Article

Assessment of the SM12, SM8, and SMD Solvation Models for Predicting Limiting Activity Coefficients at 298.15 K

Sydnee N. Roese[1], Justin D. Heintz[1], Cole B. Uzat[1], Alexa J. Schmidt[1,2], Griffin V. Margulis[1], Spencer J. Sabatino[1], and Andrew S. Paluch[1,*]

[1] Department of Chemical, Paper, and Biomedical Engineering, Miami University, Oxford, Ohio 45056, USA
[2] Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520, USA
[*] Correspondence: PaluchAS@MiamiOH.edu; Tel.: +1-(513)-529-0784

Abstract: The SMx (x= 12, 8, or D) universal solvent models are implicit solvent models which using electronic structure calculations can compute solvation free energies at 298.15 K. While solvation free energy is an important thermophysical property, within the thermodynamic modeling of phase equilibrium, limiting (or infinite dilution) activity coefficients are preferred since they may be used to parameterize excess Gibbs free energy models to model phase equilibrium. Conveniently, the two quantities are related. Therefore the present study was performed to assess the ability to use the SMx universal solvent models to predict limiting activity coefficients. Two methods of calculating the limiting activity coefficient where compared: 1) The solvation free energy and self-solvation free energy were both predicted and 2) the self-solvation free energy was computed using readily available vapor pressure data. Overall the first method is preferred as it results in a cancellation of errors, specifically for the case in which water is a solute. The SM12 model was compared to both UNIFAC and MOSCED. MOSCED was the highest performer, yet had the smallest available compound inventory. UNIFAC and SM12 exhibited comparable performance. Therefore further exploration and research should be conducted into the viability of using the SMx models for phase equilibrium calculations.

1. Introduction

Limiting (or infinite dilution) activity coefficients are of particular interest both in their fundamental and practical applications. As will be shown, the log limiting activity coefficient for solute i infinitely dilute in solvent j is proportional to the transfer free energy of i from a solution of pure i to pure j (in which it is infinitely dilute). Therefore fundamentally, the limiting activity coefficient sheds insight into how favorable the interactions of i with j are relative to i with itself. This fundamental molecular-level understanding provides the framework that directly impacts design schemes for chemical processes [1].

The practical applications are vast for they are used in the design of separation equipment, solvent selection, the design of pharmaceuticals, stripping operations, leaching operations, and the extraction of dilute materials [2,3]. Previous studies have also shown that limiting activity coefficients can be used to predict the formation of an azeotrope [4–8]. The equilibrium composition of a dilute mixture can also be determined, thus having industrial applications in azeotropic distillation and high-purity extraction [9]. This can serve as a means for process intensification in the environmental sphere, since environmental engineering addresses the implications of a chemical very dilute in a phase, often requiring dilute separation processes [10]. For example, limiting activity coefficients can be used to

determine partition coefficients for a dilute species between two phases, which is critical in predicting the bio-accumulation of long-lived chemicals in the environment and food chain [2].

Of particular interest, it has been demonstrated that the limiting activity coefficient of each component in a binary mixture can be used to parameterize a binary interaction excess Gibbs free energy model to make phase equilibrium calculations over an entire range of concentrations [8,11–13]. These parameterizations are particularly advantageous for maximizing the efficiency of separation processes, which comprise a large majority of chemical plant operations and costs. The underlying principle is that limiting activity coefficients are a direct measure of the interactions in solution, and correspond to maximum deviations from ideality. It is therefore reasonable to assume we can extrapolate to more concentrated solutions where the solution would be more ideal (i.e., closer to the pure component limit). Consider for example the use of Wilson's equation [13]. For a binary system we can compute the composition dependent activity coefficient of component 1 and 2, at mole fraction composition x_1 and $x_2 = 1 - x_1$ as:

$$\ln \gamma_1 = -\ln (x_1 + \Lambda_{12} x_2) + x_2 \left(\frac{\Lambda_{12}}{x_1 + \Lambda_{12} x_2} - \frac{\Lambda_{21}}{\Lambda_{21} x_1 + x_2} \right) \tag{1}$$
$$\ln \gamma_2 = -\ln (x_2 + \Lambda_{21} x_1) - x_1 \left(\frac{\Lambda_{12}}{x_1 + \Lambda_{12} x_2} - \frac{\Lambda_{21}}{\Lambda_{21} x_1 + x_2} \right)$$

where Λ_{12} and Λ_{21} are adjustable parameters which may be related to the binary (intermolecular) interaction parameters (BIPs) of the system (a_{12} and a_{21}):

$$\Lambda_{12} = \frac{v_2}{v_1} \exp \left[-\frac{a_{12}}{RT} \right] \tag{2}$$
$$\Lambda_{21} = \frac{v_1}{v_2} \exp \left[-\frac{a_{21}}{RT} \right]$$

where R is the molar gas constant, T is the absolute temperature, and v_1 and v_2 are the molar volume of pure component 1 and 2, respectively. At infinite dilution, eq. (1) reduces to:

$$\ln \gamma_1^\infty = -\ln (\Lambda_{12}) + 1 - \Lambda_{21} \tag{3}$$
$$\ln \gamma_2^\infty = -\ln (\Lambda_{21}) + 1 - \Lambda_{12}$$

It follows that with knowledge of the limiting activity coefficient of each component in the binary mixture, eq. (3) reduces to a system of two equations with two unknowns that can be used to parameterize Wilson's equation. This would allow one to calculate composition dependent activity coefficients which could be used, for example, to model vapor/liquid equilibrium. Moreover, this approach may readily be generalized to multi-component systems. Wilson's equation has two binary parameters for each pair in a multi-components system, where limiting activity coefficients may be used to parameterize each binary pair [13].

Limiting activity coefficients can be experimentally determined through a variety of methods including: gas chromatography, gas stripping, ebulliometry, and the dew point method [9,14]. However when there is little or no experimental data present, predictive models and tools can be implemented. Some popular methods include group contributions methods, solubility parameter based methods, molecular simulations and electronic structure calculations. The Universal Quasichemical Functional-group Activity Coefficients (UNIFAC) method is a group contribution-based binary interaction excess Gibbs free energy model designed by breaking molecules into functional groups and calculating the configuration energy of the molecule as the sum of group interaction energies [15,16]. Similarly, mod-UNIFAC (Dortmund) is a revision of the original UNIFAC model that

accounts for the temperature dependence of activity coefficients by including temperature-dependent interaction parameters within the model [17,18]. Solubility parameter based methods such as the modified separation of cohesive energy density (MOSCED) [8,19,20] and Hansen solubility parameter (HSP) methods [21], while they limit the user to modeling systems for which solute parameters are available, they give molecular-level insight into the the behavior of a binary system. Molecular simulations and electronic structure calculations can be computationally rigorous, yet are advantageous for they only require the input of a molecular structure [22–26].

The pursuit of accurate, efficient predictive models is vital as it could reduce the experimental and design cost associated with chemical processes. This is of particular interest in separation processes, as they may account for 40-70 percent of the capital and operating cost of a chemical plant [27]. Predictive computational tools have become an essential asset in conducting feasibility and comparative studies, as they are significantly quicker and more cost-efficient than experimental determinations [28]. They are necessary when property data is unavailable, specifically in the case of dealing with new products and compounds. However within computational advancement, there is a risk in losing the intuitive molecular understanding of the processes that lead to observed outcomes. Thus, including thermodynamic parameters such as limiting activity coefficients in the computational process provides molecular insight that could greatly assist in resultant recommendations for industrial design [10].

This study seeks to assess the capability of the SMx (x= 12, 8, or D) universal solvent models for predicting limiting activity coefficients [29–31]. The SMx universal solvent models are implicit or continuum solvent models that can compute solvation free energies at 298.15 K using electronic structure calculations. The models have long been under active development with their accuracy continuously improving and are implemented in a number of software packages [29,32–34]. The only input needed from the user is a structure of the molecule; therefore the SMx models are set apart from other methods in the sheer simplicity for the user. Of specific interest to us is their potential future use to parameterize binary interaction excess Gibbs free energy models for phase equilibrium predictions. In addition, a major goal of this work is to determine how best to predict limiting activity coefficients using solvation free energies computed using the SMx universal solvent models. Specifically, as will be shown, the prediction of the limiting activity coefficient is dependent on both the solvation free energy and self-solvation free energy. While the self-solvation free energy may be computed directly, it may additionally be computed from readily available vapor pressure data.

Previous work to use the SMx universal solvent models to predict limiting activity coefficients is limited. Recently, Lisboa and Pliego [26] used SMD to compute 46 limiting activity coefficients. In the present study, we compute over 3000 limiting activity coefficients with comparison to reference data, for each SM12, SM8 and SMD. In addition, we compare two methods of calculating the limiting activity coefficient: 1) The solvation free energy and self-solvation free energy were both predicted and 2) the self-solvation free energy was computed using readily available vapor pressure data. To better understand the two methods, we compare the individual contributions of the solvation and self-solvation free energy to available reference data. In general, there is not a significant difference between the limiting activity coefficients computed using the two methods. However, for the cases in which water is a solute, the results are significantly better when both the solvation and self-solvation free energy are predicted using SMx. This results from a cancellation of errors in the two terms. We therefore recommend that in general SMx be used to predict both the solvation and self-solvation free energy when computing limiting activity coefficients. Comparison is additionally made to predictions using UNIFAC and MOSCED. SM12 and UNIFAC are found to exhibit comparable performance, motivating future exploration and research into the viability of using the SMx models for phase equilibrium calculations.

2. Method

Performing electronic structure calculations in the SMx ($x=$ 12, 8, or D) universal solvent model, one calculates the solvation free energy of a solute (component i) at infinite dilution in a solvent (component j), $\Delta G_{i,j}^{\text{solv}}$, where $i =\{1$ or $2\}$ and $j =\{1$ or $2\}$ [29–31,35,36]. The solvation free energy in this context is defined as taking a solute from a non-interacting ideal gas state to solution at the same molecular density (or concentration). We have shown previously that the solvation free energy is readily related to the limiting activity coefficient ($\gamma_{i,j}^{\infty}$) as [24,37–42]:

$$\ln \gamma_{i,j}^{\infty} (T,P) = \frac{1}{RT} \left[\Delta G_{i,j}^{\text{solv}} (T,P) - \Delta G_{i,i}^{\text{self}} (T,P) \right] + \ln \frac{v_i (T,P)}{v_j (T,P)} \tag{4}$$

where $\Delta G_{i,i}^{\text{self}}$ is the solute "self"-solvation free energy, R is the molar gas constant, and v_i and v_j correspond to the pure liquid molar volume of component i and j, respectively. The difference $\Delta G_{i,j}^{\text{solv}} - \Delta G_{i,i}^{\text{self}}$ is equivalent to the transfer free energy of i from a solution of pure i to pure j (in which it is infinitely dilute). When $i = j$ we obtain the correct limiting behavior that $\gamma_{i,j}^{\infty} = 1$. The self-solvation free energy, $\Delta G_{i,i}^{\text{self}}$, may readily be calculated by performing a solvation free energy calculation for component i in itself. Alternatively, we can relate $\Delta G_{i,i}^{\text{self}}$ to the pure liquid fugacity of component i, f_i^0, as [38]:

$$\ln f_i^0 (T,P) = \frac{1}{RT} \Delta G_{i,i}^{\text{self}} (T,P) + \ln \frac{RT}{v_i (T,P)} \tag{5}$$

We can expand f_i^0 as [13]:

$$f_i^0 (T,P) = \phi_i^{\text{sat}} (T) P_i^{\text{sat}} (T) \exp \left\{ \int_{P^{\text{sat}}}^{P} \frac{v_i (T,P)}{RT} dP \right\} \tag{6}$$

where ϕ_i^{sat} and P_i^{sat} are the fugacity coefficient and vapor pressure of pure component i at saturation at T, and the term in brackets is the Poynting correction, and accounts for the change in fugacity in going from P^{sat} to P at constant T. If we assume that the vapor phase is an ideal gas and that the Poynting correction is negligible, we have [13,38,43,44]:

$$f_i^0 (T,P) = P_i^{\text{sat}} (T) \tag{7}$$

If we are at low pressures well removed from the critical point, and we have a non-self associating fluid (i.e., no carboxylic acid), use of this expression is generally reasonable. Often times, the pure component property P_i^{sat} is known. This presents an opportunity to use reference data to compute $\Delta G_{i,i}^{\text{self}}$ as:

$$\frac{1}{RT} \Delta G_{i,i}^{\text{self}} (T,P) = \ln \frac{P_i^{\text{sat}} (T) v_i (T,P)}{RT} \tag{8}$$

and then using electronic structure calculations in the SMx universal solvent models to compute $\Delta G_{i,j}^{\text{solv}}$, which is a property of the mixture.

3. Computational Details

All solvation free energy calculations were performed in the SMx ($x=$ 12, 8, or D) universal solvent models using QChem 5.1.2 [45,46]. We selected systems by searching through the Minnesota Solvation Database version 2012 [47] for binary systems for which QChem had (SMx) solvent parameters for both components. The Minnesota Solvation Database was selected as our reference for comparison as it was used in the parameterization of SM12 [29], and earlier versions were used in the parameterization of SM8 and SMD [30,31]. The data in the Minnesota Solvation Database is restricted to 298.15 K, and it follows that calculations using the SMx solvent models are restricted to 298.15 K. Also, the Minnesota Solvation Database contains only solvation and transfer free energy data. We computed reference

limiting activity coefficients from the solvation free energies using eq. (4), where the self solvation free energy was computed via eq. (8) using reference vapor pressure and molar volume data [48]. Note that iodine and carboxylic acid containing molecules were excluded from our search. We excluded the seven iodine hydrocarbons because SM8 and SMD were not parameterized to treat iodine containing solutes [30,31]. The six molecules containing a carboxylic acid functional group were excluded as they are known to dimerize in the vapor phase [49], and we wish to compare the use of eq. (5) and eq. (7) when computing limiting activity coefficients. Excluding these 13 molecules, QChem contains SMx parameters for 148 additional molecules (solvents).

This resulted in a data set of solvation free energies (and hence limiting activities) for 1140 systems, consisting of 107 unique solutes and 73 unique solvents. The Minnesota Solvation Database contained self-solvation free energies at 298.15 K for only 62 of the 148 possible molecules. We therefore instead computed a set of reference self-solvation free energies for all 148 molecules using eq. (8) with reference vapor pressure and molar volume data [48]. For the 62 molecules for which values were available in the Minnesota Solvation Database, the computed values were found to be consistent. The inclusion of both solvation free energies and self-solvation free energies will allow us to look at the individual contributions to the limiting activity coefficient.

As already mentioned, the Minnesota Solvation Database was used in the parameterization of the SMx universal solvent models. Its use was motivated by the desire to assess the prediction of the individual contributions, $\Delta G_{i,j}^{\text{solv}}$ and $\Delta G_{i,i}^{\text{self}}$, and the combined effect on $\ln \gamma_{i,j}^{\infty}$. As a secondary test, we curated reference data from Parts 1 to 6 of DECHEMA's "Activity Coefficients at Infinite Dilution" volumes [50–55]. We again restricted ourselves to binary systems for which both components were in our set of 148 molecules (solvents) for which QChem contains SMx parameters. This resulted in limiting activity coefficients for an additional 2163 systems, consisting of 117 unique solutes and 77 unique solvents. Note that as compared to the Minnesota Solvation Database wherein there was only a single entry for each unique system (i.e., solute-solvent pair), duplicates were present in DECHEMA. We retained all of the data, including duplicates, in the final set.

When curating reference data, we only included binary systems for which QChem had (SMx) solvent parameters for both components. The reason for this is two-fold. (1) First, if one were to compute both $\Delta G_{i,j}^{\text{solv}}$ and $\Delta G_{i,i}^{\text{self}}$ to predict $\ln \gamma_{i,j}^{\infty}$ via eq. (4), a solvation free energy would be required when both i and j take on the role of solvent, requiring that SMx solvent parameters be available for both components. (2) On the other hand, if $\Delta G_{i,i}^{\text{self}}$ was instead computed using reference vapor pressure data, then in theory solvent parameters for i would not be necessary. However, even if this is the case, we are interested in the potential future application of using SMx computed solvation free energies to parameterize a binary interaction excess Gibbs free energy model for phase equilibrium calculations. For this case, for a binary pair both $\gamma_{i,j}^{\infty}$ and $\gamma_{j,i}^{\infty}$ are required; no matter how the self-solvation free energy is computed, a solvation free energy calculation is required in both solvents, making necessary the availability of SMx solvent parameters for both components.

As we will discuss, we find that challenges occur in predicting limiting activity coefficients for aqueous systems. We therefore sourced an additional data set composed of organics in water and water in organics at 298.15 K from "Yaws' Handbook of Properties for Aqueous Systems" [56]. Data was available for 121 systems where water is the solvent and 90 systems where water is the solute, with no duplicates.

The calculations proceeded by first obtaining a daylight SMILES [57,58] for each of the 148 molecules from Pubchem [59]. We then generated 3-D structures from the SMILES using Open Babel 2.3.2 [60,61]. Open Babel was then used to perform a systematic conformational search to identify the lowest energy conformer, followed by geometry optimization. The conformational search and geometry optimization were conducted using the General Amber Force Field [62,63] with Gasteiger partial charges [64]. The final structures were then further refined by performing a geometry optimization in vacuum using QChem 5.1.2 at the M06-2X/cc-pVDZ level of theory/basis set [35,65,66]. This final structure was then used for all of the solvation free energy calculations.

To calculate the solvation (including self-solvation) free energies, single point energy calculations were performed in vacuum and in the SM12, SM8 and SMD universal solvent models for each solvent of interest at the M06-2X/6-31G(d) level of theory/basis set using QChem 5.1.2. Using QChem 5.1.2, the calculation of the solvation free energy is handled completely internally by the software. We emphasize that we used the same conformation for the solute in the ideal gas (vacuum) and solution phases, consistent with the SM12, SM8, and SMD parameterizations [29–31]. It is possible to consider different configurations for the two phases, and it is possible to incorporate multiple configurations into the calculation of the solvation free energy which may be important for some systems [67]. However, this was not pursued here for simplicity.

All of the SMx computed self-solvation free energies ($\Delta G_{i,i}^{\text{self}}$), solvation free energies ($\Delta G_{i,j}^{\text{solv}}$), and limiting activity coefficients ($\gamma_{i,j}^{\infty}$), along with the reference values and results of the reference calculations (UNIFAC and MOSCED) to which comparison is made, are tabulated in the supporting information accompanying the electronic version of this manuscript. As mentioned, we are interested in the potential future application of using SMx computed solvation free energies to parameterize a binary interaction excess Gibbs free energy model for phase equilibrium calculations. Therefore, in the supporting information the results are additionally categorized following the system classification used in DECHEMA's "Vapor-Liquid Equilibrium Data Collection" volumes [68].

4. Results and Discussion

The central goal of this work is to assess the ability of electronic structure calculations with the SMx (x= 12, 8, or D) universal solvent models to predict limiting activity coefficients. As seen in eq. (4), this is composed of contributions due to the solvation free energy ($\Delta G_{i,j}^{\text{solv}}$) and self solvation free energy ($\Delta G_{i,i}^{\text{self}}$). We will therefore first assess each individual contribution. We will quantify the agreement between the reference data and the SMx predictions using the root mean square error (*RMSE*):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(Y^{\text{pred}} - Y^{\text{ref}}\right)^2} \tag{9}$$

the average absolute percent deviation (*AAPD*):

$$AAPD = \frac{100}{N} \sum_{i=1}^{N} \left| \frac{Y^{\text{pred}} - Y^{\text{ref}}}{Y^{\text{ref}}} \right| \tag{10}$$

and the mean unsigned error (*MUE*):

$$MUE = \frac{1}{N} \sum_{i=1}^{N} \left| Y^{\text{pred}} - Y^{\text{ref}} \right| \tag{11}$$

where the superscript "pred" and "ref" correspond to the predicted and reference values, respectively, Y corresponds to the property of interest, and N corresponds to the number of systems for which comparison is being made.

4.1. Self-Solvation Free Energy

We begin with $\Delta G_{i,i}^{\text{self}}$. A parity plot of the predicted versus reference values of $\Delta G_{i,i}^{\text{self}} / (RT)$ and a summary of the computed statistics is provided in fig. 1 and table 1, respectively. The statistics in table 1 are computed for $Y = \Delta G_{i,i}^{\text{self}} / (RT)$, with the summation carried out over all $N = 148$ molecules for which reference data is available. Note that we are interested in the dimensionless $\Delta G_{i,i}^{\text{self}}$ as it is the quantity of interest in eq. (4).
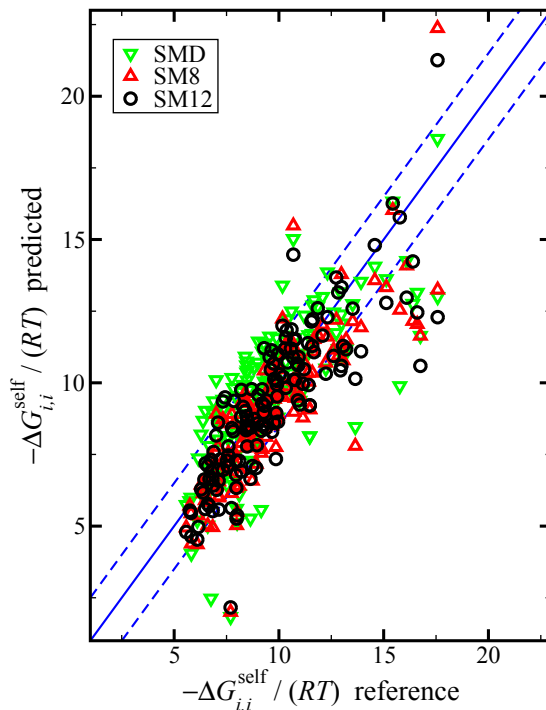
Figure 1. Parity plot of the predicted versus reference values of $-\Delta G_{i,i}^{\text{self}} / (RT)$ using the SM12, SM8, and SMD universal solvent model, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.5 and are drawn as a reference; the value of 1.5 was used based on the *RMSE* for the SM12 predictions.

Table 1. A summary of the squared correlation coefficient (R^2), root mean square error (*RMSE*), average absolute percent deviation (*AAPD*), and mean unsigned error (*MUE*) for the values of $\Delta G_{i,i}^{\text{self}} / (RT)$ predicted using the SM12, SM8, and SMD universal solvent model as compared to the reference data.

	$\Delta G_{i,i}^{\text{self}} / (RT)$		
	SM12	SM8	SMD
R^2	0.713	0.707	0.566
RMSE	1.526	1.600	1.820
AAPD	10.888	11.230	14.655
MUE	1.077	1.131	1.372

Overall, we find that SM12 and SM8 outperform SMD, with SM12 performing slightly better than SM8. The top performance of SM12 is in agreement with the recent SM12 parameterization [29]. Visually, it is clear in fig. 1 that SM12 and SM8 offer a noticeably smaller spread in the data as compared to SMD, corresponding to a smaller overall error.

During the parameterization of SM12, it was pointed out that the largest deviation was observed for the self-solvation free energy of water, which was overly negative by 1 to 2 kcal/mol (or 1.7 to 3.4 *RT* in dimensionless units) using all methods, and the solvation free energy of the water dimer in water was overly negative by 3 to 4 kcal/mol (or 5.1 to 6.8 *RT* in dimensionless units) using all methods [29]. We likewise observe large errors for the self-solvation free energy of water, although our predictions are slightly more negative. The reference value of $\Delta G_{i,i}^{\text{self}} / (RT)$ for water is –10.689, while we predict –14.473, –15.485, and –15.024 with SM12, SM8, and SMD, respectively. This would correspond to predicting a vapor pressure that is too small, and that we are over estimating the affinity of water for the liquid phase or equivalently the strength of the water-water interactions.

4.2. Solvation Free Energy

Next we consider $\Delta G_{i,j}^{\mathrm{solv}}$, making comparison for the 1140 systems, consisting of 107 unique solutes and 73 unique solvents, from the Minnesota Solvation Database. A parity plot of the predicted versus reference values of $\Delta G_{i,j}^{\mathrm{solv}} / (RT)$ and a summary of the computed statistics is provided in fig. 2 and table 2. The statistics in table 2 are computed for $Y = \Delta G_{i,j}^{\mathrm{solv}} / (RT)$, with the summation carried out over all $N = 1140$ systems for which reference data is available.
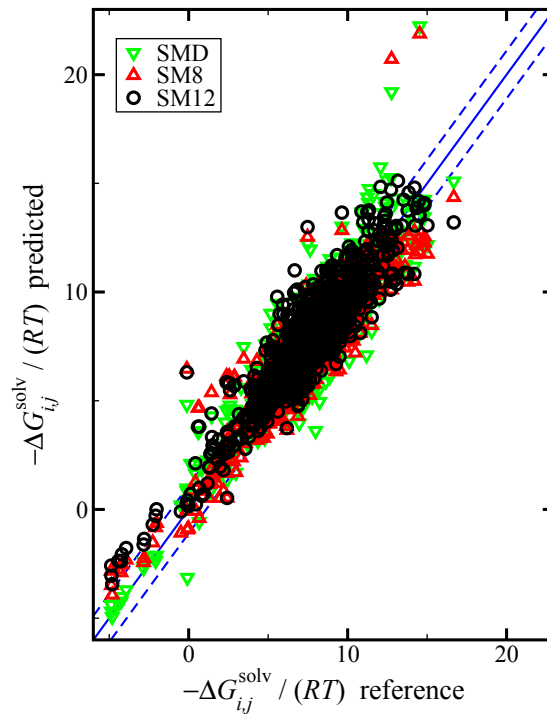
Figure 2. Parity plot of the predicted versus reference values of $-\Delta G_{i,j}^{\mathrm{solv}} / (RT)$ using the SM12, SM8, and SMD universal solvent model, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference; the value of 1.1 was used based on the *RMSE* for the SM12 predictions. The reference values are from the Minnesota Solvation Database.

Table 2. A summary of the squared correlation coefficient (R^2), root mean square error (*RMSE*), average absolute percent deviation (*AAPD*), and mean unsigned error (*MUE*) for the values of $\Delta G_{i,j}^{\mathrm{solv}} / (RT)$ predicted using the SM12, SM8, and SMD universal solvent model as compared to the reference data from the Minnesota Solvation Database.
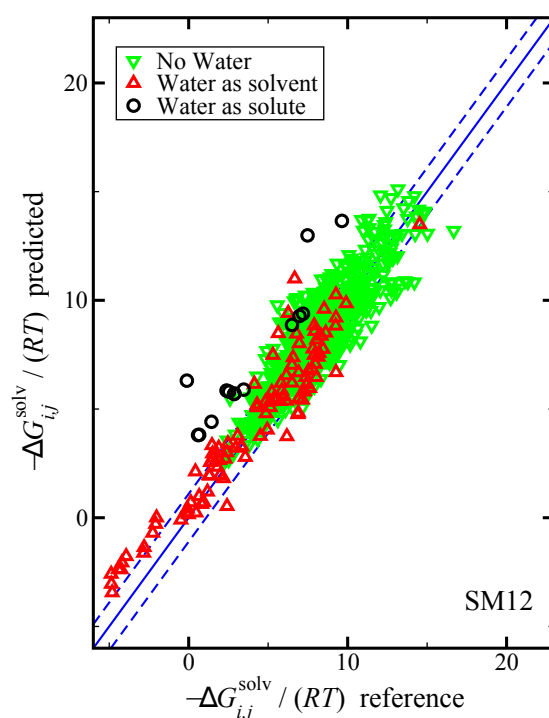
	$\Delta G_{i,j}^{\mathrm{solv}} / (RT)$		
	SM12	SM8	SMD
R^2	0.838	0.826	0.820
RMSE	1.120	1.125	1.181
AAPD	20.985	24.593	26.093
MUE	0.828	0.843	0.882

We again find that SM12 is the top performing model, although the difference in performance between the models is much smaller for $\Delta G_{i,j}^{\mathrm{solv}}$ as compared to $\Delta G_{i,i}^{\mathrm{self}}$. Moreover, we find that except for the *AAPD*, the overall errors are all improved for $\Delta G_{i,j}^{\mathrm{solv}}$ relative to $\Delta G_{i,i}^{\mathrm{self}}$. Comparing figs. 1 and 2, there appears to be a smaller spread in the $\Delta G_{i,j}^{\mathrm{solv}}$ data, which would correspond to the smaller *RMSE* and *MUE*. This at first had us surprised. In theory, $\Delta G_{i,i}^{\mathrm{self}}$ corresponds to a single component, homogeneous system. On the other hand, $\Delta G_{i,j}^{\mathrm{solv}}$ corresponds to a single solute infinitely dilute in a

solvent. In this case we need to account for solute-solvent and solvent-solvent interactions which are not the same. We would therefore assume that the latter case is more difficult to predict and would result in a lower accuracy. However, this was not the case. We do not offer an explanation for this difference, but would like to point out two differences in the reference data sets. 1) Our data set for $\Delta G_{i,j}^{\text{solv}}$ consists of 1140 systems, while we have just 148 values of $\Delta G_{i,i}^{\text{self}}$. And of the 148 values of $\Delta G_{i,i}^{\text{self}}$, only 62 were included in the Minnesota Solvation Database. 2) Comparing figs. 1 and 2, we see there is a difference in the range of values in each data set. While a majority of the values of $\Delta G_{i,i}^{\text{self}} / (RT)$ are over the range –15 to –5, there is an appreciable number of values of $\Delta G_{i,j}^{\text{solv}} / (RT)$ greater than –5. We also point out that this last point contributes to the larger $AAPD$ for $\Delta G_{i,j}^{\text{solv}}$. The computed errors and top performance of SM12 are consistent with those published in the recent SM12 parameterization [29].

Considering again the original SM12 parameterization, it was found that the errors for the solvation free energy (both aqueous and non-aqueous) for the solute class containing molecular hydrogen, ammonia, water, and the water dimer were relatively large compared to the other solute classes [29]. As already mentioned, the original SM12 parameterization pointed out a particular challenge with the self-solvation free energy of water which was predicted to be overly negative. To explore this further, we broke down $\Delta G_{i,j}^{\text{solv}}$ to cases where water was the solvent ($j =$water), water was the solute ($i =$water), and when neither the solvent or solute was water. Note that $\Delta G_{i,i}^{\text{self}}$ was not include in the set of $\Delta G_{i,j}^{\text{solv}}$. This resulted in 104 systems wherein water was the solvent, 14 systems wherein water was the solute, and 1022 systems wherein water was neither the solvent or solute. A parity plot of the predicted versus reference values of $\Delta G_{i,j}^{\text{solv}} / (RT)$ with SM12 wherein these three cases are differentiated is provided in fig. 3.

Figure 3. Parity plot of the predicted versus reference values of $-\Delta G_{i,j}^{\text{solv}} / (RT)$ using the SM12 universal solvent model for the case where water was the solvent, water was the solute, and where water was neither the solvent or solute, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference; the value of 1.1 was used based on the *RMSE* for the SM12 predictions. The reference values are from the Minnesota Solvation Database.

We find that in general when water is the solute, $\Delta G_{i,j}^{\text{solv}}$ is predicted to be overly negative. We find that $\Delta G_{i,j}^{\text{solv}} / (RT)$ is overly negative on average by 3.4, 3.5, and 2.4 (dimensionless units) when using SM12, SM8, and SMD, respectively. Compare this to the case of $\Delta G_{i,i}^{\text{self}} / (RT)$ for water predicted to be overly negative by 3.8, 4.8, and 4.3 (dimensionless units) for SM12, SM8, and SMD, respectively.

4.3. Limiting Activity Coefficients

When predicting limiting activity coefficients ($\gamma_{i,j}^{\infty}$) with the SMx (x= 12, 8, or D) universal solvent models, two different strategies were evaluated. 1) We predicted both $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$, then using reference values of the molar volume [48], $\gamma_{i,j}^{\infty}$ was computed via eq. (4). 2) In the second approach, since the pure component vapor pressure (P_i^{sat}) is commonly available [48], we used this reference data to compute $\Delta G_{i,i}^{\text{self}}$ via eq. (8). We then combined this reference value of $\Delta G_{i,i}^{\text{self}}$ with our SMx predicted value of $\Delta G_{i,j}^{\text{solv}}$ to predict $\gamma_{i,j}^{\infty}$. When computing statistics, for *AAPD* we use $Y = \gamma_{i,j}^{\infty}$. For all other cases ($R^2$, *RMSE*, and *MUE*) we use $Y = \ln \gamma_{i,j}^{\infty}$.

4.3.1. Minnesota Solvation Database

Limiting activity coefficients ($\gamma_{i,j}^{\infty}$) were computed for the same 1140 systems from the Minnesota Solvation Database for which $\Delta G_{i,j}^{\text{solv}}$ was computed. Reference values of $\gamma_{i,j}^{\infty}$ were computed via eq. (4) using the reference values of $\Delta G_{i,j}^{\text{solv}}$ from the Minnesota Solvation Database, the reference values of $\Delta G_{i,i}^{\text{self}}$ already described, and reference values of the molar volume [48]. A parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^{\infty}$ is provided in fig. 4. In the top pane results are shown when we predict both $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$, and the bottom pane shows the case when $\Delta G_{i,i}^{\text{self}}$ was computed using reference data.
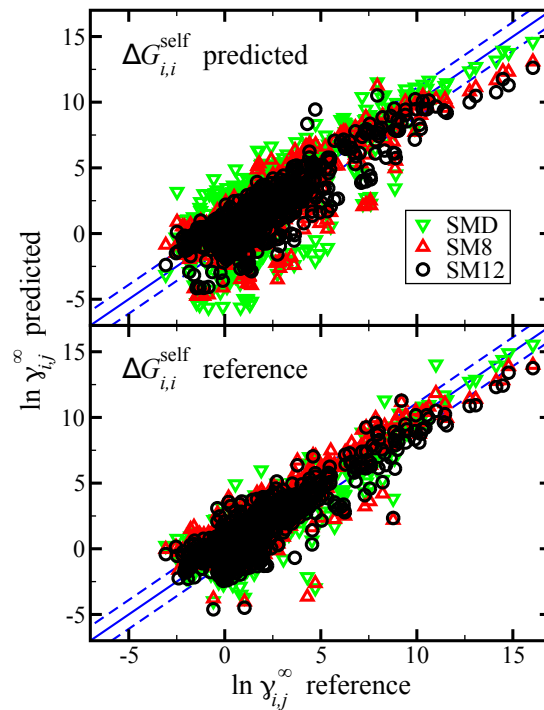
Figure 4. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^{\infty}$ from the Minnesota Solvation Database version 2012 [47]. Predictions are made using the SM12, SM8, and SMD universal solvent model, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference; the value of 1.1 was used based on the *RMSE* for the SM12 predictions. In the top pane we present results when $\Delta G_{i,i}^{\text{self}}$ was predicted using the SM12, SM8, or SMD universal solvent model. In the bottom pane we present results when $\Delta G_{i,i}^{\text{self}}$ was computed using reference vapor pressure data.

In general, there does not appear to be a significant difference between the predictions wherein $\Delta G_{i,i}^{\text{self}}$ was predicted or computed using reference data. The spread in results does appear to decrease when $\Delta G_{i,i}^{\text{self}}$ is computed using reference data, which may reflect the spread in the predicted values of $\Delta G_{i,i}^{\text{self}}$ (see fig. 1). A comparison of the statistics is provided in table 3.

Table 3. A summary of the squared correlation coefficient (R^2), root mean square error (*RMSE*), and mean unsigned error (*MUE*) for $\ln \gamma_{i,j}^{\infty}$, and the average absolute percent deviation (*AAPD*) for $\gamma_{i,j}^{\infty}$ predicted using the SM12, SM8, and SMD universal solvent model as compared to the reference data from the Minnesota Solvation Database. For the predictions, we compare the case where 1) $\Delta G_{i,i}^{\text{self}}$ is predicted using the SM12, SM8, and SMD universal solvent model, and where 2) $\Delta G_{i,i}^{\text{self}}$ is computed using reference data, as indicated.

	$\Delta G_{i,i}^{\text{self}}$ predicted			$\Delta G_{i,i}^{\text{self}}$ reference		
	SM12	SM8	SMD	SM12	SM8	SMD
R^2	0.796	0.740	0.627	0.806	0.796	0.788
RMSE	1.155	1.288	1.801	1.120	1.125	1.181
AAPD	89.856	107.821	357.706	92.482	134.797	131.626
MUE	0.840	0.883	1.376	0.828	0.842	0.882

Again, while the difference in the two sets of predictions does not appear to be large, in all cases we find that R^2 increases and the computed overall errors decrease when using reference values of $\Delta G_{i,i}^{\text{self}}$. One might expect this result. Consider eq. (4) and the errors summarized in tables 1 to 3, where we will take *RMSE* as a measure of the error in the predictions. When using reference values of $\Delta G_{i,i}^{\text{self}}$,

the only predicted quantity in eq. (4) is $\Delta G_{i,j}^{\text{solv}}$. As a result, we find for this case that the *RMSE* for the predicted values of $\ln \gamma_{i,j}^{\infty}$ is identical to that for $\Delta G_{i,j}^{\text{solv}} / (RT)$. Next, consider the case of when $\Delta G_{i,i}^{\text{self}}$ is predicted. Using propagation of error, we would expect the error (taken to be *RMSE*) in $\ln \gamma_{i,j}^{\infty}$ to be equal to $\sqrt{RMSE_{\text{solv}}^2 + RMSE_{\text{self}}^2}$, where $RMSE_{\text{solv}}$ and $RMSE_{\text{self}}$ are the *RMSE* for the predicted values of $\Delta G_{i,j}^{\text{solv}} / (RT)$ and $\Delta G_{i,i}^{\text{self}} / (RT)$, respectively. This would result in an *RMSE* of 1.893, 1.970, and 2.170 for SM12, SM8 and SMD, respectively. However, we find that the actual error is less than the predicted error. The expression used assumes that the error is random [69]. The smaller observed uncertainty suggests the presence of systematic error.

With this in mind, let us again look at the specific case of water. We found earlier that when water is the solute, on average $\Delta G_{i,j}^{\text{solv}} / (RT)$ is predicted to be overly negative by 3.4, 3.5, and 2.4 (dimensionless units) when using SM12, SM8, and SMD, respectively. Further, $\Delta G_{i,i}^{\text{self}} / (RT)$ for water is predicted to be overly negative by 3.8, 4.8, and 4.3 (dimensionless units) for SM12, SM8, and SMD, respectively. From eq. (4), this leads to a convenient cancellation of errors in the calculation of $\ln \gamma_{i,j}^{\infty}$. For the case of SM12, using the predicted value of $\Delta G_{i,i}^{\text{self}}$, this would suggest that on average the value of $\ln \gamma_{i,j}^{\infty}$ is overly positive by just 0.4. On the other hand, using the reference value of $\Delta G_{i,i}^{\text{self}}$, the corresponding value would be overly negative by 3.4. This is illustrated for SM12 in fig. 5 where the predictions were again broken down to cases where water was the solvent ($j =$water), water was the solute ($i =$water), and when neither the solvent or solute was water.
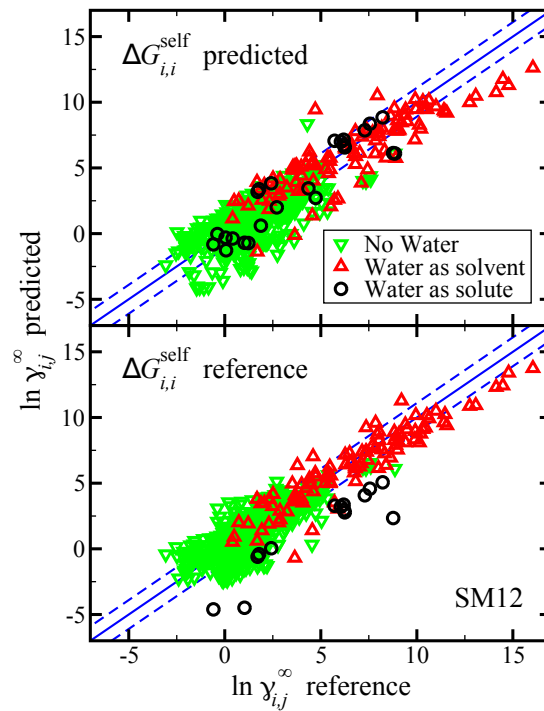
Figure 5. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^{\infty}$ using the SM12 universal solvent model for the case where water was the solvent, water was the solute, and where water was neither the solvent or solute, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference; the value of 1.1 was used based on the *RMSE* for the SM12 predictions. The reference values are from the Minnesota Solvation Database. In the top pane we present results when $\Delta G_{i,i}^{\text{self}}$ was predicted using the SM12 universal solvent model. In the bottom pane we present results when $\Delta G_{i,i}^{\text{self}}$ was computed using reference vapor pressure data.

Again, in general we find that the predicted values of $\ln \gamma_{i,j}^{\infty}$ do not change substantially when $\Delta G_{i,i}^{\text{self}}$ was predicted or computed using reference vapor pressure data. However, when water is the

solute, there is a clear difference. This would suggest that when water is the solute of interest, both $\Delta G_{i,j}^{solv}$ and $\Delta G_{i,i}^{self}$ should be computed consistently in the calculation of $\ln \gamma_{i,j}^\infty$.

In fig. 5 we additionally point out that we observe an apparent break in our trend for reference values of $\ln \gamma_{i,j}^\infty > 9$. For all cases when the reference values of $\ln \gamma_{i,j}^\infty$ are greater than 9, we under-predict the value of $\ln \gamma_{i,j}^\infty$. These would correspond to highly non-ideal (unfavorable) systems. For all 26 cases, water is the solvent. A possible source of error could be due to the difficulty of experimentally measuring such non-ideal values [70].

4.3.2. DECHEMA

Next, as a secondary test, values of $\gamma_{i,j}^\infty$ at 298.15 K were computed for 2163 systems, consisting of 117 unique solutes and 77 unique solvents, from Parts 1 to 6 of DECHEMA's "Activity Coefficients at Infinite Dilution" volumes [50–55]. A parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^\infty$ is provided in fig. 6. In the top pane of fig. 6 results are shown when we predict both $\Delta G_{i,i}^{self}$ and $\Delta G_{i,j}^{solv}$, and the bottom pane shows the case when $\Delta G_{i,i}^{self}$ was computed using reference data. A summary of the statistics is provided in table 4.
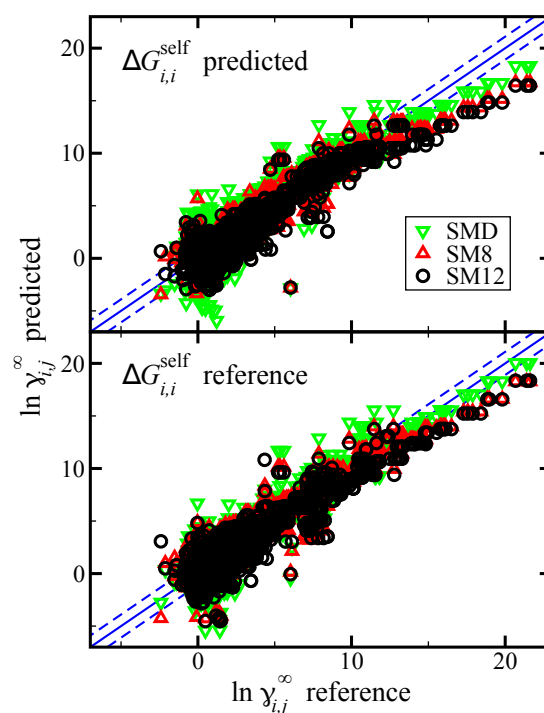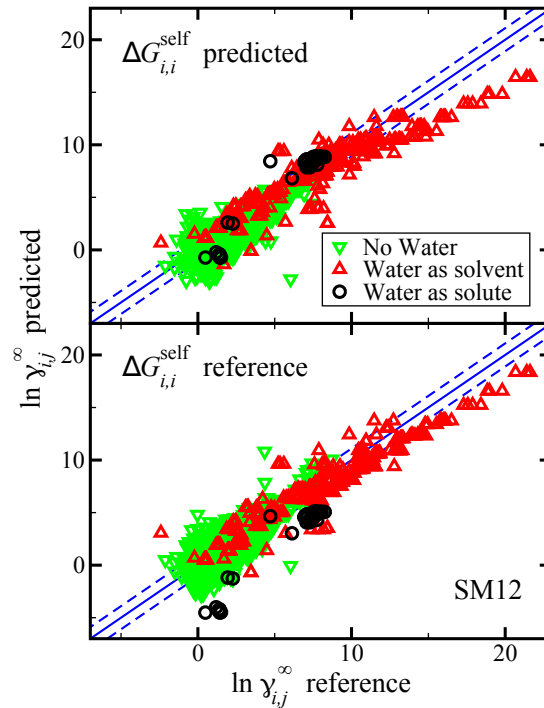

Figure 6. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^\infty$ from the DECHEMA. Predictions are made using the SM12, SM8, and SMD universal solvent model, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference. In the top pane we present results when $\Delta G_{i,i}^{self}$ was predicted using the SM12, SM8, or SMD universal solvent model. In the bottom pane we present results when $\Delta G_{i,i}^{self}$ was computed using reference vapor pressure data.

Table 4. A summary of the squared correlation coefficient (R^2), root mean square error (*RMSE*), and mean unsigned error (*MUE*) for $\ln \gamma_{i,j}^{\infty}$, and the average absolute percent deviation (*AAPD*) for $\gamma_{i,j}^{\infty}$ predicted using the SM12, SM8, and SMD universal solvent model as compared to the reference data from DECHEMA. For the predictions, we compare the case where 1) $\Delta G_{i,i}^{\text{self}}$ is predicted using the SM12, SM8, and SMD universal solvent model, and where 2) $\Delta G_{i,i}^{\text{self}}$ is computed using reference data, as indicated.

	$\Delta G_{i,i}^{\text{self}}$ predicted			$\Delta G_{i,i}^{\text{self}}$ reference		
	SM12	SM8	SMD	SM12	SM8	SMD
R^2	0.911	0.913	0.874	0.865	0.889	0.886
RMSE	1.163	1.065	1.382	1.354	1.196	1.271
AAPD	86.823	120.158	308.173	177.079	166.345	289.780
MUE	0.847	0.768	1.012	1.061	0.875	0.910

We again breakdown the predictions to cases where water was the solvent ($j =$water), water was the solute ($i =$water), and when neither the solvent or solute was water, and plot the results for SM12 in fig. 7. The breakdown resulted in 437 systems where water is the solvent, 52 systems where water is the solute, and 1668 systems where water was neither the solvent or solute.

Figure 7. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^{\infty}$ using the SM12 universal solvent model for the case where water was the solvent, water was the solute, and where water was neither the solvent or solute, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference. The reference values are from DECHEMA. In the top pane we present results when $\Delta G_{i,i}^{\text{self}}$ was predicted using the SM12 universal solvent model. In the bottom pane we present results when $\Delta G_{i,i}^{\text{self}}$ was computed using reference vapor pressure data.

In general, our observations are the same as when we compared to the Minnesota Solvation Database. First, there again does not appear to be a significant difference between the predictions wherein $\Delta G_{i,i}^{\text{self}}$ was predicted or computed using reference data. However, for the specific case of water as a solute, there is a noticeable difference. When $\Delta G_{i,i}^{\text{self}}$ was computed using reference vapor pressure data, $\ln \gamma_{i,j}^{\infty}$ when water was a solute was under-predicted on average by 3.02, 3.68, and 2.84 with SM12, SM8, and SMD, respectively. For of theses cases, the value of $\ln \gamma_{i,j}^{\infty}$ was always under-predicted except

for water in tetrachloroethene with SMD. Additionally, we again notice an abrupt deviation in our trend for reference values of $\ln \gamma_{i,j}^\infty$ greater than 9. As before, all of these highly non-ideal (unfavorable) systems, water was the solvent. Interestingly, the value of R^2 does increase. However, we are dealing with a larger data set that does contain duplicates.

4.3.3. Yaws' Aqueous Systems

Having identified challenges in predicting $\ln \gamma_{i,j}^\infty$ for aqueous systems, we sourced an additional data set composed of organics in water and water in organics at 298.15 K from "Yaws' Handbook of Properties for Aqueous Systems" [56]. Data was available for 121 systems where water is the solvent and 90 systems where water is the solute, with no duplicates. A parity plot of the results for SM12 is provided in fig. 8. In the top pane of fig. 8 results are shown when we predict both $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$, and the bottom pane shows the case when $\Delta G_{i,i}^{\text{self}}$ was computed using reference data. A summary of the statistics is provided in table 5, broken down into the case where water is the solvent and water is the solute.
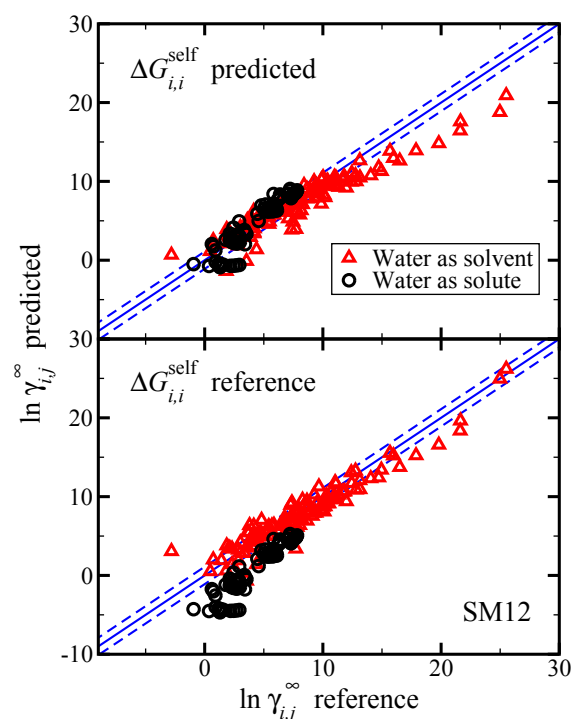
Figure 8. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^\infty$ using the SM12 universal solvent model for the case where water was the solvent and water was the solute, as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference. The reference values are from "Yaws' Handbook of Properties for Aqueous Systems" [56]. In the top pane we present results when $\Delta G_{i,i}^{\text{self}}$ was predicted using the SM12 universal solvent model. In the bottom pane we present results when $\Delta G_{i,i}^{\text{self}}$ was computed using reference vapor pressure data.

Table 5. A summary of the squared correlation coefficient (R^2), root mean square error (*RMSE*), and mean unsigned error (*MUE*) for $\ln \gamma_{i,j}^{\infty}$, and the average absolute percent deviation (*AAPD*) for $\gamma_{i,j}^{\infty}$ predicted using the SM12, SM8, and SMD universal solvent model as compared to the reference data from "Yaws' Handbook of Properties for Aqueous Systems" [56]. For the predictions, we compare the case where 1) $\Delta G_{i,i}^{\text{self}}$ is predicted using the SM12, SM8, and SMD universal solvent model, and where 2) $\Delta G_{i,i}^{\text{self}}$ is computed using reference data, as indicated. The results are additionally separated into case where water was the solvent and where water was the solute, as indicated.

| | $\Delta G_{i,i}^{\text{self}}$ predicted | | | $\Delta G_{i,i}^{\text{self}}$ reference | | |
	SM12	SM8	SMD	SM12	SM8	SMD
	water as solute			water as solute		
R^2	0.889	0.793	0.645	0.889	0.793	0.645
RMSE	1.339	1.836	2.482	3.721	3.600	3.175
AAPD	168.491	645.592	2598.544	94.691	93.947	97.359
MUE	1.116	1.625	2.168	3.485	3.397	2.733
	water as solvent			water as solvent		
R^2	0.913	0.928	0.907	0.916	0.946	0.951
RMSE	1.854	1.658	1.652	1.464	1.300	1.090
AAPD	117.058	160.651	327.313	409.655	197.774	176.344
MUE	1.353	1.191	1.317	1.115	1.017	0.792

The results here are consistent with the other data sets. Considering first $\ln \gamma_{i,j}^{\infty}$ for the case of water as a solute when $\Delta G_{i,i}^{\text{self}}$ was computed using reference vapor pressure data. For all cases with SM12 and SM8 $\ln \gamma_{i,j}^{\infty}$ is under predicted with an average error of 3.485 and 3.397, respectively. With SMD, $\ln \gamma_{i,j}^{\infty}$ is under predicted for 85 of the 90 systems, with an average error of 2.834. On the other hand, when $\Delta G_{i,i}^{\text{self}}$ was also predicted, with SM12, SM8, and SMD the number of systems for which $\ln \gamma_{i,j}^{\infty}$ was under predicted was 22, 13, and 17, respectively. And most importantly, with SM12, SM8, and SMD, the overall *MUE* noticeably decreased.

When water is the solvent, we find that with SM12, SM8, and SMD, the value of R^2 increases, and the value of the *MUE* and *RMSE* decreases for $\ln \gamma_{i,j}^{\infty}$. The difference in behavior results from a systematic error when water is the solute, where when $\Delta G_{i,i}^{\text{self}}$ for water is also predicted, we obtain a convenient cancellation of errors. We note that we find the values of *AAPD* to be less informative as the value can be dominated by just a few systems.

4.3.4. Free SM*x* $\gamma_{i,j}^{\infty}$ Calculator

To facilitate the use of the SM12, SM8, and SMD solvation models to predict $\gamma_{i,j}^{\infty}$ and ultimately phase equilibrium, we are actively developing HTML (HyperText Markup Language) based software that can run in a web browser. The software is currently housed on GitHub at: https://github.com/SpencerSabatino/SolvationFreeEnergyWebsite. For the 148 molecules for which QChem 5.1.2 [45,46] has SM*x* parameters (which excludes iodine and carboxylic acid containing molecules), the user can select any of the 148 molecules to form a binary pair and compute $\Delta G_{i,i}^{\text{self}}$, $\Delta G_{i,j}^{\text{solv}}$, and $\gamma_{i,j}^{\infty}$ for each component. The QChem 5.1.2 calculation results used by the software were generated in the present study. The software will continue to be developed with additional features added.

4.3.5. Reference Calculations

Overall, we found that the SM*x* (*x*= 12, 8, or D) predicted values of $\ln \gamma_{i,j}^{\infty}$ did not change substantially when $\Delta G_{i,i}^{\text{self}}$ was predicted compared to when it was computed using reference vapor pressure data, except in the case when water is the solute. Therefore, moving forward the selected method for calculating $\ln \gamma_{i,j}^{\infty}$ will be to compute $\Delta G_{i,j}^{\text{solv}}$ and $\Delta G_{i,i}^{\text{self}}$ consistently; put differently, both $\Delta G_{i,j}^{\text{solv}}$ and $\Delta G_{i,i}^{\text{self}}$ will be predicted.

To provide further insight into the potential utility of electronic structure calculations with SM*x* universal solvent models to predict $\ln \gamma_{i,j}^{\infty}$, the top performing SM12 was compared to two well-known

methods, the Universal Quasichemical Functional-group Activity Coefficients (UNIFAC) method [15,16,71] and the Modified Separation of Cohesive Energy Density (MOSCED) model [19,20,72]. UNIFAC was selected as a comparative method because it is perhaps the most common predictive excess Gibbs free energy method [73]. Additionally, it is widely used and integrated into most process simulators. Brouwer and Schuur [73] recently conducted a comparative study of eight predictive methods to compute $\ln \gamma_{i,j}^{\infty}$ at 298.15 K. Of the methods evaluated, the solubility parameter method MOSCED was the top performer. This strong performance of MOSCED is consistent with the findings of the 2005 re-parameterization [20,74] and our recent work [8,75]. MOSCED was therefore selected as an additional comparison alongside UNIFAC. Note that for water, we adopted the revised MOSCED parameters from ref. 75

The Minnesota Solvation Database, DECHEMA, and Yaws' reference databases were constrained to match the compound availability provided by the two models, UNIFAC and MOSCED. In comparison to the Minnesota Solvation Database, UNIFAC was able to make predictions for 1065 (of 1140) systems with 99 unique solutes and 68 unique solvents. MOSCED was able to make predictions for 732 (of 1140) systems from the Minnesota Solvation Database with 58 unique solutes and 54 unique solvents. In regards to DECHEMA, UNIFAC was able to model 1988 (of 2163) systems with 112 unique solutes and 70 unique solvents. MOSCED was able to model 1818 (of 2163) systems with 71 unique solutes and 59 unique solvents. We again note that the DECHEMA data set includes repeated systems. All of the UNIFAC calculations were performed with CHEMCAD 7 [76]. A simple GNU Octave/MATLAB [77,78] M-file to predict limiting activity coefficients using MOSCED, a screen cast tutorial and additional references are available on the YouTube Channel of Andrew Paluch [79].

A parity plot of the results for SM12, UNIFAC and MOSCED is provided in fig. 9. In the top pane of fig. 9 results for the Minnesota Solvation Database are displayed, and in the bottom pane the corresponding DECHEMA results are displayed. A summary of the statistics is included in table 6.
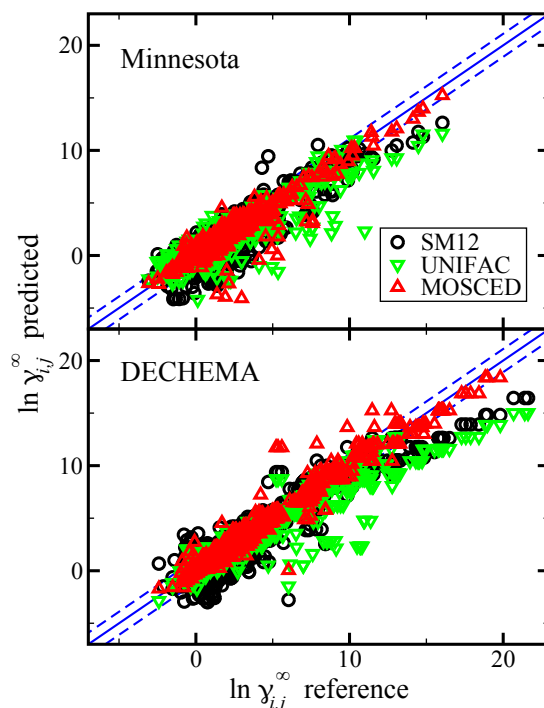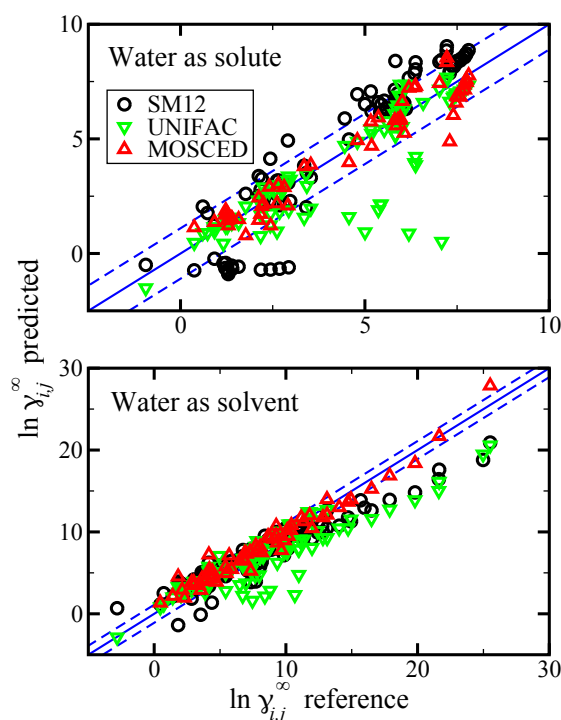

Figure 9. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^{\infty}$ at 298.15 K using the SM12 universal solvent model, UNIFAC, and MOSCED versus reference data from the Minnesota Solvation Database version 2012 (top pane) [47] and DECHEMA (bottom pane) [50–55], as indicated. The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference. For the SM12 universal solvent model predictions, $\Delta G_{i,i}^{\text{self}}$ was predicted.

Table 6. A summary of the squared correlation coefficient (R^2), root mean square error (*RMSE*), and mean unsigned error (*MUE*) for $\ln \gamma_{i,j}^{\infty}$, and the average absolute percent deviation (*AAPD*) for $\gamma_{i,j}^{\infty}$ predicted using UNIFAC and MOSCED as compared to the reference data from the Minnesota Solvation Database version 2012 [47] and DECHEMA [50–55].

	Minnesota		DECHEMA	
	UNIFAC	MOSCED	UNIFAC	MOSCED
R^2	0.820	0.884	0.915	0.976
RMSE	1.077	0.802	1.157	0.524
AAPD	52.595	41.251	41.044	131.758
MUE	0.634	0.422	0.559	0.241

For both the Minnesota Solvation Database and DECHEMA, MOSCED achieves the largest R^2 and smallest *RMSE* and *MUE* for $\ln \gamma_{i,j}^{\infty}$. For DECHEMA, MOSCED does have a noticeably larger *AAPD* for $\gamma_{i,j}^{\infty}$, but we again note that the value *AAPD* can be heavily influenced by a few systems. Comparing to SM12 (see tabs. 3 and 4), UNIFAC does slightly outperform SM12. However, considering R^2 and *RMSE*, the UNIFAC and SM12 results are comparable, especially with regards to the DECHEMA data set.

In following the same methodology conducted previously, we sought to analyze the models' ability to predict limiting activity coefficients when water is the solute and when water is the solvent. The reference data from "Yaws' Handbook of Properties for Aqueous Systems" again served as the basis for systems that were modeled using the SM12 universal continuum solvent model, UNIFAC and MOSCED. A parity plot is provided in fig. 10 where the top pane contains data where water is the solute and the bottom pane where water is the solvent. The corresponding statistics are recorded in table 7.

Figure 10. Parity plot of the predicted versus reference values of $\ln \gamma_{i,j}^{\infty}$ at 298.15 K using the SM12 universal solvent model, UNIFAC, and MOSCED versus reference data from "Yaws' Handbook of Properties for Aqueous Systems" [56]. In the top pane results as shown for the case of water as the solute (water in organics), and in the bottom pane results are shown for the case of water as the solvent (organics in water). The blue lines corresponds to the $y = x$ line, and the dashed blue lines correspond to ± 1.1 and are drawn as a reference. For the SM12 universal solvent model predictions, $\Delta G_{i,i}^{\text{self}}$ was predicted.

Table 7. A summary of the squared correlation coefficient (R^2), root mean square error ($RMSE$), and mean unsigned error (MUE) for $\ln \gamma_{i,j}^{\infty}$, and the average absolute percent deviation ($AAPD$) for $\gamma_{i,j}^{\infty}$ predicted using UNIFAC and MOSCED as compared to the reference data from "Yaws' Handbook of Properties for Aqueous Systems" [56]. The results are separated for the case of water as the solute (water in organics) and water as the solvent (organics in water).

| | water as solute | | water as solvent | |
	UNIFAC	MOSCED	UNIFAC	MOSCED
R^2	0.713	0.923	0.816	0.962
$RMSE$	1.423	0.688	2.479	1.000
$AAPD$	51.395	55.749	77.925	138.245
MUE	0.824	0.532	1.647	0.761

Within the Yaws' reference set where water is the solute, the SM12 universal continuum solvent model made predictions for 90 systems, UNIFAC modeled 87 systems and MOSCED 53 systems. When water was the solvent, the SM12 universal continuum solvent model made predictions for 121 systems, UNIFAC 114 systems and MOSCED 61 systems. MOSCED in both cases was the highest performer, yet modeled the lowest number of systems due to its limited compound inventory. We additionally note that the MOSCED parameters for water were optimized using limiting activity coefficients from "Yaws' Handbook of Properties for Aqueous Systems" [75]. Overall, all models seem to have more difficulty predicting $\ln \gamma_{i,j}^{\infty}$ when water is the solute. UNIFAC is the lowest performer in both cases. A plausible reason for the SM12 universal continuum solvent model's improved accuracy over UNIFAC could be due to the cancellation of errors that occurs in computing $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$ consistently.

Again, when water is the solvent with reference values of $\ln \gamma_{i,j}^{\infty} > 9$, we notice a tapering or an under prediction in $\ln \gamma_{i,j}^{\infty}$ for both the SM12 and UNIFAC models. As stated previously, these are highly non-ideal systems.

The SMx universal solvent models, MOSCED, and UNIFAC are all limited in the system they can treat. MOSCED can only be used to model binary systems for which the the molar volume and five molecular descriptors are available for each component. At present, MOSCED parameters are available for 130 organic solvents and water [20,75]. Efforts have been made to use electronic structure calculations to predict missing MOSCED parameters, including the use of solvation free energy calculations with the SMx universal solvent models [42,80–84]. While we would not expect the predicted MOSCED parameters to be as good as those regressed using experimental data, the results of the present study nonetheless would suggest they would yield acceptable results, especially for early stage process conceptualization and design applications. Interestingly, ref. 42 suggested that results obtained with MOSCED parameters predicted using solvation free energy calculations with the SMx universal solvent model were superior to predictions made with SMx alone. This may result from implicitly including reference data by using MOSCED parameters that were regressed using reference data.

UNIFAC is a group contribution-based method, and in general is able to cover a wider range of chemical compound space. Nonetheless, some parameters are still missing from the parameter matrix. Recently, it has likewise been demonstrated how electronic structure calculations, specifically COSMO-based (conductor-like screening model) calculations, can be used to predict missing UNIFAC parameters [85–87]. So while the SMx universal solvent models, MOSCED, and UNIFAC are all limited in the system they can treat, there is opportunity for synergy among the methods.

5. Summary and Conclusion

The SMx (x= 12, 8, or D) universal solvent models are implicit solvent models that can be used to compute the solvation free energy at 298.15 K using electronic structure calculations. Fundamentally, the solvation free energy is an important thermophysical property, and corresponds to the transfer free energy of taking a single solute from a non-interacting ideal gas state to solution (at infinite dilution) at the same molar density. In going from a non-interacting state to a state with a single solute molecule in solution, the solvation free energy is representative of the solute-solvent interactions in solution. If the solvation free energy of a solute is computed in two different solvents, they can readily be used (via a thermodynamic cycle) to compute the transfer free energy between the two solvents. This transfer free energy is directly related to the partition coefficient, which can be used to relate to the relative composition of the solute in the two (immiscible) solvents at equilibrium, or more generally can be used to compare the relative affinity of the solute for the two solvents.

In conventional chemical engineering thermodynamic modeling of phase equilibrium, we are interested in the activity coefficient of the components in the mixture, and the excess Gibbs free energy of the mixture. Coincidentally, values of the solvation free energy are not directly measured experimentally. Rather, it is common to measure limiting activity coefficients from which one may obtain the solvation free energy. The limiting activity coefficient is related to the transfer free energy of a single solute from a system of pure liquid solute to a solution phase at infinite dilution. Moreover, limiting activity coefficients may be related directly to the parameters in common binary interaction excess Gibbs free energy models used to model phase equilibrium.

In the present study we therefore assessed the ability to use the SMx universal solvent models to predict limiting activity coefficients. We considered 148 molecules for which SMx (solvent) parameters were available in QChem 5.1.2. As seen in eq. 4 the calculation of limiting activity coefficients is dependent upon both the self-solvation free energy ($\Delta G_{i,i}^{\text{self}}$) and solvation free energy terms ($\Delta G_{i,j}^{\text{solv}}$). Therefore we assessed the contribution of each individual term. The Minnesota Solvation Database was used as our reference set, since it was used to parameterize the SM12 model, which was consistently the top performer throughout the analysis.

In assessing the contribution of $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$ terms, water was found to be a large source of deviation. The $\Delta G_{i,i}^{\text{self}}$ of water was predicted to be overly negative, suggesting we over-estimate the strength of the water-water interactions. We gained further insight into this contribution by breaking down the $\Delta G_{i,j}^{\text{solv}}$ analysis into cases where water was the solute (i =water), water was the solvent (j =water), and where water was neither the solute or solvent. We found that when water was the solute, $\Delta G_{i,j}^{\text{solv}} / (RT)$ is predicted to be overly negative on average by 3.4 (dimensionless units) in the SM12 model as compared to the self-solvation free energy term $\Delta G_{i,i}^{\text{self}} / (RT)$ which was overly negative by 3.8 (dimensionless units). This deviation ends up being advantageous in discerning how to calculate limiting activity coefficients, as there were two methods in which to do so: 1) prediction of $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$ and use of reference values for the molar volume, and 2) compute $\Delta G_{i,i}^{\text{self}}$ using pure component vapor pressure data, which is readily available. We evaluated these two methods using three reference data sets: the Minnesota Solvation Database, Parts 1 to 6 of DECHEMA's "Activity Coefficients at Infinite Dilution" volumes [50–55], and "Yaws' Handbook of Properties for Aqueous Systems"[56]. All three methods told the same story, where at first glance there does not appear to be a significant difference in the predictions when $\Delta G_{i,i}^{\text{self}}$ is predicted or computed from reference data. However, upon isolating systems where water is the solute, we see a that $\ln \gamma_{i,j}^{\infty}$ is under-predicted when $\Delta G_{i,i}^{\text{self}}$ is computed from reference vapor pressure data. Yet when both $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$ are predicted there is a convenient cancellation of errors in the $\ln \gamma_{i,j}^{\infty}$. Therefore this suggests that the best method to calculate limiting activity coefficients moving forward is to compute $\Delta G_{i,i}^{\text{self}}$ and $\Delta G_{i,j}^{\text{solv}}$ consistently; that is, to predict both terms.

In conclusion, the findings were compared to two popular and successful predictive methods, UNIFAC and MOSCED. Comparison to these models allows us to assess the viability of taking the SMx models into further chemical engineering applications. A recent 2019 study conducted by Brouwer and Schuur [73], previously showed MOSCED was the top performing model in an comparative study to predict limiting activity coefficients at 298.15 K. Our comparison further confirms that MOSCED is the top performing model in comparison to UNIFAC and the SM12 universal continuum solvent model, yet it is limited by the number of compounds and systems it can model. UNIFAC and the SM12 universal continuum solvent model are comparable in performance among the three reference sets. This suggests that the SMx models have potential to be used and tested in chemical engineering applications such as the prediction of phase equilibrium There is also opportunity for synergy between SM12, UNIFAC, and MOSCED.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

1. Reddy, B.M.; Raghavan, K.V. *Industrial Catalysis and Separations: Innovations for Process Intensification.*; Apple Academic Press, 2014.
2. Sandler, S.I. Infinite dilution activity coefficients in chemical, environmental and biochemical engineering. *Fluid Phase Equilib.* **1996**, *116*, 343 – 353.

3. Nanu, D.E.; Mennucci, B.; de Loos, T.W. Predicting infinite dilution activity coefficients with the group contribution solvation model: an extension of its applicability to aqueous systems. *Fluid Phase Equilib.* **2004**, *221*, 127 – 137.

4. Missen, R.W. On Criteria for Occurrence of Azeotropes in Isothermal and Isobaric Binary Systems. *Can. J. Chem. Eng.*, *83*, 667–674.

5. Missen, R.W. On Determination of Azeotrope Coordinates from gE for Binary Isothermal and Isobaric Systems. *Can. J. Chem. Eng.* **2006**, *84*, 316–321.

6. Brandani, V. Use of infinite-dilution activity coefficients for predicting Azeotrope formation at constant temperature and partial miscibility in binary liquid mixtures. *Ind. Eng. Chem. Fundam.* **1974**, *13*, 154–156.

7. Shulgin, I.; Fischer, K.; Noll, O.; Gmehling, J. Classification of homogeneous binary azeotropes. *Ind. Eng. Chem. Res.* **2001**, *40*, 2742–2747.

8. Dhakal, P.; Roese, S.N.; Stalcup, E.M.; Paluch, A.S. Application of MOSCED To Predict Limiting Activity Coefficients, Hydration Free Energies, Henry's Constants, Octanol/Water Partition Coefficients, and Isobaric Azeotropic Vapor–Liquid Equilibrium. *J. Chem. Eng. Data* **2018**, *63*, 352–364.

9. Kojima, K.; Zhang, S.; Hiaki, T. Measuring methods of infinite dilution activity coefficients and a database for systems including water. *Fluid Phase Equilib.* **1997**, *131*, 145 – 179.

10. Freund, H.; Sundmacher, K. Process Intensification, 1. Fundamentals and Molecular Level. In *Ullmann's Encyclopedia of Industrial Chemistry*; John Wiley and Sons, Inc., 2011. doi:10.1002/14356007.o22_o02.

11. Schacht, C.S.; Zubeir, L.; de Loos, T.W.; Gross, J. Application of Infinite Dilution Activity Coefficients for Determining Binary Equation of State Parameters. *Ind. Eng. Chem. Res.* **2010**, *49*, 7646–7653.

12. Schreiber, L.B.; Eckert, C.A. Use of Infinite Dilution Activity Coefficients with Wilson's Equation. *Ind. Eng. Chem. Process Des. Dev.* **1971**, *10*, 572–576.

13. Prausnitz, J.M.; Lichtenthaler, R.N.; de Azevedo, E.G. *Molecular Thermodynamics of Fluid-phase Equilibria*, 2nd ed.; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1986.

14. Eckert, C.A.; Sherman, S.R. Measurement and prediction of limiting activity coefficients. *Fluid Phase Equilib.* **1996**, *116*, 333 – 342.

15. Fredenslund, A.; Jones, R.L.; Prausnitz, J.M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21*, 1086–1099.

16. Gmehling, J.; Rasmussen, P.; Fredenslund, A. Vapor-liquid equilibriums by UNIFAC group contribution. Revision and extension. 2. *Ind. Eng. Chem. Process Des. Dev.* **1982**, *21*, 118–127.

17. Gmehling, J.; Lohmann, J.; Jakob, A.; .; Li, J.; Joh, R. A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Ind. Eng. Chem. Res.* **1998**, *37*, 4876–4882.

18. Lohmann, J.; Joh, R.; Gmehling, J. From UNIFAC to Modified UNIFAC (Dortmund). *Ind. Eng. Chem. Res.* **2001**, *40*, 957–964.

19. Thomas, E.R.; Eckert, C.A. Prediction of limiting activity coefficients by a modified separation of cohesive energy density model and UNIFAC. *Ind. Eng. Chem. Proc. Des. Dev.* **1984**, *23*, 194–209.

20. Lazzaroni, M.J.; Bush, D.; Eckert, C.A.; Frank, T.C.; Gupta, S.; Olson, J.D. Revision of MOSCED Parameters and Extension to Solid Solubility Calculations. *Ind. Eng. Chem. Res.* **2005**, *44*, 4075–4083.

21. Hansen, C.M. The Universality of the Solubility Parameter. *Ind. Eng. Chem. Prod. Res. Dev.* **1969**, *8*, 2–11.

22. Diedenhofen, M.; Eckert, F.; Klamt, A. Prediction of Infinite Dilution Activity Coefficients of Organic Compounds in Ionic Liquids Using COSMO-RS. *J. Chem. Eng. Data* **2003**, *48*, 475–479.

23. Hempel, S.; Fischer, J.; Paschek, D.; Sadowski, G. Activity Coefficients of Complex Molecules by Molecular Simulation and Gibbs-Duhem Integration. *Soft Mater.* **2012**, *10*, 26–41.

24. Ley, R.T.; Fuerst, G.B.; Redeker, B.N.; Paluch, A.S. Developing a Predictive Form of MOSCED for Nonelectrolyte Solids Using Molecular Simulation: Application to Acetanilide, Acetaminophen, and Phenacetin. *Ind. Eng. Chem. Res.* **2016**, *55*, 5415–5430.

25. Duarte Ramos Matos, G.; Calabró, G.; Mobley, D.L. Infinite Dilution Activity Coefficients as Constraints for Force Field Parameterization and Method Development. *J. Chem. Theory Comput.* **2019**, *15*, 3066–3074.

26. Lisboa, F.M.; Pliego, Jr., J.R. Infinite dilution activity coefficient from SMD calculations: accuracy and performance for predicting liquid-liquid equilibria. *J. Mol. Model.* **2018**, *24*, 56.

27. Wankat, P.C. *Separation Process Engineering: Includes Mass Transfer*, 3rd ed.; Pearson Education, Inc.: Upper Saddle River, NJ, 2012.

28. Onken, U.; Fischer, K.; Rarey, J.; Kleiber, M. Estimation of Physical Properties. In *Ullmann's Encyclopedia of Industrial Chemistry*; John Wiley and Sons, Inc., 2008. doi:10.1002/14356007.b01_06.pub2.

29. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

30. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

31. Marenich, A.V.; Olson, R.M.; Kelly, C.P.; Cramer, C.J.; Truhlar, D.G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *J. Chem. Theory. Comput.* **2007**, *3*, 2011–2033.

32. Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Parametrized Model for Aqueous Free Energies of Solvation Using Geometry-Dependent Atomic Surface Tensions with Implicit Electrostatics. *J. Phys. Chem. B* **1997**, *101*, 7147–7157.

33. Hawkins, G.D.; Liotard, D.A.; Cramer, C.J.; Truhlar, D.G. OMNISOL: Fast Prediction of Free Energies of Solvation and Partition Coefficeints. *J. Org. Chem.* **1998**, *63*, 4305–4313.

34. MINNESOTA SOLVATION MODELS AND SOLVATION SOFTWARE. https://comp.chem.umn.edu/solvation/ (accessed August 8, 2019).

35. Cramer, C.J. *Essentials of Computational Chemistry*; John Wiley & Sons Ltd: Chichester, West Sussex, England, 2002.

36. Chipot, C.; Pohorille, A., Eds. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Vol. 86, *Springer Series in Chemical Physics*, Springer: New York, NY, 2007.

37. Paluch, A.S.; Maginn, E.J. Predicting the Solubility of Solid Phenanthrene: A Combined Molecular Simulation and Group Contribution Approach. *AIChE J.* **2013**, *59*, 2647–2661.

38. Fuerst, G.B.; Ley, R.T.; Paluch, A.S. Calculating the Fugacity of Pure, Low Volatile Liquids via Molecular Simulation with Application to Acetanilide, Acetaminophen, and Phenacetin. *Ind. Eng. Chem. Res.* **2015**, *54*, 9027–9037.

39. Noroozi, J.; Paluch, A.S. Microscopic Structure and Solubility Predictions of Multifunctional Solids in Supercritical Carbon Dioxide: A Molecular Simulation Study. *J. Phys. Chem. B* **2017**, *121*, 1660–1674.

40. Long, G.E.; Dhakal, P.; Redeker, B.N.; Paluch, A.S. Using limiting activity coefficients to efficiently evaluate the ability of fixed-charge force fields to model miscible water plus cosolvent mixtures. *Mol. Simul.* **2019**, *45*, 322–335.

41. Paluch, A.S.; Maginn, E.J. Efficient Estimation of the Equilibrium Solution-Phase Fugacity of Soluble Nonelectrolyte Solids in Binary Solvents by Molecular Simulation. *Ind. Eng. Chem. Res.* **2013**, *52*, 13743–13760.

42. Phifer, J.R.; Solomon, K.J.; Young, K.L.; Paluch, A.S. Computing MOSCED parameters of nonelectrolyte solids with electronic structure methods in SMD and SM8 continuum solvents. *AIChE J.* **2017**, *63*, 781–791.

43. Winget, P.; Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Predicting the Vapor Pressures from Self-Solvation Free Energies Calculated by the SM5 Series of Universal Solvation Models. *J. Phys. Chem. B* **2000**, *104*, 4726.

44. Horn, H.W.; Swope, W.C.; Pitera, J.W. Characterization of the TIP4P-Ew water model: Vapor pressure and boiling point. *J. Chem. Phys.* **2005**, *123*, 194504.

45. Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A.T.B.; Wormit, M.; Kussmann, J.; Lange, A.W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P.R.; Jacobson, L.D.; Kaliman, I.; Khaliullin, R.Z.; Kús, T.; Landau, A.; Liu, J.; Proynov, E.I.; Rhee, Y.M.; Richard, R.M.; Rohrdanz, M.A.; Steele, R.P.; Sundstrom, E.J.; Woodcock III, H.L.; Zimmerman, P.M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G.J.O.; Bernard, Y.A.; Berquist, E.; Brandhorst, K.; Bravaya, K.B.; Brown, S.T.; Casanova, D.; Chang, C.M.; Chen, Y.; Chien, S.H.; Closser, K.D.; Crittenden, D.L.; Diedenhofen, M.; DiStasio Jr., R.A.; Dop, H.; Dutoi, A.D.; Edgar, R.G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M.W.D.; Harbach, P.H.P.; Hauser, A.W.; Hohenstein, E.G.; Holden, Z.C.; Jagau, T.C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R.A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C.M.; Lao, K.U.; Laurent, A.; Lawler, K.V.; Levchenko, S.V.; Lin, C.Y.; Liu, F.; Livshits, E.; Lochan, R.C.; Luenser, A.; Manohar, P.; Manzer, S.F.; Mao, S.P.; Mardirossian, N.; Marenich, A.V.; Maurer, S.A.; Mayhall, N.J.; Oana, C.M.; Olivares-Amaya, R.; O'Neill, D.P.; Parkhill, J.A.; Perrine, T.M.; Peverati, R.; Pieniazek, P.A.; Prociuk, A.; Rehn, D.R.; Rosta, E.; Russ, N.J.; Sergueev, N.; Sharada, S.M.; Sharmaa, S.; Small, D.W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.C.;

Thom, A.J.W.; Tsuchimochi, T.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M.A.; Wenzel, J.; White, A.; Williams, C.F.; Vanovschi, V.; Yeganeh, S.; Yost, S.R.; You, Z.Q.; Zhang, I.Y.; Zhang, X.; Zhou, Y.; Brooks, B.R.; Chan, G.K.L.; Chipman, D.M.; Cramer, C.J.; Goddard III, W.A.; Gordon, M.S.; Hehre, W.J.; Klamt, A.; Schaefer III, H.F.; Schmidt, M.W.; Sherrill, C.D.; Truhlar, D.G.; Warshel, A.; Xua, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A.T.; Besley, N.A.; Chai, J.D.; Dreuw, A.; Dunietz, B.D.; Furlani, T.R.; Gwaltney, S.R.; Hsu, C.P.; Jung, Y.; Kong, J.; Lambrecht, D.S.; Liang, W.; Ochsenfeld, C.; Rassolov, V.A.; Slipchenko, L.V.; Subotnik, J.E.; Van Voorhis, T.; Herbert, J.M.; Krylov, A.I.; Gill, P.M.W.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.

46. Q-CHEM. http://www.q-chem.com/. (accessed December 16, 2019).

47. Marenich, A.V.; Kelly, C.P.; Thompson, J.D.; Hawkins, G.D.; Chambers, C.C.; Giesen, D.J.; Winget, P.; Cramer, C.J.; Truhlar, D.G. *Minnesota Solvation Database - version 2012*; University of Minnesota: Minneapolis, MN. (accessed November 13, 2018), http://comp.chem.umn.edu/mnsol.

48. Yaws, C.L. *Yaws' Critical Property Data for Chemical Engineers and Chemists*; Knovel, 2012; 2013; 2014. https://app.knovel.com/hotlink/toc/id:kpYCPDCECD/yaws-critical-property/yaws-critical-property.

49. Verevkin, S.P. Measurement and Prediction of the Monocarboxylic Acids Thermochemical Properties. *J. Chem. Eng. Data* **2000**, *45*, 953–960.

50. Tiegs, D.; Gmehling, J.; Medina, A.; Soares, M.; Bastos, J.; Alessi, P.; Kikic, I., Eds. *Activity Coefficients at Infinite Dilution, Part 1: C_1–C_9*; DECHEMA: Frankfurt am Main, Germany, 1986.

51. Tiegs, D.; Gmehling, J.; Medina, A.; Soares, M.; Bastos, J.; Alessi, P.; Kikic, I., Eds. *Activity Coefficients at Infinite Dilution, Part 2: C_{10}–C_{36}*; DECHEMA: Frankfurt am Main, Germany, 1986.

52. Gmehling, J.; Menke, J.; Schiller, M., Eds. *Activity Coefficients at Infinite Dilution, Part 3: C_1–C_9*; DECHEMA: Frankfurt am Main, Germany, 1994.

53. Gmehling, J.; Menke, J.; Schiller, M., Eds. *Activity Coefficients at Infinite Dilution, Part 4: C_{10}–C_{36} with O_2S and H_2O*; DECHEMA: Frankfurt am Main, Germany, 1994.

54. Gmehling, J.; Menke, J., Eds. *Activity Coefficients at Infinite Dilution, Part 5: C_1–C_{16}*; DECHEMA: Frankfurt am Main, Germany, 2007.

55. Gmehling, J.; Menke, J., Eds. *Activity Coefficients at Infinite Dilution, Part 6: C_{17}–C_{78} with D_2O and H_2O*; DECHEMA: Frankfurt am Main, Germany, 2008.

56. Yaws, C.L. *Yaws' Handbook of Properties for Aqueous Systems*; Knovel, 2012. https://app.knovel.com/hotlink/toc/id:kpYHPAS006/yaws-handbook-properties/yaws-handbook-properties.

57. Daylight Chemical Information Systems, Inc. https://www.daylight.com/. (accessed June 26, 2019).

58. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

59. PubChem. https://pubchem.ncbi.nlm.nih.gov/. (accessed June 26, 2019).

60. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchinson, G.R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.

61. Open Babel: The Open Source Chemistry Toolbox. http://openbabel.org/wiki/Main_Page. (accessed June 26, 2019).

62. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

63. Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Molec. Graph. Model* **2006**, *25*, 247–260.

64. Gasteiger, J.; Marsili, M. A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **1978**, *34*, 3181–3184.

65. Dunning, Jr., T.H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

66. Zhao, Y.; Truhlar, D.G. The M06 theory of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account* **2008**, *120*, 215–241.

67. Cramer, C.J.; Truhlar, D.G. Quantum Chemical Conformational Analysis of 1,2-Ethanediol: Correlation and Solvation Effects on the Tendency To Form Internal Hydrogen Bonds in the Gas Phase and in Aqueous Solutions. *J. Am. Chem. Soc.* **1994**, *116*, 3892–3900.

68. Chemistry Data Series, Volume I: Vapor-Liquid Equilibrium Data Collection. https://dechema.de/en/Media/Publications/Chemistry+Data+Series/Volume+I.html.

69. Harris, D.C. *Quantitative Chemical Analysis*, 6th ed.; W. H. Freeman and Company: New York, NY; pp. 66–68.

70. Sherman, S.R.; Trampe, D.B.; Bush, D.M.; Schiller, M.; Eckert, C.A.; Dallas, A.J.; Li, J.; Carr, P.W. Compilation and Correlation of Limiting Activity Coefficients of Nonelectrolytes in Water. *Ind. Eng. Chem. Res.* **1996**, *35*, 1044–1058.

71. Domańska, U. Experimental Data of Fluid Phase Equilibria- Correlation and Prediction Models: A Review. *Processes* **2019**, *7*, 277.

72. Howell, W.J.; Karachewski, A.M.; Stephenson, K.M.; Eckert, C.A.; Park, J.H.; Carr, P.W.; Rutan, S.C. An Improved MOSCED Equation for the Prediction and Application of Infinite Dilution Activity Coefficients. *Fluid Phase Equilib.* **1989**, *52*, 151–160.

73. Brouwer, T.; Schuur, B. Model Performances Evaluated for Infinite Dilution Activity Coefficients Prediction at 298.15 K. *Ind. Eng. Chem. Res.* **2019**, *58*, 8903–8914.

74. Anderson, T.C.F.J.J.; Olson, J.D.; Eckert, C.A. Application of MOSCED and UNIFAC to screen hydrophobic solvents for extraction of hydrogen-bonding organics from aqueous solution. *Ind. Eng. Chem. Res.* **2007**, *46*, 4621–4625.

75. Dhakal, P.; Paluch, A.S. Assessment and Revision of the MOSCED Parameters for Water: Applicability to Limiting Activity Coefficients and Binary Liquid-Liquid Equilibrium. *Ind. Eng. Chem. Res.* **2018**, *57*, 1689–1695.

76. CHEMCAD 7.1.0.9402. http://www.chemstations.com.

77. Eaton, J.W.; Bateman, D.; Hauberg, S. *GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations*; CreateSpace Independent Publishing Platform, 2009. ISBN 1441413006, http://www.gnu.org/software/octave/doc/interpreter.

78. The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 9.4.0.813654 (R2018a)*, 2018.

79. Andrew Paluch: MOSCED. https://www.youtube.com/c/AndrewPaluch (accessed August 8, 2019).

80. Cox, C.E.; Phifer, J.R.; da Silva, L.F.; Nogueira, G.G.; Ley, R.T.; O'Loughlin, E.J.; Barbosa, A.K.P.; Rygelski, B.T.; Paluch, A.S. Combining MOSCED with molecular simulation free energy calculations or electronic structure calculations to develop an efficient tool for solvent formulation and selection. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 183–199.

81. Phifer, J.R.; Cox, C.E.; da Silva, L.F.; Nogueira, G.G.; Barbosa, A.K.P.; Ley, R.T.; Bozada, S.M.; O'Loughlin, E.J.; Paluch, A.S. Predicting the equilibrium solubility of solid polycyclic aromatic hydrocarbonds and dibenzothiophene using a combination of MOSCED plus molecular simulation or electronic structure calculations. *Mol. Phys.* **2017**, *115*, 1286–1300.

82. Diaz-Rodriguez, S.; Bozada, S.M.; Phifer, J.R.; Paluch, A.S. Predicting cyclohexane/water distribution coefficients for the SAMPL5 challenge with MOSCED and the SMD solvation model. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1007–1017.

83. Dhakal, P.; Roese, S.N.; Lucas, M.A.; Paluch, A.S. Predicting Limiting Activity Coefficients and Phase Behavior from Molecular Structure: Expanding MOSCED to Alkanediols Using Group Contribution Methods and Electronic Structure Calculations. *J. Chem. Eng. Data* **2018**, *63*, 2586–2598.

84. Gnap, M.; Elliott, J.R. Estimation of MOSCED parameters from the COSMO-SAC database. *Fluid Phase Equilib.* **2018**, *470*, 241–248.

85. Dong, Y.; Zhu, R.; Guo, Y.; Lei, Z. A United Chemical Thermodynamic Model: COSMO-UNIFAC. *Ind. Eng. Chem. Res.* **2018**, *57*, 15954–15958.

86. Zhu, R.; Taheri, M.; Zhang, J.; Lei, Z. Extension of the COSMO-UNIFAC Thermodynamic Model. *Ind. Eng. Chem. Res.* **2020**, *59*, 1693–1701.

87. Bell, I.H.; Mickoleit, E.; Hsieh, C.M.; Lin, S.T.; Vrabec, J.; Breitkopf, C.; Jäger, A. A Benchmark Open-Source Implementation of COSMO-SAC. *J. Chem. Theory Comput.* **2020**. doi:10.1021/acs.jctc.9b01016.