELECT set are selected on the basis of their chemical potential and can have any number of intramolecular H-bonds.

In general, the COSMO energy of a conformer containing multiple intramolecular H-bonds is lower (more favorable) than that of a conformer containing no hydrogen bonds (see section S1.1 and Tables S1 and S2 of the Supporting Information). In COSMO*therm* calculations, conformers (chosen for the calculation based on their COSMO energies) are weighted on the basis of the sum of their COSMO energy and chemical potential in the solution using Boltzmann distribution and assuming an equilibrium. Selecting conformers for COSMO-*therm* calculations solely on the basis of their COSMO energies, the high energy 0 H-bond conformers are not necessarily considered in the calculation. This depends on the total number

of conformers and the flexibility of the molecule. Limiting the number of intramolecular H-bonds in the conformers that are selected to the COSMO*therm* calculation, will therefore increase the number of conformers containing fewer intramolecular H-bonds.

Additionally, we use the CONF_SELECT method to select relevant citric acid conformers in the binary citric acid−water system. CONF_SELECT is a COSMO*therm* calculation step implemented in COSMO*conf*, to select conformers on the basis of their weights in different solvents. The default set of solvents contains 31 different inorganic and organic compounds from the COSMO*base*.[36] COSMO*base* is a database of precalculated cosmo-files needed for COSMO-RS calculations. Here, we use water and citric acid (COSMO*base* conformers) as the solvents instead of the default 31 different compounds, in order to find conformers of citric acid that are relevant in binary citric acid−water solutions. The CONF_SELECT method finds the highest weighted conformer at infinite dilution in each of the solvents and adds each of those conformers to the final conformer set. In addition, CONF_SELECT uses a threshold parameter that describes the allowed change in the free energy of the compound in all of the different solvents, when a single conformer is removed from the full set of conformers. The free energy of the compound is calculated in the different solvents using the full set of conformers and after omitting a single conformer. If the change in free energy calculated using these two sets of conformers is below the threshold in all of the solvents, the conformer is removed from the conformer set (as it has only a small effect on the total free energy in each solvent). This means that decreasing the threshold value increases the number of conformers that the method saves.

Neutral citric acid contains four hydrogen bond donating functional groups (three carboxylic acid groups and one hydroxy group). In a deprotonated citric acid molecule, one of the protons of a carboxylic acid group is removed, leaving three hydrogen bond donors (the structures are shown in Figure S3 of the Supporting Information). Similarly, the isoprene-derived $C_5H_{12}O_6$ contains 4 hydrogen bond donating functional groups but all of the 500 lowest-energy conformers contain at least one intramolecular hydrogen bond[14] due to the flexibility of the molecule and the large total number of conformers (the highest energy conformers were omitted from the COSMO*conf* calculations due to the 500 conformer cutoff). We therefore rerun the COSMO*conf* calculation for the two diastereomers (iso1-S,S and iso1-S,R) without any cutoffs in order to find conformers with no intramolecular hydrogen bonds (using COSMO*conf* version 4.3[37]).

For citric, tartaric, malic, and maleic acids, we also compare the effect of using the cosmo-files available in COSMO*base* on the estimated condensed-phase properties. COSMO*base* contains only at most 10 conformers of each compound. We therefore use all of the conformers from the COSMO*base* instead of omitting conformers containing intramolecular hydrogen bonds. The α-pinene- and isoprene-derived compounds studied here are not available in the COSMO*base*.

COSMO*therm* Calculations. We use COSMO*therm* to calculate solubility and activity coefficients for each organic compound and water in their binary mixtures. In addition, we calculate the first acid dissociation constant (pK_a) of citric acid in water, with K_a corresponding to the first deprotonation equilibrium $C_6H_8O_7 + H_2O \rightleftharpoons C_6H_7O_7^- + H_3O^+$. Equations used to calculate each of these properties are summarized below,

more details can be found from the COSMO*therm* Reference Manual.[38]

Activity Coefficients. The activity coefficient (γ_i) of compound i with mole fraction x_i is calculated from the pseudochemical potential[39] μ_i^* using the pure compound as the reference state (convention I[40]):

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

The pseudochemical potential is defined using the standard chemical potential at the reference state $\mu_i^\circ(T, P)$:

$$\mu_i^*(x_i) = \mu_i^\circ(T, P) + RT \ln \gamma_i \tag{2}$$

where R is the gas constant, T is the temperature, and the reference pressure P is 10^5 Pa. The same temperature is used for the reference state and the actual state $\{x_i\}$.

The value of the activity coefficient in a given actual state $\{x_i\}$ depends on the chosen reference state and by definition, $\gamma_i = 1$ in the reference state $\{x_i\}^\circ$. Often, experimental activity coefficients are derived with respect to the infinite dilution of the solute. To change the reference state of the activity coefficient calculation of a solute from pure compound (γ_i^I) to infinite dilution (γ_i^{II}, labeled as convention II[40]), the activity coefficient calculated at the actual state $\{x_i\}$ is divided by the activity coefficient calculated at the infinite dilution state[40] ($x_i \rightarrow 0$) with respect to the pure component reference:

$$\begin{aligned} \ln \frac{\gamma_i^I(x_i)}{\gamma_i^I(x_i \rightarrow 0)} &= \ln \gamma_i^I(x_i) - \ln \gamma_i^I(x_i \rightarrow 0) \\ &= \frac{\mu_i^*(x_i) - \mu_i^{*\circ,I}(T, P)}{RT} \\ &\quad - \frac{\mu_i^*(x_i = 0) - \mu_i^{*\circ,I}(T, P)}{RT} = \frac{\mu_i^*(x_i) - \mu_i^*(x_i = 0)}{RT} \\ &= \frac{\mu_i^*(x_i) - \mu_i^{*\circ,II}(T, P)}{RT} \\ &= \ln \gamma_i^{II}(x_i) \end{aligned} \tag{3}$$

The reference state of water activity coefficients is the same in convention I and convention II ($x_w = 1$).

The experimental values for citric acid activity coefficients[21,25,41] are given on a molality basis, as opposed to the mole fraction basis that is used by COSMO*therm*. To compare experimental molality basis activity coefficients (γ_b^{II}) with our mole fraction basis activity coefficients, the experimental activity coefficients are converted to the mole fraction basis as

$$\gamma_i^{II} = \gamma_{b,i}^{II}(1 + 0.001M_w b_i) \tag{4}$$

where b is the molality (mol kg^{-1}) of the compound and M_w is the molar mass (g mol^{-1}) of the solvent water.[42]

Solubility. With relatively high melting temperatures (see Table 1), citric, tartaric, malic, and maleic acids are solid at atmospherically relevant temperatures. The aqueous solubilities (x_{SOL}) are therefore estimated by solving the solid−liquid equilibrium (SLE) of the binary acid−water system:

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

Table 1. Experimental Melting Points ($T_{\text{melt,i}}$) and Heats of Fusion ($\Delta H_{\text{fus,i}}$) of Citric, Tartaric, Malic, and Maleic Acids

acid	melting point (K)	heat of fusion (kJ mol^{-1})
citric	427.8[a]	40.32[a]
tartaric	445.04[a]	36.31[a]
malic	376[b]; 403[c]	23.012;[b] 29.031[c]
maleic	411.85[d]	26.9[d]

[a]Meltzer and Pincu.[29] [b]Leclercq et al.[43] [c]Booth et al.[44] [d]Guo et al.[45]

where $\Delta G_{\text{fus,i}}(T)$ is the temperature dependent molar free energy of fusion of the solute. When ΔG_{fus} is unknown, it is estimated by COSMO*therm* from the heat of fusion ($\Delta H_{\text{fus,i}}$), melting temperature ($T_{\text{melt,i}}$), and the heat capacity of fusion ($\Delta C_{p,\text{fus,i}}$) of the compound:

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

We use experimentally determined values of the heat of fusion and melting point, and the heat capacity of fusion is estimated from the two:

$$\Delta C_{p,\text{fus,i}} = \frac{\Delta H_{\text{fus,i}}}{T_{\text{melt,i}}} \tag{7}$$

Combining eqs 1 and 5 gives the equilibrium condition

$$\ln a_i^{\text{I}}(x_{\text{SOL,i}}) = \frac{\Delta G_{\text{fus,i}}(T)}{RT} \tag{8}$$

by which the activity ($a_i^{\text{I}} = x_i \gamma_i^{\text{I}}(x_i)$) of a compound at the solubility limit only depends on the free energy of fusion and the temperature.

When a liquid–liquid equilibrium (LLE) exists, the system consists of two distinct liquid phases, a solvent-rich phase (α) and a solute-rich phase (β), and the LLE forms between these two phases, as opposed to a pure solute phase and a solvent-rich phase. Solubilities for liquid compounds (isoprene- and α-pinene-derived compounds) are therefore estimated by solving the liquid–liquid equilibrium condition:

$$a_i^{\text{I},\alpha} = a_i^{\text{I},\beta} \tag{9}$$

For acidic compounds, solubilities can furthermore be corrected by taking the dissociation of the acid in the solvent water into account. For this, we need to know the acid constant (pK_a) in water. We evaluate the performance of the COSMO*therm* pK_a calculation by comparing estimated pK_a values of citric acid to those found in the literature. Only the first pK_a is calculated, since according to experiments, the second pK_a of citric acid is much higher than the first,[46] so that the second deprotonation step only has a small effect on the equilibrium position of the first deprotonation. COSMO*therm* estimates the pK_a of compound i from the molar free energy (G) of the neutral and ionic species at infinite dilution, using the linear free energy relationship (LFER):

$$pK_a^i = c + d(G_i^{\text{anion}} - G_i^{\text{neutral}}) \tag{10}$$

We use LFER parameters c (=−130.152) and d (=0.116 mol kJ^{-1}) for water as solvent from COSMO*therm*'s parameter file.

The concentration of ionic species from dissociation ($c_i^{A^-}$) is calculated using the molar concentration of dissolved, undissociated acid, $c_i^{\text{HA}} = x_i \rho / M$.

$$c_i^{A^-} = -0.5 \times 10^{-\text{pH}} + \sqrt{0.25 \times 10^{-2\text{pH}} + c_i^{\text{HA}} 10^{-pK_a}} \tag{11}$$

The density (ρ) and average molar mass (M) of aqueous citric acid solutions used in the conversion between mole fractions and molar concentrations are given in section S1.2 and Figure S4 of the Supporting Information.

■ RESULTS AND DISCUSSION

Carboxylic Acids. *Solubility.* We estimated the temperature dependent binary aqueous solubility for each organic compound, using the different sets of conformers in the COSMO*therm* calculation. The SLE for the carboxylic acid–water solutions (anhydrous, amorphous solid acid, and liquid-phase water) were found in 10 K intervals between 278.15 and 338.15 K. Figure 2 shows a comparison between the calculated
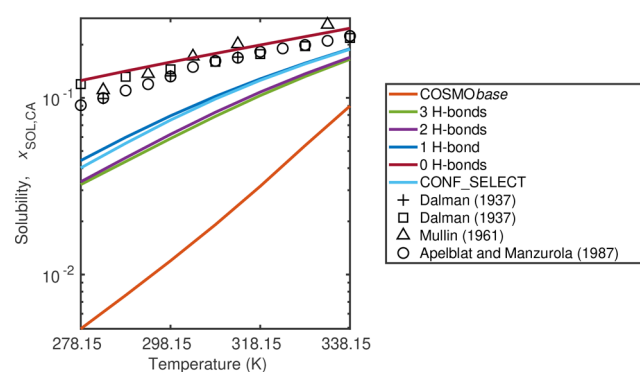
Figure 2. Solubility of citric acid as a function of temperature resulting from using different sets of conformers in the COSMO*therm* calculation. Experimental values are from Dalman (plus,[16] square[17]) Mullin[18] and Apelblat and Manzurola.[19]

and experimental[16−19] solubilities for citric acid as a function of temperature. The aqueous solubilities of tartaric, malic, and maleic acids are shown in Figure S5a−c of the Supporting Information, respectively. In general, using the COSMO*base* conformers produces the lowest solubility estimates at all temperatures. One exception is maleic acid, for which our systematic conformer sampling scheme found one less unique conformer than what is found in the COSMO*base*, leading to slightly higher aqueous solubility when the COSMO*base* conformers were used, compared to using all conformers found from systematic conformer sampling. The CONF_SELECT conformer set (citric acid) leads to a solubility estimate similar to that of the 1 H-bond conformer set, which can be explained by the high fraction (74%) of the conformers containing a single intramolecular H-bond in the CONF_SELECT conformer set.

For citric acid, the experimental solubilities are underestimated by 1 order of magnitude at 298.15 K using the COSMO*base* conformers. We ran additional calculations to see what value of ΔG_{fus} would produce agreement between the experimental citric acid solubilities (Apelblat and Manzurola[19]) and COSMO*therm* estimates calculated using the COSMO*base* conformers. The free energy of fusion calculated from the experimental heat of fusion and melting temperature using eq 6

is 2.2−5.2 kJ mol^{-1} higher (Figure S6 in section S1.3 of the Supporting Information) than what is required for an agreement between the experimental and estimated solubilities using the COSMO*base* conformers.

Using conformers with no intramolecular hydrogen bonds gives the best agreement with the experimental solubility values of citric, tartaric, and maleic acids. This is reasonable, since water is a highly polar solvent that is able to interact with both the hydrogen bond donating and accepting functional groups of the solute. When a conformer has fewer intramolecular H-bonds, more of the hydrogen bonding area is available for intermolecular interactions with the solvent. The experimental heat of fusion and melting point values of malic acid have large variations, leading to large differences in the COSMO*therm*-estimated solubilities of malic acid. Using the higher melting point and heat of fusion values,[44] we get a good agreement between experimental[19] and calculated solubilities using the 0 H-bond conformer set, if we assume that $\Delta C_{p,\text{fus}} = 0$ kJ mol^{-1} K^{-1}. Using the $\Delta C_{p,\text{fus}} = \Delta H_{\text{fus}}/T_{\text{melt}}$ estimate, the best agreement between experiments and COSMO*therm* estimates is found using all conformers from systematic conformer sampling.

Aqueous solubilities calculated for citric acid using $\Delta C_{p,\text{fus}} = 0$ kJ mol^{-1} K^{-1}, instead of the $\Delta C_{p,\text{fus}}$ estimate of eq 7, are shown in Figure S7 and Table S3 of the Supporting Information. This assumption leads to lower solubility estimates for all the used conformer sets. The difference in solubilities calculated using the two $\Delta C_{p,\text{fus}}$ values decreases with increasing temperature and is on average a factor of 3.1 at 278.15 K and 1.3 at 338.15 K. Assuming $\Delta C_{p,\text{fus}} = 0$ kJ mol^{-1} K^{-1} underestimates the aqueous solubility of Apelblat and Manzurola[19] by a factor of 1.3.

We also calculated dissociation corrected solubilities for citric acid at 298.15 K using the different conformer sets (0, 1, 2, and 3 H-bonds) and first pK_a of citric acid estimated by COSMO*therm* (see section S1.4 and Figure S8 of the Supporting Information for the pK_a calculations). The pK_a values were estimated using the corresponding conformer sets for the neutral and deprotonated citric acid; e.g., the 0 H-bond conformer set was used for both the neutral and the deprotonated citric acid to calculate the pK_a value used to correct the solubility calculated using the 0 H-bond conformer set. The estimated pK_a is the highest using the 0 H-bond conformer set and the lowest using the 3 H-bond conformer set. The experimental pK_a (3.13[20]) falls between the estimates calculated using the 0 and 1 H-bond conformers. For the COSMO*base* conformer set, we used the experimentally determined value of pK_a since the deprotonated species needed for the pK_a calculation was not available in the COSMO*base*. Density calculations for aqueous citric acid solutions (needed for the conversion from mole fraction to molar concentration) using COSMO*therm* are given in section S1.2 of the Supporting Information. The increase in dissociation corrected (mole fraction) solubility, compared to the uncorrected estimates, is smallest (1.4%) using the 0 H-bond conformer set due to the high pK_a (3.6) relative to the other conformer sets and experiments. The estimated pK_a is the lowest (1.8) using the 3 H-bond conformer set, which leads to a 13.9% increase in the estimated solubility when dissociation is included in the calculation. At 298.15 K, the dissociation corrected 3 H-bond solubility is still underestimating the experimental values[19] by a factor of 4.5, as opposed to a factor of 5.2 for the uncorrected solubility estimate.
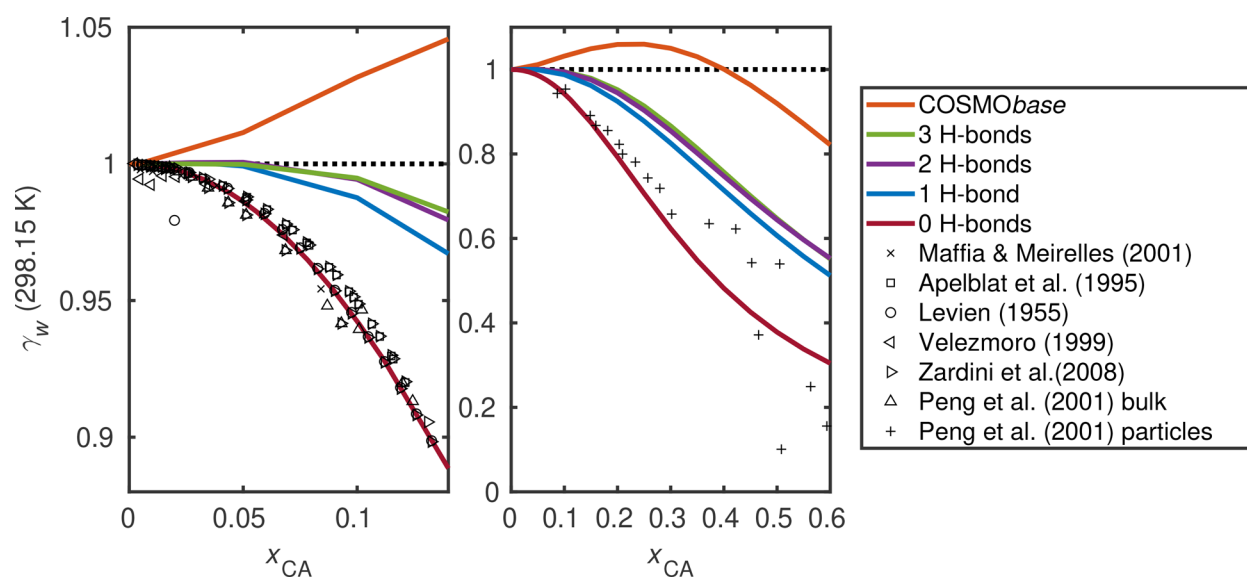
Activity. We calculated the activity coefficients of citric acid and water in binary citric acid−water mixtures and compared

our estimated values with experimentally determined activity coefficients.[21,25,41,47−50] For binary aqueous mixtures of tartaric, malic, and maleic acids, we found experimental acid activity data only for malic and maleic acids measured in a short acid mole fraction range ($0.009 \leq x_{\text{acid}} \leq 0.05$) using the 1 molal ($x_{\text{acid}} = 0.018$) solution as a reference.[51] We are therefore not comparing tartaric, malic, and maleic acid activity coefficient estimates with experiments. Parts a and b of Figure 3 compare estimated and experimentally determined[21,25,41] activity *coefficients* of citric acid in water.
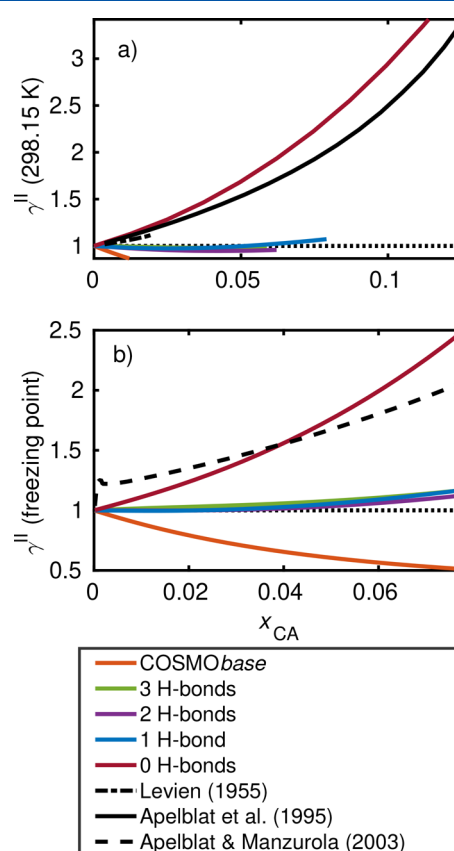
Figure 3. Estimated and experimentally determined (Levien,[21] Apelblat et al.,[41] and Apelblat and Manzurola[25]) activity coefficients (γ^{II}) of citric acid at (a) 298.15 K and (b) the freezing points[25] of binary aqueous mixtures, as a function of citric acid mole fraction x_{CA}. The black dotted line represents ideal mixing, $\gamma^{\text{II}} = 1$.

All sets of experimentally determined activity coefficients were given on a molality basis and for comparison converted to a mole fraction basis using eq 4. Derivation of the experimental activity coefficients is explained in section S1.5 of the Supporting Information. To facilitate comparison to these experimental values, COSMO*therm* estimates for citric acid activity coefficients were made at the experimentally determined freezing temperatures of each binary mixture.[25] Since the experimental values are derived from freezing points, we here assumed that the reference state temperature is the freezing temperature of the reference state solution ($T° = 273.125$ K). However, the reference state temperatures of COSMO*therm* estimates are the freezing temperatures of each of the calculated mixtures. This difference in reference state temperature may lead to small discrepancies between experimental values and COSMO*therm* estimates.

Figure 4. Estimated and experimentally determined water activity coefficients (γ_w) at 298.15 K in binary citric acid−water mixtures, as a function of citric acid mole fraction x_{CA}. The left panel shows comparison with bulk measurements,[21,41,47−50] and the right panel shows that with particle measurements.[50] The black dotted line represents ideal mixing, $\gamma_w = 1$.

The experimentally determined activity coefficients (convention II) of citric acid in aqueous solutions are above 1 at all binary mixing states, similar to the estimates using the 0 H-bond conformer set. For the other conformer sets from the systematic conformer search, COSMO*therm*-estimated activity coefficients are slightly above (at the freezing point) or below (298.15 K) ideality ($\gamma^{II} = 1$). The experimental activity coefficients for citric acid derived at the freezing points span a wide composition range, including also very dilute citric acid states. In the data sets based on osmotic coefficients the lowest citric acid mole fraction is an order of magnitude higher. Due to this lack of experimental points for very dilute solutions in the Levien[21] and Apelblat et al.[41] data sets, the behavior at $x_{CA} < 1.8 \times 10^{-3}$ cannot be resolved in these cases and the extrapolation to lower mole fractions does not predict similar behavior, as seen by Apelblat and Manzurola.[25] The estimated citric acid activity coefficients at infinite dilution and convention I are 0.019, 0.197, 0.281, 0.288, and 1.614, calculated using 0, 1, 2, 3 H-bond and COSMO*base* conformer sets, respectively.

Water activities in aqueous citric acid solutions have been measured in both bulk- and particle-phase experiments.[21,41,47−50] A comparison between experimentally determined water activity coefficients and COSMO*therm* estimates is shown in Figure 4. We see also for water in the mixtures how the estimated activity coefficients calculated using the 0 H-bond conformer set agree very well with the experimental values, while the other conformer sets predict activity coefficients close to ideal ($\gamma_w = 1$).

Peng et al.[50] measured water activities in binary tartaric acid−water and malic acid−water mixtures, Choi and Chan[52] measured water activities in binary maleic acid−water mixtures. In addition, Maffia and Meirelles[47] have reported water activities of tartaric and malic acids in acid mole fractions below the solubility limits of the acids ($x_{acid} < 0.12$). A comparison between experimental and COSMO*therm*-estimated water activity coefficients for tartaric, malic, and maleic acids is shown in Figure S9a−c of the Supporting Information, respectively. For all three acids, the best agreement between experiments and calculation is found using the 0 H-bond

conformer set. In the bulk phase, COSMO*therm*-estimated water activities are very close to the experiments, while in the particle phase, COSMO*therm* tends to overestimate the experimental water activity coefficients (see Figure S9 in the Supporting Information). The water and acid activity coefficients calculated using the 0 H-bond conformer set at 298.15 K are shown in Table S4 of the Supporting Information.

The conformer distribution of multifunctional compounds in equilibrium depends highly on the solution. As an example, we have used a set of 10 conformers (one half from the 3 H-bond conformer set and the other half containing no intramolecular H-bonds) to investigate the conformer distributions of citric acid in binary citric acid−water solutions. In pure citric acid, conformers containing no intramolecular H-bonds have 0.07 weight whereas the conformers containing multiple H-bonds have 0.93 weight of the total conformer distribution. In a dilute aqueous solution, the corresponding weights are 0.81 and 0.19 of the total conformer distribution, respectively (see section S1.6, Figure S10, and Table S5 of the Supporting Information). Using the infinite dilution reference state and calculating properties for citric acid in states close to the reference state, the pure compound is not taken into account. In this case, it is especially critical to include conformers that have a high weight at infinite dilution, to improve the description of the strong intermolecular interaction between citric acid and the solvent water. Using conformers containing intramolecular H-bonds in COSMO*therm* calculations clearly leads to the underestimation of the intermolecular interactions between citric acid and solvent water, compared to what is seen in experiments.

We additionally calculated activities of both citric acid and water in binary mixtures ranging from pure water to the solubility limit of citric acid and 278.15−338.15 K. Activities calculated using different sets of conformers are shown in Figure 5a−e and Figure 6a−e for citric acid and water, respectively. The activity coefficient values of Figures 5e and 6e can be found in Table S6 of the Supporting Information.

At each temperature, the value of the activity of citric acid at the aqueous solubility limit $a_{CA}(x_{SOL,CA})$ is independent of the conformers used in the COSMO*therm* calculations, as can be
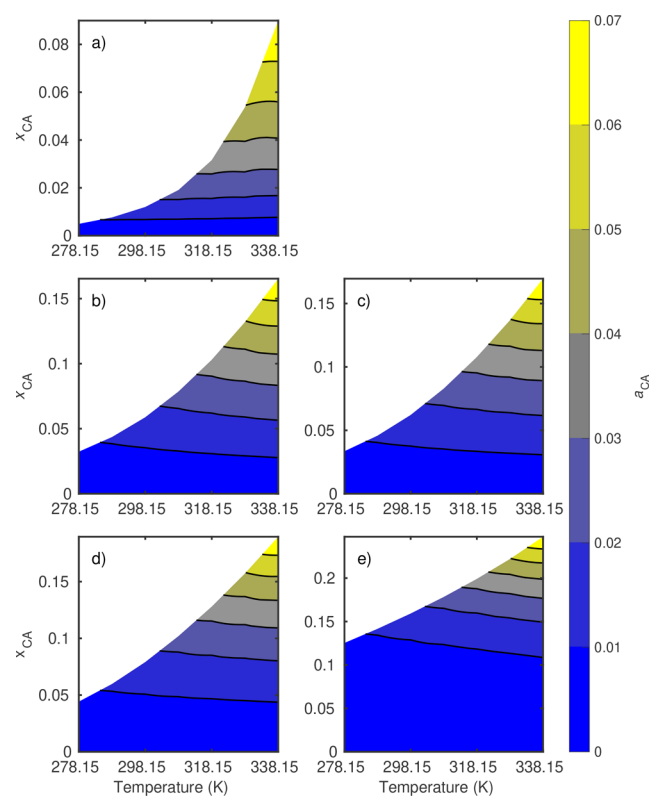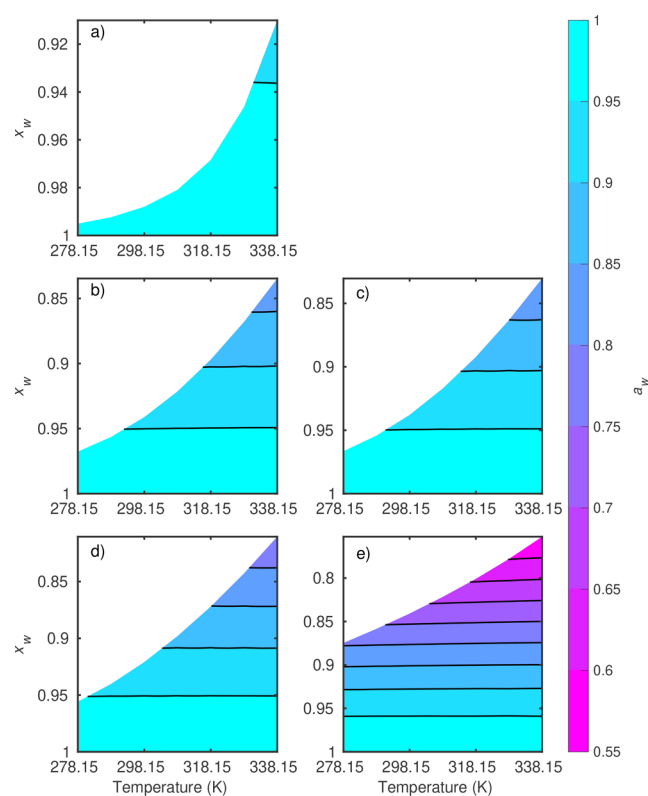
Figure 5. COSMO*therm*-estimated activity of citric acid (a_{CA}) in binary citric acid−water mixtures with different mole fractions of citric acid (x_{CA}) calculated using (a) COSMO*base* conformers and conformers from the (b) 3 H-bond, (c) 2 H-bond, (d) 1 H-bond, and (e) 0 H-bond conformer sets.

Figure 6. COSMO*therm*-estimated water activity (a_w) in binary citric acid−water mixtures with different mole fractions of water (x_w) calculated using (a) COSMO*base* conformers and conformers from the (b) 3 H-bond, (c) 2 H-bond, (d) 1 H-bond, and (e) 0 H-bond conformer sets. Note the reverse *y*-axis.

seen from eq 8. However, as the solubility ($x_{SOL,i}$) itself depends on the conformers used, the activity *coefficients* at the solubility limit ($\gamma_{CA}^{I}(x_{SOL,CA}) = a_{CA}^{I}(x_{SOL,CA})/x_{SOL,CA}$) in turn do as well. For example, at 298.15 K, $x_{SOL,CA}$ varies between 1.2×10^{-2} and 1.6×10^{-1} using all of the different conformer sets but the value of the activity at each solubility limit is 1.6×10^{-2}. In Figure 5 this can be seen from the contour lines, which intersect the solubility limit (the border between the white and the colored areas) at the same temperatures in all of the panels. The variation in citric acid activity with conformers used can be seen at all of the other binary mixing ratios ($x_{CA} < x_{SOL}$).

Using the COSMO*base* conformers, the estimated activity of citric acid is higher than its mole fraction ($\gamma_{CA}^{I} > 1$) at all of the temperatures and mixing states, except for the highest mole fractions at 338.15 K (the yellow region). For each of the other conformer sets, citric acid activity is below the corresponding mole fraction ($\gamma_{CA}^{I} < 1$) at all temperatures and mixing states. Decreasing the number of intramolecular H-bonds in the conformers used in the COSMO*therm* calculation decreases the aqueous activity of citric acid, meaning that mixing interactions are more favorable, compared to pure citric acid. Placing too much emphasis on stabilizing intramolecular H-bonds therefore leads to underestimation of the stabilizing interaction between the citric acid and water.

The water activity in binary citric acid−water solutions is close to the ideal value ($a_w \sim x_w$) at all mixing states, except when calculated using the 0 H-bond conformer set. This deviation from ideality can be explained by the estimated solubility of citric acid, which is higher using the 0 H-bond conformer set

compared to the other conformer sets. At low concentrations of citric acid ($x_{CA} < 0.05$) the water activity is close to ideal, similar to estimates using the other conformer sets. We also note that COSMO*therm*-estimated water activities have no clear temperature dependence in the binary citric acid−water mixtures.

For citric, tartaric, maleic, and malic acids, we were able to compare our estimates with experimental condensed-phase data. We also investigated the effect of taking conformer selection into account for estimated solubilities and activities of α-pinene-derived organosulfates and isoprene-derived multifunctional oxidized compounds. For these atmospherically relevant compounds, to our best knowledge, no corresponding experimental data is available in the literature.

α-Pinene-Derived Organosulfates. The aqueous solubilities and activities of the α-pinene-OS have previously been estimated at 298.15 K by Hyttinen et al.[30] using the 0 H-bond conformer set. Here, we show the aqueous solubility estimates and activity coefficients at varying temperatures and compare the values calculated using different conformer sets in COSMO*therm*.

Solubilities for α-pinene-derived organosulfates are shown in Figure 7a−f. Generally, the estimated solubilities of organosulfates increase when the number of intramolecular H-bonds is limited in the COSMO*therm* calculation. This is in agreement with what was seen in the solubility estimates of citric, tartaric, malic, and maleic acids using the different conformer sets. Using conformers that interact more with the polar solvent water increases the solubility estimate compared to when conformers containing intramolecular H-bonds are used. The only exception is α-pinene-OS-4, which has three carbonyl groups
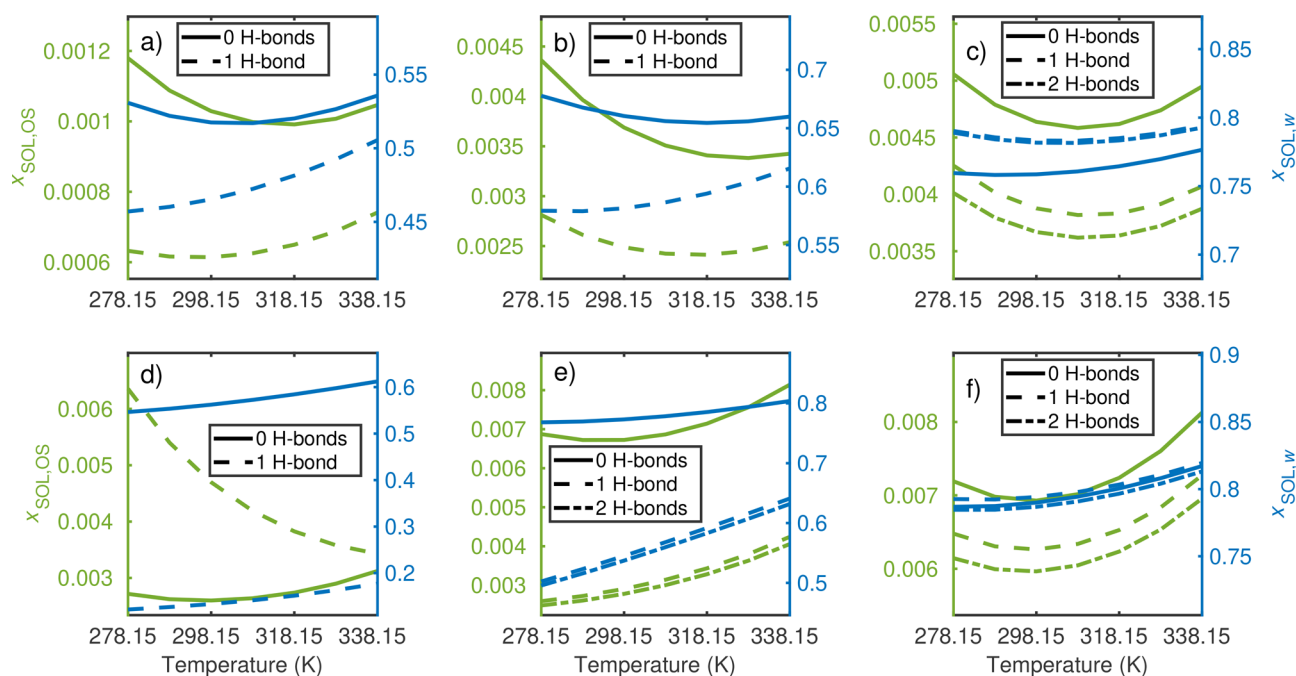
Figure 7. Solubilities of α-pinene-derived organosulfates in pure water ($x_{\text{SOL,OS}}$, green lines) and of water in the corresponding organosulfate-rich phases ($x_{\text{SOL,w}}$, blue lines) as a function of temperature: (a) α-pinene-OS-1, (b) α-pinene-OS-2, (c) α-pinene-OS-3, (d) α-pinene-OS-4, (e) α-pinene-OS-5, and (f) α-pinene-OS-6. The solubility estimates at 298.15 K, calculated using the 0 H-bond conformers, have previously been reported by Hyttinen et al.[30]

(H-bond acceptors) in addition to the sulfate group. At low temperatures ($T \leq 338.15$ K), using the 0 H-bond conformer set gives a lower solubility estimate than the 1 H-bond conformer set.

COSMO*conf* identified only 15 and 22 unique conformers for α-pinene-OS-1 and 2, respectively (see Table S7 of the Supporting Information). No conformers containing the maximum number of 2 H-bonds were found for either of these two structural isomers. With at most 40 conformers in the COSMO*therm* calculations, all of the conformers of α-pinene-OS-1 and 2 are included. In calculations where all conformers are used (at 298.15 K), the weight of the 0 H-bond conformers is 0.24−0.40 and 0.47−0.81 for α-pinene-OS-1 and 2, respectively, depending on the mixing state (see Figure S11 of the Supporting Information). For comparison, the fraction of 0 H-bond conformers in the full conformer set is 0.47 and 0.50 for α-pinene-OS-1 and 2, respectively. Due to the low number of conformers, differences in the solubilities calculated using the different conformer sets are likely caused both by the difference in the number of conformers in the calculation (the calculation has not converged) and by the effect of including/excluding the energetically more favorable hydrogen bond containing conformers.

On the contrary, none of the 0 H-bond conformers of α-pinene-OS-4 are included in the 40 lowest-energy conformers (due to the flexibility of the molecule and the high number of conformers). This explains the large difference in the solubility of water in the OS, and the different temperature dependence of the aqueous solubility of OS, between calculations using the two different conformer sets (0 H-bonds and 1 H-bond).

Activities for water (a_w) and organosulfates (a_{OS}) were calculated at the stable mixing states of each binary solution. Parts a and b of Figure 8 show water activity in the organic-rich phase of the α-pinene-OS-4−water binary system calculated using the 0 H-bond and 1 H-bond conformer sets, respectively.
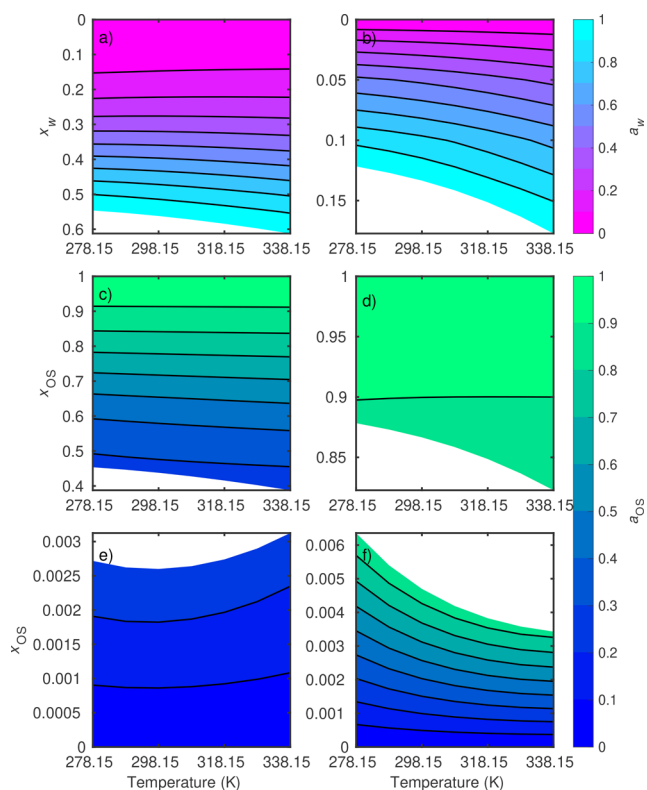
Figure 8. Activities of water (a_w, panels a and b) and α-pinene-OS-4 (a_{OS}, panels c−f) as a function of temperature at different mole fractions of water and OS (x_w and x_{OS}, respectively). Activity coefficients were calculated using the 0 H-bond (a, c, and e) and 2 H-bond (b, d and f) conformer sets. The corresponding activity coefficient values can be found in Table S13 of the Supporting Information.

Similarly, parts c and d of Figure 8 show OS activity in the organic-rich phase, and parts e and f of Figure 8 are for the water-rich phase. The water activity is very close to unity in all stable mixing states of the water-rich phase and is therefore not shown in Figure 8. Water and OS activities of the remaining α-pinene-derived organosulfates are shown in Figures S12a–f–S16a–f (and Tables S8–S12) of the Supporting Information.

There is a clear difference in activities of both water and α-pinene-OS-4 between the different conformer sets in each of the water-rich and organic-rich phases of the binary system. COSMO*therm* estimates much higher water activities using the 0 H-bond conformer set than using the 1 H-bond conformer set. The estimated OS activities in the OS-rich phase are similar when calculated using either of the conformer sets, but due to the large difference in the estimated aqueous solubility in the OS, the activity of the OS at the solubility limit is very different.

There are no large differences in either water or OS activities calculated using the different conformer sets for α-pinene-OS-1, 2, 3, and 6 (see Figures S12–S14 and S16 of the Supporting Information). This is likely due to the 0 H-bond conformers being included in the 40 lowest-energy conformers of the 2 H-bond conformer set. Furthermore, the lowest-energy conformers containing no intramolecular H-bonds are third, fifth, eighth, and fourth lowest-energy conformers in the set of all conformers, respectively. This gives the 0 H-bond conformers higher weights in the COSMO*therm* calculations compared to the lowest-energy 0 H-bond conformer of α-pinene-OS-5. α-Pinene-OS-5 has several 0 H-bond conformers within the 40 lowest-energy conformers but the lowest-energy conformer of the 0 H-bond conformer set has the 18th lowest energy overall. This leads to simlar differences in the water and OS activities calculated using the different conformer sets (see Figure S15 in the Supporting Information) as was seen in α-pinene-OS-4 (Figure 8).

Isoprene-Derived Oxidized Multifunctional Compounds. COSMO*therm* predicts that each of the two isoprene-derived oxidation products ($C_5H_{10}O_6$ and $C_5H_{12}O_6$) are miscible with pure water at temperatures ranging from 278.15 to 338.15 K. As an example, activities of water and $C_5H_{12}O_6$ iso1-S,R in all binary mixing states at 298.15 K and predicted using different sets of conformers in the COSMO*therm* calculations are shown in Figure 9a–d. Activities for the other isoprene-derived compounds are shown in Figures S17a–d–S23a–d of the Supporting Information.

We see that there is very little difference in estimated activities of both organic and water when conformers containing multiple intramolecular hydrogen bonds are included in the COSMO*therm* calculation (Figure 9a,b). Furthermore, activities are close to ideal ($a = x$) in all mixing states. When the number of intramolecular H-bonds is limited to 1 or 0 (Figure 9c,d, respectively), both water and organic activity decrease. This change is consistent with what was seen for the solubility estimates when only conformers containing no intramolecular H-bonds are used, that the mixing of water and the oxidized organic is more favorable.

We also compare activity coefficients (γ^I) of the $C_5H_{10}O_6$ structural isomers (Figure 10a–d) and water (Figure S24a–d) estimated when the different conformer sets of $C_5H_{10}O_6$ are used in the COSMO*therm* calculation. Using all conformers of $C_5H_{10}O_6$ leads to large variations between activity coefficients for the different $C_5H_{10}O_6$ isomers (Figure 10a) and for water in the binary mixtures of the different $C_5H_{10}O_6$ isomers (Figure S24a). Differences between the different structural isomers are
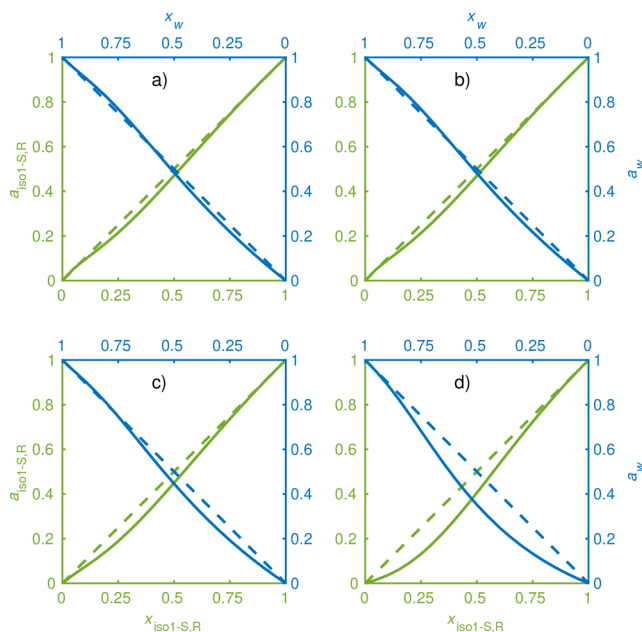
Figure 9. Activities of $C_5H_{12}O_6$ iso1-S,R and water ($a_{iso1-S,R}$ and a_w, respectively) in the binary solutions at 298.15 K, as a function of mole fraction x. The conformer sets of $C_5H_{12}O_6$ used in the COSMO*therm* calculation are (a) 3 H-bonds, (b) 2 H-bonds, (c) 1 H-bond, and (d) 0 H-bonds.
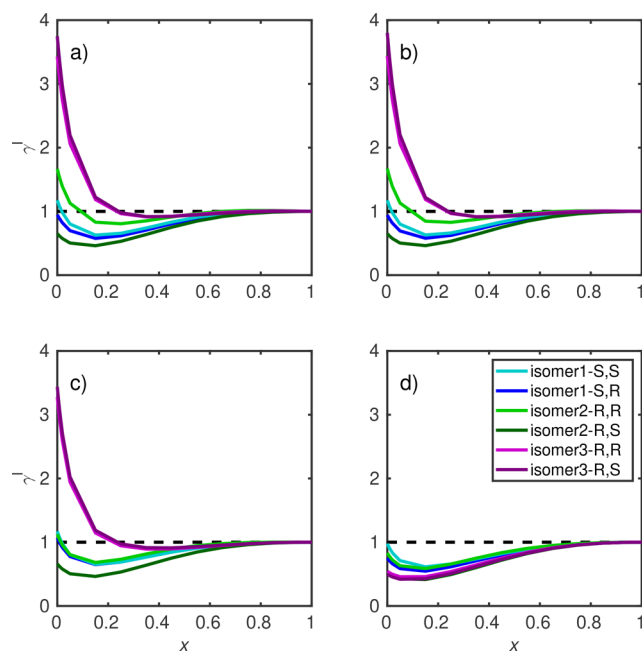
Figure 10. Activity coefficients (γ^I) of $C_5H_{10}O_6$ in water at 298.15 K, as a function of organic mole fraction x. The conformer sets of $C_5H_{10}O_6$ used in the COSMO*therm* calculation are (a) 3 H-bonds, (b) 2 H-bonds, (c) 1 H-bond, and (d) 0 H-bonds.

much smaller when the $C_5H_{10}O_6$ conformers containing intramolecular H-bonds are omitted from the COSMO*therm* calculation (Figures 10d and S24d). Similarly for the $C_5H_{12}O_6$ diastereomers, estimated activity coefficients for water and organic isomers are most similar when the 0 H-bond conformer set is used. The same was seen for saturation vapor pressures of the two iso1 diastereomers by Kurtén et al.[14] However, the ratio of the highest and the lowest saturation vapor pressures of the

$C_5H_{10}O_6$ isomers increased when the number of intramolecular H-bonds was limited in the calculation (from 3.8 calculated using the 2 H-bond conformers to 8.2 using the 0 H-bond conformers).[14]

There is a clear difference in the binary aqueous activity coefficients of the isomer3 diastereomers (Figure S1 of the Supporting Information) calculated using the 0 H-bond conformer set and the other conformer sets. For isomer3, the lowest-energy conformers in the 0 H-bond conformer set are the 158th (R,S) and 184th (R,R) lowest-energy conformers out of all conformers found with systematic conformer search. These conformers are not included in any of the other conformer sets due to their high relative energy. This is likely due to the strong hydrogen bonding between the hydroxyl group and the adjacent carbonyl group (see Figure S1 of the Supporting Information). For the other structural isomers, the lowest-energy conformer of the 0 H-bond conformer set is included in the 40 lowest-energy conformers and thus in all of the conformer sets. This indicates that multifunctional compounds may be described adequately (especially in systems close to the pure compound reference state) if the set of lowest-energy conformers selected to the calculation includes a sufficient number of conformers without any intramolecular H-bonding.

CONCLUSIONS

We have investigated the effect of selecting conformers, on the basis of their intramolecular hydrogen bonding, on the condensed-phase properties of SOA representatives in aqueous solutions calculated using the COSMO*therm* program. There is a general scarcity of experimental data for properties needed to describe formation, transformation, and climate impact of SOA in the atmosphere. The ability to produce reliable estimates from theoretical approaches could therefore be a critical step to increase the accuracy of large-scale predictions of atmospheric chemistry and climate effects of SOA. To benchmark our calculations, we therefore studied citric, tartaric, malic, and maleic acids for which experimental aqueous solubility and activity data are available in the literature, in addition to atmospherically relevant highly oxidized α-pinene-derived organosulfates and isoprene-derived compounds. At the moment, no experimental thermophysical data are available for these atmospheric SOA compounds and these estimates can therefore provide a first insight to their multiphase behavior.

We find that in polar solvents, such as atmospherically ubiquitous water, conformers with H-bond donors that have free hydrogen bonding areas are able to interact with the solvent, which increases the solubility. If conformers containing no intramolecular H-bonds are not included in the COSMO*therm* calculations, or if their weight in the calculation is small, the estimated solubility is significantly lower. For most of the compounds studied here, the estimated solubility therefore increases when only conformers with 0 intramolecular H-bonds are selected. For other solvents, the relevant conformer distribution can be determined by finding conformer weights in each solution using a set of conformers that contain different number of intramolecular H-bonds. For example, in organic nonpolar solvents, H-bond donors of solutes are not able to interact with the solvent and conformers containing multiple intramolecular H-bonds are likely more favorable than conformers containing no intramolecular H-bonds.

The comparison between experimental aqueous solubilities and COSMO*therm* estimates of citric, tartaric, malic, and maleic acids shows better agreement when only conformers containing

no intramolecular hydrogen bonds are used in the COSMO*therm* calculation (on average a factor of 1.2 overestimation as opposed to an underestimation by a factor of 5 using the COSMO*base* conformers). For the α-pinene-derived organosulfates, the estimated solubilities of both the organic compound and water generally increase in a similar fashion when the number of intramolecular H-bonds is limited in the COSMO*therm* calculation. Limiting the number of intramolecular hydrogen bonds in the conformers also reduces differences in estimated activity coefficients for both diastereomers and structural isomers of the same chemical formula. The effect of a favorable positioning of the functional groups, with regard to intramolecular hydrogen bond formation, is neglected when the conformers containing hydrogen bonds are removed from the calculation.

Estimating the saturation vapor pressures of the isoprene-derived $C_5H_{12}O_6$, Kurtén et al.[14] found that the BP/def2-TZVPD level of theory overestimates the strength of intramolecular H-bonds in the gas phase. This causes errors in the saturation vapor pressure estimates, where an accurate gas-phase energy is needed to describe the stability of the compound both in the condensed and in the gas phase. However, using a coupled cluster energy correction to correct the gas-phase energies only decreased the saturation vapor pressure estimate by a factor of 2, while the uncorrected estimate overestimated the experiments by 2 orders of magnitude.[14] Since the gas-phase energy is not needed for estimating condensed-phase properties, improving the gas-phase energy description does not affect properties such as solubility and activity.

Selecting conformers on the basis of the number of intramolecular H-bonds, Kurtén et al.[14] found that decreasing the number of H-bonds by 1 decreases the saturation vapor pressure estimate roughly by a factor of 5. Here, we find that at 298.15 K one intramolecular H-bond (difference between 0 and 1 H-bond) changes the solubility by a factor of 1.8 (citric acid and α-pinene-OS). Although this effect is numerically smaller than what was seen in the saturation vapor pressures of $C_5H_{10}O_6$ and $C_5H_{12}O_6$ by Kurtén et al.,[14] the global impact on SOA formation and climate effects may be equally large, considering the critical role of SOA–water interactions. For activity coefficients, the effect of intramolecular hydrogen bonds is highest for mixing states farthest from the reference state, i.e., at infinite dilution of the organic (in convention I) and at concentrated organic for water activities. At most, activity coefficients decrease by a factor of 6.9 (at infinite dilution), when the number of intramolecular H-bonds is limited from one to zero. On average, the activity coefficients of the organics ($C_5H_{10}O_6$ and $C_5H_{12}O_6$) and water decrease by a factor of 2.9 and 1.4, respectively, switching from the 1 H-bond conformer set to the 0 H-bond conformer set. In general, no significant differences are seen in the corresponding activity coefficients between using the 4, 3, 2, and 1 H-bond conformer sets.

This work shows that COSMO*therm* can give quite reasonable predictions of solubility and activity for highly oxidized atmospheric organics, such as citric, tartaric, malic, and maleic acids, if care is taken to not bias estimates with an overestimation of the effect of intramolecular hydrogen bonding. These results can be applied to other polycarboxylic acids and possibly other atmospherically relevant multifunctional organic compounds, such as peroxides. The importance of conformers containing no intramolecular H-bonds is seen in the weights of conformers in different solutions. For example, the weight of 0 H-bond conformers of α-pinene-OS-1 and 2 is high

in aqueous solutions, similar to results for citric acid. For the highly oxidized isoprene-derived compounds the weight of 0 H-bond conformers is even higher (above 0.77 for iso1-S,R when the conformer set contains 6, 5, and 5 lowest-energy conformers form 3, 1, and 0 H-bond conformer sets, respectively) in the total conformer distribution in all mixing states with water. For other types of compounds, the validity of using only conformers containing no intramolecular H-bonds can be tested by finding which types of conformers have the highest weights in the conformer distribution of the studied solution.

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.0c04285.

Figures S1−S3, isomers of isoprene- and α-pinene-derived multifunctional compounds, and citric acid; Figure S5, aqueous solubilities of tartaric, malic, and maleic acids; Figure S9, water activity coefficients in tartaric, malic, and maleic solutions; Table S4, activity coefficients in aqueous tartaric, malic, and maleic acid solutions; Section S1, additional details on citric acid calculations and results (S1.1 conformers, S1.2 density, S1.3 free energy of fusion estimate, S1.4 pK_a, S1.5 conversion of experimental activity coefficient, S1.6 effect of intramolecular hydrogen bonds); Section S2, additional results of α-pinene-derived OS; Section S3, additional results of isoprene-derived compounds (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Noora Hyttinen − *Nano and Molecular Systems Research Unit, University of Oulu, 90014 Oulu, Finland;* ⓞ orcid.org/0000-0002-6025-5959; Email: noora.hyttinen@oulu.fi

Nønne L. Prisle − *Nano and Molecular Systems Research Unit, University of Oulu, 90014 Oulu, Finland;* Email: nonne.prisle@oulu.fi

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpca.0c04285

Notes

The authors declare no competing financial interest.

.cosmo files of citric, tartaric, malic, and malonic acids that are not available in the COSMO*base* version 17, and of $C_5H_{12}O_6$, are available from the corresponding author upon reasonable request.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Boucher, O.; Randall, D.; Artaxo, P.; Bretherton, C.; Feingold, G.; Forster, P.; Kerminen, V.-M.; Kondo, Y.; Liao, H.; Lohmann, U.; et al. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Doschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. M., Eds.; Cambridge University Press: Cambridge, U.K., 2013; pp 571−657.

(2) Prisle, N. L.; Raatikainen, T.; Laaksonen, A.; Bilde, M. Surfactants in cloud droplet activation: mixed organic-inorganic particles. *Atmos. Chem. Phys.* **2010**, *10*, 5663−5683.

(3) Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(4) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(5) Eckert, F.; Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* **2002**, *48*, 369−385.

(6) Arp, H. P. H.; Goss, K.-U. Ambient gas/particle partitioning. 3. Estimating partition coefficients of apolar, polar, and ionizable organic compounds by their molecular structure. *Environ. Sci. Technol.* **2009**, *43*, 1923−1929.

(7) Schröder, B.; Santos, L. M.; Marrucho, I. M.; Coutinho, J. A. Prediction of aqueous solubilities of solid carboxylic acids with COSMO-RS. *Fluid Phase Equilib.* **2010**, *289*, 140−147.

(8) Wang, C.; Lei, Y. D.; Endo, S.; Wania, F. Measuring and modeling the salting-out effect in ammonium sulfate solutions. *Environ. Sci. Technol.* **2014**, *48*, 13238−13245.

(9) Wang, C.; Goss, K.-U.; Lei, Y. D.; Abbatt, J. P.; Wania, F. Calculating equilibrium phase distribution during the formation of secondary organic aerosol using COSMO*therm*. *Environ. Sci. Technol.* **2015**, *49*, 8585−8594.

(10) Wania, F.; Lei, Y.; Wang, C.; Abbatt, J.; Goss, K.-U. Novel methods for predicting gas−particle partitioning during the formation of secondary organic aerosol. *Atmos. Chem. Phys.* **2014**, *14*, 13189−13204.

(11) Wania, F.; Lei, Y.; Wang, C.; Abbatt, J.; Goss, K.-U. Using the chemical equilibrium partitioning space to explore factors influencing the phase distribution of compounds involved in secondary organic aerosol formation. *Atmos. Chem. Phys.* **2015**, *15*, 3395−3412.

(12) Toivola, M.; Prisle, N. L.; Elm, J.; Waxman, E. M.; Volkamer, R.; Kurtén, T. Can COSMO*therm* predict a salting in effect? *J. Phys. Chem. A* **2017**, *121*, 6288−6295.

(13) Reinisch, J.; Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of the temperature dependence of a polyether−water mixture using COSMO*therm*. *Fluid Phase Equilib.* **2011**, *310*, 7−10.

(14) Kurtén, T.; Hyttinen, N.; D'Ambro, E. L.; Thornton, J.; Prisle, N. L. Estimating the saturation vapor pressures of isoprene oxidation products $C_5H_{12}O_6$ and $C_5H_{10}O_6$ using COSMO-RS. *Atmos. Chem. Phys.* **2018**, *18*, 17589−17600.

(15) Bianchi, F.; Kurtén, T.; Riva, M.; Mohr, C.; Rissanen, M. P.; Roldin, P.; Berndt, T.; Crounse, J. D.; Wennberg, P. O.; Mentel, T. F.; et al. Highly oxygenated organic molecules (HOM) from gas-phase autoxidation involving peroxy radicals: A key contributor to atmospheric aerosol. *Chem. Rev.* **2019**, *119*, 3472−3509.

(16) Dalman, L. H. Ternary systems of urea and acids. IV. Urea, citric acid and water. V. Urea, acetic acid and water. VI. Urea, tartaric acid and water. *J. Am. Chem. Soc.* **1937**, *59*, 775−779.

(17) Dalman, L. H. The solubility of citric and tartaric acids in water. *J. Am. Chem. Soc.* **1937**, *59*, 2547−2549.

(18) Mullin, J. *Crystallization*; Butterworths: London, 1961; p 251.

(19) Apelblat, A.; 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.* **1987**, *19*, 317−320.

(20) Bates, R. G.; Pinching, G. D. Resolution of the dissociation constants of citric acid at 0 to 50°, and determination of certain related thermodynamic functions. *J. Am. Chem. Soc.* **1949**, *71*, 1274−1283.

(21) Levien, B. J. A physicochemical study of aqueous citric acid solutions. *J. Phys. Chem.* **1955**, *59*, 640−644.

(22) Laguerie, C.; Aubry, M.; Couderc, J. P. Some physicochemical data on monohydrate citric acid solutions in water: solubility, density,

viscosity, diffusivity, pH of standard solution, and refractive index. *J. Chem. Eng. Data* **1976**, *21*, 85−87.

(23) Evangelista, R. A.; Chen, F.-T. A.; Guttman, A. Reductive amination of N-linked oligosaccharides using organic acid catalysts. *J. Chromatogr. A* **1996**, *745*, 273−280.

(24) Forster, A.; Hempenstall, J.; Tucker, I.; Rades, T. Selection of excipients for melt extrusion with two poorly water-soluble drugs by solubility parameter calculation and thermal analysis. *Int. J. Pharm.* **2001**, *226*, 147−161.

(25) Apelblat, A.; Manzurola, E. Cryoscopic studies in the citric acid−water system. *J. Mol. Liq.* **2003**, *103*, 201−210.

(26) Schunk, A.; Maurer, G. Activity of water in aqueous solutions of sodium citrate and in aqueous solutions of (an inorganic salt and citric acid) at 298.15 K. *J. Chem. Eng. Data* **2004**, *49*, 944−949.

(27) Kharat, S. J. Density, viscosity, conductivity, ultrasonic velocity, and refractive index studies of aqueous solutions of citric acid at different temperatures. *Int. J. Appl. Chem.* **2008**, *4*, 223−236.

(28) Wyrzykowski, D.; Hebanowska, E.; Nowak-Wiczk, G.; Makowski, M.; Chmurzyński, L. Thermal behaviour of citric acid and isomeric aconitic acids. *J. Therm. Anal. Calorim.* **2011**, *104*, 731−735.

(29) Meltzer, V.; Pincu, E. Thermodynamic study of binary mixture of citric acid and tartaric acid. *Open Chem.* **2012**, *10*, 1584−1589.

(30) Hyttinen, N.; Elm, J.; Malila, J.; Calderón, S. M.; Prisle, N. L. Thermodynamic properties of isoprene- and monoterpene-derived organosulfates estimated with COSMO*therm*. *Atmos. Chem. Phys.* **2020**, *20*, 5679−5696.

(31) *COSMOtherm*, version C3.0, Release 19; COSMO logic GmbH & Co. KG: Leverkusen, Germany, 2019.

(32) *Spartan'14*; Wavefunction Inc.: Irvine, CA, 2014.

(33) *COSMO* conf 4.2; COSMOlogic GmbH & Co. KG: Leverkusen, Germany, 2013.

(34) Kurtén, T.; Hyttinen, N.; D'Ambro, E. L.; Thornton, J.; Prisle, N. L. Supplementary data for the manuscript "Estimating the saturation vapor pressures of isoprene oxidation products $C_5H_{12}O_6$ and $C_5H_{10}O_6$ using COSMO-RS", submitted to Atmospheric Chemistry and Physics (Discussions) (Data set). *Zenodo* **2018**, DOI: 10.5281/zenodo.1344890.

(35) Hyttinen, N.; Elm, J.; Malila, J.; Calderón, S. M.; Prisle, N. L. Supplementary data for the manuscript "Thermodynamic properties of isoprene and monoterpene derived organosulfates estimated with COSMO*therm*" (Version 1) (Data set). *Zenodo* **2019**, DOI: 10.5281/zenodo.3552309.

(36) *COSMObase*, version 17.01; COSMOlogic GmbH & Co. KG: Leverkusen, Germany. 2011.

(37) *COSMO* conf 4.3; COSMOlogic GmbH & Co. KG: Leverkusen, Germany. 2013.

(38) Eckert, F.; Klamt, A. *COSMOtherm Reference Manual*, version C30, Release 19; COSMOlogic GmbH & Co., 2019.

(39) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York and London, 1987.

(40) Levine, I. N. *Physical Chemistry*, 6th ed.; McGraw-Hill: New York, USA, 2009; pp 295−302.

(41) Apelblat, A.; Dov, M.; Wisniak, J.; Zabicky, J. Osmotic and activity coefficients of $HO_2CCH_2C(OH)(CO_2H)CH_2CO_2H$ (citric acid) in concentrated aqueous solutions at temperatures from 298.15 to 318.15 K. *J. Chem. Thermodyn.* **1995**, *27*, 347−353.

(42) Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. *Molecular thermodynamics of fluid-phase equilibria*, 3rd ed.; Pearson Education, 1998; p 510.

(43) Leclercq, M.; Collet, A.; Jacques, J. Etude des melanges d'antipodes optiques—XII: mesure de la stabilite des racemiques vrais. *Tetrahedron* **1976**, *32*, 821−828.

(44) Booth, A. M.; Barley, M. H.; Topping, D. O.; McFiggans, G.; Garforth, A.; 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.* **2010**, *10*, 4879−4892.

(45) Guo, K.; Sadiq, G.; Seaton, C.; Davey, R.; Yin, Q. Co-crystallization in the caffeine/maleic acid system: lessons from phase equilibria. *Cryst. Growth Des.* **2010**, *10*, 268−273.

(46) Rumble, J. R., Ed. *CRC Handbook of Chemistry and Physics*, 99th ed.; CRC Press/Taylor & Francis: Boca Raton, FL, 2018 (Internet Version 2018).

(47) Maffia, M. C.; Meirelles, A. J. A. Water activity and pH in aqueous polycarboxylic acid systems. *J. Chem. Eng. Data* **2001**, *46*, 582−587.

(48) Velezmoro Sanchez, C. E. Modelagem e predição de atividade de agua em fluidos alimenticios. *Ph.D. Thesis*, 1999.

(49) Zardini, A. A.; Sjogren, S.; Marcolli, C.; Krieger, U.; Gysel, M.; Weingartner, E.; Baltensperger, U.; Peter, T. A combined particle trap/HTDMA hygroscopicity study of mixed inorganic/organic aerosol particles. *Atmos. Chem. Phys.* **2008**, *8*, 5589−5601.

(50) Peng, C.; Chan, M. N.; Chan, C. K. The hygroscopic properties of dicarboxylic and multifunctional acids: Measurements and UNIFAC predictions. *Environ. Sci. Technol.* **2001**, *35*, 4495−4501.

(51) Davies, M.; Thomas, D. K. Isopiestic studies of aqueous dicarboxylic acid solutions. *J. Phys. Chem.* **1956**, *60*, 41−44.

(52) Choi, M. Y.; Chan, C. K. Continuous measurements of the water activities of aqueous droplets of water-soluble organic compounds. *J. Phys. Chem. A* **2002**, *106*, 4566−4572.