



HAL
open science

Prediction of solvation energies at infinite dilution by the tc-PR cubic equation of state with advanced mixing rule based on COSMO-RS as gE model

Francisco Carlos Paes, Romain Privat, Jean-Noël Jaubert, Baptiste Sirjean

► To cite this version:

Francisco Carlos Paes, Romain Privat, Jean-Noël Jaubert, Baptiste Sirjean. Prediction of solvation energies at infinite dilution by the tc-PR cubic equation of state with advanced mixing rule based on COSMO-RS as gE model. *Journal of Molecular Liquids*, 2023, 386, pp.122480. 10.1016/j.molliq.2023.122480 . hal-04166077

HAL Id: hal-04166077

<https://hal.science/hal-04166077>

Submitted on 19 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prediction of solvation energies at infinite dilution by the *tc*-PR cubic equation of state with advanced mixing rule based on COSMO-RS as g^E model.

Francisco Carlos PAES⁽¹⁾, Romain PRIVAT^{(1,2)*}, Jean-Noël JAUBERT^(1,2), Baptiste SIRJEAN⁽¹⁾

¹ *Université de Lorraine, CNRS, LRGP, F-54000 Nancy, France*

² *École Nationale Supérieure des Industries Chimiques, 1 rue Grandville, 54000 Nancy, France*

*Corresponding author: romain.privat@univ-lorraine.fr

ABSTRACT

The prediction of solvation quantities is an important issue in several fields of chemical and process engineering. In this regard, this work was conducted to increase the prediction power of the equation-of-state (EoS) approach by coupling the *tc*-PR (*translated-consistent* Peng-Robinson) EoS with COSMO-RS (Conductor-like Screening Model for Real Solvents) as g^E model through an advanced mixing rule. Since the only input parameter of the COSMO-RS model is the charge density (σ) profile of each molecule, the resulting EoS (called *tc*-PR/COSMO-RS) does not require any binary interaction parameter contrary to other predictive cubic EoS, such as VTPR, PPR78, PSRK, and UMR-PRU. This feature allows the quantification of solvation data of a great number of systems from readily available pure-component properties (critical temperature and pressure, acentric factor and σ -profile). Satisfactory results were obtained for the prediction of solvation Gibbs energy data that were extracted from the extensive CompSol database (the average absolute deviation is 0.30 kcal/mol). The proposed model represents a very flexible alternative for the prediction of solvation data, since it can deal with a large variety of solutes in either a pure or mixed solvent, and for the entire phase diagram (including the supercritical region).

KEYWORDS

COSMO theory; equation of state; solvation; prediction; charge density profile

1 Introduction

Solvation Gibbs energy is defined as *the change in the chemical potential when a molecule devoid of kinetic energy is transferred from the perfect gas phase to a condensed phase* [1]. This thermodynamic property has many applications in chemical and process engineering. For instance, solvation Gibbs energy is a useful information for the selection of the optimal solvent for a separation process or a chemical reaction [2]. In the field of product development, this quantity plays an important role in the design of new drugs [3,4]. Moreover, predictions of solvation Gibbs energy are also required in the development of detailed kinetic models in liquid phase [5–8]

Regarding the theoretical models available to predict solvation quantities, the simplest alternative is the Linear Solvation Energy Relationships (LSER), such as the Abraham model [9,10]. Although being the approach with the lowest computational cost, the Abraham model has the disadvantage to predict solvation quantities at room temperature only (298 K), for a limited set of pure solvents. The mean absolute deviation for this model is usually between 0.6 and 0.9 kcal/mol, depending on the database used for its evaluation [10,11]. Recently, Vermeire and Green [12] proposed a more robust approach based on transfer learning. This method is able to provide very accurate results (mean absolute error of 0.21 kcal/mol), but is also limited to pure solvents at room temperature. To adapt both Abraham model and the transfer learning method for temperatures other than 298 K, the correction proposed by Chung et al. can be used [13].

Among the methods that allow the calculation of solvation quantities in a mixture of solvents, one of the possible alternatives lies in the family of continuum solvation models, like PCM, SMD, and COSMO [11,14–17]. However, these models are restricted to be used with low-pressure systems, far from the critical region. In terms of accuracy, models like COSMO-RS and COSMO-SAC can provide accurate predictions, with average deviations close to 0.3 kcal/mol [11,17]. Molecular dynamics could also be an option [18,19], but the computational cost is much higher. This approach is often used to predict solvation quantities of complex molecules found in drug design.

Moreover, the use of an equation of state (EoS) is among the simplest, fastest and most accurate ways to predict solvation Gibbs energies, given that this quantity can be directly related to the fugacity of the solutes and the density of their solvents, which can be straightforwardly estimated from this type of model [1,20]. The advantage of having a

calculation framework of solvation quantities based on an EoS is the possibility to easily work with co-solvent systems, in the entire phase diagram (low- and high-pressure regions, including supercritical conditions). In addition, an EoS opens the door to the prediction of many other properties, such as solubilities, caloric properties and phase equilibria. The EoS approach has already been investigated in previous works of our group [17,21] for the prediction of solvation Gibbs energies. It was found that predictive versions of cubic EoS, namely PSRK and UMR-PRU, are able to provide results with average deviation below 0.4 kcal/mol.

Both PSRK and UMR-PRU, and other equations of the same type, are coupled with activity coefficient models through the application of an advanced mixing rule. The activity coefficient models embedded in the EoS are usually based on predictive group-contribution methods, and in particular, the UNIFAC (Universal Functional Activity Coefficient) approach [22]. Basically, UNIFAC requires volume and surface area information of the functional groups for the calculation of the combinatorial (or entropic) contribution, in addition to the binary interaction parameters between groups for the calculation of the residual (or enthalpic) contribution.

However, despite the flexibility of UNIFAC for representing a wide variety of systems, some weaknesses can be pointed out [23]. The first one is that the number of functional groups is limited, which can lead to the impossible decomposition of some complex molecules [24]. Another limitation is the need for a large amount of experimental data to fit the group interaction parameters. As a result, not only are many interaction parameters between existing groups still missing [25], but the addition of new groups is a tedious task. Moreover, the current available parameters were usually not fitted to represent experimental data at very low concentrations, which can be a source of error to calculate solvation quantities at infinite dilution [26].

Constantinescu et. al [27] proposed an alternative to circumvent those limitations by using COSMO-RS [28,29] instead of UNIFAC as the activity coefficient model in the mixing rule. Later on, others studies, for instance [30–32], have been conducted using the same approach for the prediction of phase equilibria data. The main advantage of using COSMO-based models is that one can perform thermodynamic calculations of mixtures only based on pure compound information, i.e., no binary interaction parameters are required.

We would like to insist on this major argument in favor of COSMO type models: while for classical versions of cubic and SAFT EoS, it is necessary to select appropriate mixing rules involving binary interaction parameters, a COSMO-based EoS is capable of estimating mixture properties from pure compound input data only (i.e., the critical properties of the species as well as the so-called “ σ -profile” which is the polarization charge density distribution over their molecular surface). In particular, the σ -profile of a given mixture results from a linear combination of the σ -profiles of the species weighted by their molar fraction in the mixture. Note that the σ -profiles are obtained through quantum chemistry calculations, which means that no experimental data is actually needed.

As a result, these advantages allow the treatment of a greater number of systems and molecules. This higher applicability can be illustrated through our previous study [17] dealing with the prediction of the solvation Gibbs energy data present in the CompSol database [20], using two predictive EoS relying on the UNIFAC approach, namely the UMR-PRU and PSRK models. In the same study [17], several experimental datapoints were disregarded as they were related to binary systems in which some UNIFAC interaction parameters between groups were missing. This resulted in the modelling of only 80% of the experimental data present in the CompSol database.

In the present study, our goal is to develop and assess a model coupling a cubic EoS and a COSMO-type model through advanced mixing rules for the estimation of solvation properties. While some similar models were developed for predicting vapor-liquid equilibrium data (VLE), see, e.g., [27,30–32], none of these was developed and calibrated specifically for the description of solvation properties. Such an approach could be a good alternative to UNIFAC-based models enabling to describe most of the systems for which the UNIFAC method cannot be applied (whenever a molecular group of a molecule of interest is not present in the UNIFAC list of groups or a UNIFAC group parameter is missing).

In this context, the present paper outlines a predictive approach to calculate solvation Gibbs energies by coupling a cubic EoS (the *tc*-PR EoS [33,34]) with COSMO-RS [29]; the resulting model is named the *tc*-PR/COSMO-RS EoS. As mentioned above, this model only requires as inputs the critical constants of each molecule (critical, temperature, critical pressure, and acentric factor), as well as the σ -profiles of the compounds, that are obtained from quantum calculations [35].

In addition, in order to enlarge the scope of applicability of the proposed model, it was decided to develop a method for generating the σ -profiles from a group-contribution method.

Such an approach [36,37] is an alternative to quantum calculations when a fast estimation of solvation properties is required. The corresponding version of the *tc*-PR/COSMO-RS EoS is named *tc*-PR/GC-COSMO-RS and the developed group-contribution method uses the same group-decomposition as UNIFAC.

As an important result, 99% of the CompSol database can be described accurately with the *tc*-PR/COSMO-RS while 94% of this database can be simulated with the *tc*-PR/GC-COSMO-RS (this percentage is lower for the simple reason that small molecules and molecules involving unusual atoms or molecular groups cannot be split into UNIFAC groups). In both cases, quite a reasonable accuracy was obtained.

2 Modeling

2.1 The translated-consistent Peng-Robinson (*tc*-PR) EoS

The *tc*-PR EoS [33], where *tc* stands for *translated* [38,39] and *consistent* [40,41] is given in equation (1).

$$P = \frac{RT}{(V + c_i - b_i)} - \frac{a_i(T)}{(V + c_i)(V + c_i + b_i) + b(V + c_i - b_i)}$$

$$\text{with: } \left\{ \begin{array}{l} \eta_c = \left[1 + \sqrt[3]{4 - 2\sqrt{2}} + \sqrt[3]{4 + 2\sqrt{2}} \right]^{-1} \approx 0.25308 \\ b_i = \frac{\eta_c}{\eta_c + 3} \frac{RT_{c,i}}{P_{c,i}} \approx 0.0778 \frac{RT_{c,i}}{P_{c,i}} \\ a_i(T) = \frac{40\eta_c + 8}{49 - 37\eta_c} \frac{(RT_{c,i})^2}{P_{c,i}} \alpha(T_R) \approx 0.45724 \frac{(RT_{c,i})^2}{P_{c,i}} \alpha(T_R) \end{array} \right. \quad (1)$$

T , P , and V are the temperature, pressure and molar volume. $T_{c,i}$, $P_{c,i}$ and ω_i are the experimental critical temperature, critical pressure, and acentric factor of any pure compound i . The *tc*-PR EoS is coupled with the Twu 91 α -function [42] which depends on three compound-specific parameters (namely, L , M , and N) as shown in equation (2).

$$\alpha(T_{r,i}) = T_{r,i}^{N(M-1)} \exp \left[L \left(1 - T_{r,i}^{MN} \right) \right] \quad (2)$$

where $T_{r,i} = T/T_{c,i}$

Such an α -function was selected because it is very flexible and especially because it can be made consistent (that is, pass the consistency test for alpha functions [40]) if the 3 parameters are fitted under constraint.

Optimized values for L , M and N of 1800 compounds can be found elsewhere [34]. When these three parameters are not available for a given compound, the correlations below are used instead [40].

$$\begin{aligned} L(\omega_i) &= 0.0544 + 0.7536\omega_i + 0.0297\omega_i^2 \\ M(\omega_i) &= 0.8678 - 0.1785\omega_i + 0.1401\omega_i^2 \\ N &= 2 \end{aligned} \quad (3)$$

Note that quantum fluids, i.e., He and H₂, are not treated in the same way due to quantum effects [33,40], so the above correlations are not valid. For such molecules, the α -function proposed by Soave [43,44], although not-consistent, can be used instead (see equations (4) and (5)).

$$\alpha(T_{r,i}) = \left[1 + m_i \left(1 - \sqrt{T_{r,i}} \right) \right]^2 \quad (4)$$

$$m_i = 0.37464 + 1.54226\omega_i - 0.26992\omega_i^2 \quad (5)$$

Finally, the constant volume translation of a pure compound (c_i) is determined using either an experimental value of the molar volume at a single temperature, e.g., at $T_r = 0.8$, or a correlation, such as the one given in equation (6) [34,45].

$$c_i = \frac{RT_{c,i}}{P_{c,i}} \times \left(-0.014471 + 0.067498\omega_i - 0.084852\omega_i^2 + 0.067298\omega_i^3 - 0.017366\omega_i^4 \right) \quad (6)$$

2.2 Mixing rules

Pina-Martinez et. al [46] already investigated the extension of the tc -PR EoS to mixtures by means of an advanced mixing rule inspired by the pioneering works of Huron, Vidal and Michelsen [47–50].

For the mixture co-volume (b), they proposed to use a quadratic function of the molar fractions [46]:

$$\begin{aligned} b(\mathbf{z}) &= \sum_{i=1}^{NC} \sum_{j=1}^{NC} z_i z_j b_{ij} \\ \text{with } \left\{ \begin{aligned} b_{ij} &= \left(\frac{b_i^s + b_j^s}{2} \right)^{1/s} \\ s &= 3/2 \end{aligned} \right. \end{aligned} \quad (7)$$

Where, z_i is the composition in terms of molar fraction and NC is the total number of compounds.

A linear mixing rule is assigned to the volume translation parameter (c) of the mixture, based on the volume translation parameters (c_i) of its pure compounds i :

$$c(\mathbf{z}) = \sum_{i=1}^{NC} z_i c_i \quad (8)$$

To express the attractive parameter a , they demonstrated that the ratio of the EoS attractive parameter (a) and the covolume (b) of a mixture can be expressed with respect to pure-component parameters (a_i and b_i) and the residual part of an excess Gibbs energy model ($g_{res}^{E,\gamma}$) as shown in equation (9). By “residual”, it is here meant the *non-combinatorial* (or energetic) part of the excess Gibbs energy model.

$$\frac{a(T, \mathbf{z})}{b(\mathbf{z})RT} = \sum_{i=1}^{NC} z_i \frac{a_i(T)}{b_i RT} + \frac{g_{res}^{E,\gamma}(T, \mathbf{z})}{\Lambda_{EoS} RT} \quad (9)$$

with $\Lambda_{EoS} = -0.6232$

It is recalled that the residual part of the excess Gibbs energy ($g_{res}^{E,\gamma}$) can be expressed as a function of the residual activity coefficients ($\ln \gamma_i^{res}$) through Euler’s theorem:

$$g^{E,res}(T, \mathbf{z}) = RT \sum_{i=1}^{NC} z_i \ln \gamma_i^{res}(T, \mathbf{z}) \quad (10)$$

Different options were tested for expressing the $\ln \gamma_i^{res}$ coefficients. Pina-Martinez et al. [46] obtained promising results using the Wilson [51], NRTL [52], and UNIQUAC [53] activity coefficient models for the prediction of several properties for 200 binary systems [54], such as solubilities, molar enthalpies and heat capacity changes on mixing [55].

As mentioned above, in this work, the $\ln \gamma_i^{res}$ embedded in the tc -PR EoS are generated from the COSMO-RS model for its predictive capacities. The resulting EoS is named tc -PR/COSMO-RS thereafter.

2.3 Calculating $\ln \gamma_i^{res}$ with COSMO-RS

COSMO-RS is a quantum-based method that allows the calculation of chemical potentials (and activity coefficients) in liquid phase. In COSMO-RS, the molecules are placed inside a faceted cavity embedded by a perfect conductor of infinite dielectric constant. In this fictitious

condition, the charge density (σ) of each cavity surface segment is estimated by a COSMO calculation [56]. The probability distribution of this charge density is called σ -profile, noted as $p_i(\sigma)$. In the COSMO-RS version considered in this study [35] the pure compound σ -profile is the only descriptor used for the calculation of the residual contribution $\ln \gamma_i^{res}$. Figure 1 shows some examples of σ -profiles of pure compounds.

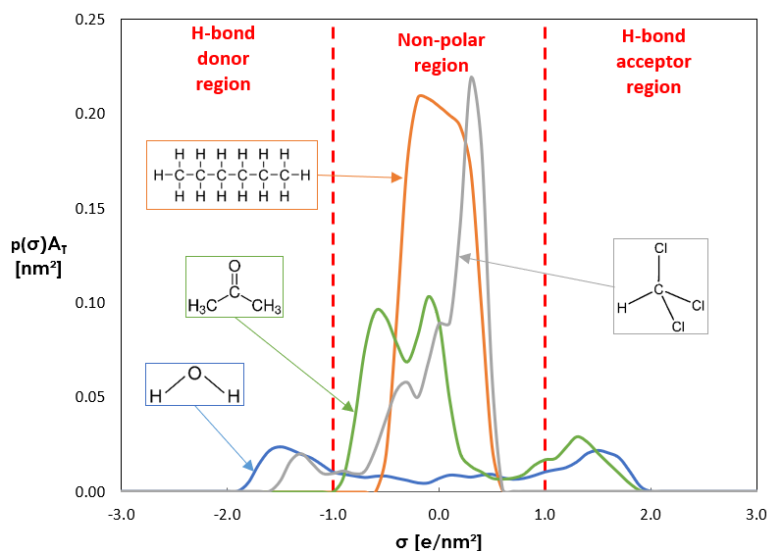


Figure 1 – BP-TZVPD-FINE σ -profiles of *n*-hexane (orange line), chloroform (gray line), acetone (green line), and water (blue line). In this plot, A_T stands for the molecular surface of the compound.

For a multi-component system, the individual normalized σ -profile of each molecule (i.e. normalized by the total molecular cavity surface) is used to compute the normalized σ -profile of the entire mixture, denoted as $p(\sigma)$, through the following mixing rule [28]. Note that the total cavity surface corresponds to moment of order 1 associated to the $p_i(\sigma)A_i$ distribution, where A_i is the molecular cavity of the compound i .

$$p(\sigma) = \sum_{i=1}^{NC} x_i p_i(\sigma) \quad (11)$$

The chemical potential of the segments $\mu(\sigma_m)$, also called σ -potential, is then estimated from statistical thermodynamics by considering all possible segments partners with charge density of σ_m and σ_n and their interaction energy, as shown in equation (12). Examples of σ -potential are given in Figure 2

$$\mu(\sigma_m) = -\frac{RT}{a_{eff}} \ln \left\{ \int_{\sigma_n} p(\sigma_n) \exp \left[\frac{a_{eff}}{RT} (\mu(\sigma_m) - e(\sigma_m, \sigma_n)) \right] d\sigma_n \right\} \quad (12)$$

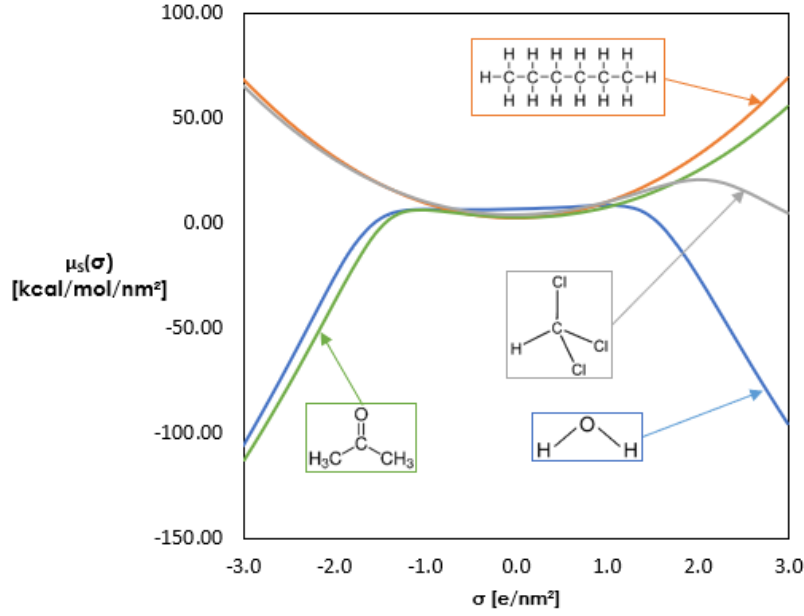


Figure 2 – BP-TZVPD-FINE σ -potentials of *n*-hexane (orange line), chloroform (gray line), acetone (green line), and water (blue line).

In equation (12), $e(\sigma_m, \sigma_n)$ represents the pairwise contact energy of two surface segments with charge density of σ_m and σ_n , and a_{eff} is the effective surface contact between two segments. This contact energy takes into account electrostatic and hydrogen-bonding interaction energies, as denoted in equation (13) [28].

$$e(\sigma_m, \sigma_n) = \frac{\alpha'}{2} (\sigma + \sigma')^2 + c_{hb}(T) \times \min \{ 0, \min \{ 0, \sigma_{acc} + \sigma_{hb} \} \max \{ 0, \sigma_{don} - \sigma_{hb} \} \} \quad (13)$$

with $\begin{cases} \sigma_{acc} = \max \{ \sigma_m; \sigma_n \} \\ \sigma_{don} = \min \{ \sigma_m; \sigma_n \} \end{cases}$

In the equation above, α' denotes the electrostatic misfit coefficient and c_{hb} is a hydrogen-bonding constant at 298 K. The constant σ_{hb} represents a cut-off value for hydrogen-bonding. The coefficient c_{hb} is supposed to be temperature-dependent due to the reduction in entropy that goes along with hydrogen bonding. This temperature dependence can be calculated by equations (14) and (15) [57].

$$f_{hb}(T) = T \ln \left[1 + \frac{1}{200} \exp \left(\frac{20 \text{ kJ/mol}}{RT} \right) \right] \quad (14)$$

$$c_{hb}(T) = c_{hb} \times \frac{f_{hb}(T)}{f_{hb}(298.15 \text{ K})} \quad (15)$$

Note that there is no dispersive contribution in the calculation of $e(\sigma, \sigma')$, because, in this version of COSMO-RS, it is assumed that the Van der Waals forces between two contacting

segments are the same as they were in the perfect conductor (reference state), so they do not need to be taken into account [29]. Finally, the residual contribution to the activity coefficient of a molecule in the mixture is given by the following equation:

$$\ln \gamma_i^{res} = \sum_{\sigma_m} p_i(\sigma_m) \left[\frac{\mu(\sigma_m) - \mu_i(\sigma_m)}{RT} \right] \quad (16)$$

Note that equation (16) above depends on the calculation of the σ -potentials for both the mixture ($\mu(\sigma_m)$) and the pure compound i ($\mu_i(\sigma_m)$). These two calculations are carried out concomitantly by solving equation (12) using a successive substitution method [58]. This characteristic makes activity coefficient calculations with COSMO-RS substantially slower than calculations with UNIFAC. In any case, equation (16) is solved in fractions of a second on any current computer.

2.4 Fast computation of σ -profiles from a group contribution method

In short, the activity coefficient calculations use as input the σ -profile of each compound in the mixture, which are obtained through quantum chemistry calculation. This step takes minutes or hours per molecule, depending on the size and the atoms. For this reason, the σ -profiles are stored in databases, so they can be readily used when required. However, every time a new molecule is not present in those databases, it is necessary to perform the quantum calculation step in order to generate its σ -profile, which can be a drawback in certain applications that demand high-throughput computations of thermodynamic properties.

With this in mind, we established a group contribution (GC) method for a fast prediction of σ -profiles, to be used as an alternative to the DFT calculation. The approach used in this work is the same as the one proposed by Mu and Gmehling [36,37]. In this formulation, σ -profiles are calculated as follows:

$$p_i(\sigma_m) A_i = \sum_k n_k p_{\sigma_m,k} \quad (17)$$

Where, n_k is the number of occurrences of the group k in the molecule i . The constant $p_{\sigma_m,k}$ is the contribution of the group k in the calculation of the σ -profile. Note that a σ -profile corresponds to a discrete histogram. That is why there is a $p_{\sigma_m,k}$ for each discrete value of σ_m , in the interval between σ_{min} and σ_{max} (-3.0 and $+3.0$ e/nm², respectively).

The version of the *tc*-PR/COSMO-RS EoS using σ -profile predicted from the GC method proposed here is named *tc*-PR/GC-COSMO-RS.

2.5 Calculation of Solvation quantities

The solvation Gibbs energy ($\Delta_{solv}\bar{g}_i$) of a given molecule i in a condensed phase can be predicted by an EoS through equation (18) [1,21].

$$\Delta_{solv}\bar{g}_i(T, P, \mathbf{z}) = RT \ln \left[\frac{P \cdot V(T, P, \mathbf{z}) \cdot \varphi_i(T, P, \mathbf{z})}{RT} \right] \quad (18)$$

This approach relies on the calculation of the fugacity coefficient (φ_i) of the solute i and the molar volume of the bulk phase (V), both estimated from the proposed EoS.

Note that the solvation Gibbs energy at infinite dilution ($\Delta_{solv}\bar{g}_i^\infty$) is obtained simply by taking the limit of $\Delta_{solv}\bar{g}_i$ when the molar fraction of the solute i tends to zero.

$$\Delta_{solv}\bar{g}_i^\infty(T, P) = \lim_{z_i \rightarrow 0} \Delta_{solv}\bar{g}_i(T, P, \mathbf{z}) \quad (19)$$

The solvation entropy ($\Delta_{solv}\bar{s}_i$) is calculated as the opposite of the temperature-derivative of the solvation energy, as follows. Such a derivative is obtained numerically by finite differences.

$$\Delta_{solv}\bar{s}_i = - \left(\frac{\partial \Delta_{solv}\bar{g}_i}{\partial T} \right)_{P, \mathbf{x}} \quad (20)$$

Finally, the solvation enthalpy ($\Delta_{solv}\bar{h}_i$) is calculated by equation (21).

$$\Delta_{solv}\bar{h}_i = \Delta_{solv}\bar{g}_i + T \Delta_{solv}\bar{s}_i \quad (21)$$

3 Technical details

3.1 Experimental data for solvation energies

The CompSol database [20] was considered as reference data in this work. This extensive database comprises 69,302 experimental solvation Gibbs energies for 14,102 binary systems at infinite dilution condition, involving a wide variety of molecules and temperature conditions.

3.2 Inputs for pure compounds

The experimental critical temperature, critical pressure, and acentric factor of molecules were extracted from the DIPPR database. For most of the molecules, the optimized parameters for the Twu-91 α -function were obtained from [34] (if not available, equation (3) was used instead).

The quantum calculated σ -profile of molecules were obtained from the databases provided by the commercial software COSMOTerm C30 v.16. In this database, the quantum calculations were performed with TURBOMOLE v.14 using the “BP” DFT functional and TZVPD basis set with the refined cavity construction (called FINE cavity) [59]. In the present study, only the most stable conformational structure was considered.

Note that the performances of the *tc*-PR/COSMO-RS model will be assessed using two different data sets:

- The data set containing all the data that can be modelled with the *tc*-PR/COSMO-RS EoS (called *data99*),
- The data set containing the data that can be modelled both with the *tc*-PR/COSMO-RS and the *tc*-PR/GC-COSMO-RS EoS (called *data94*).

It is recalled that the developed group-contribution method uses the same group-decomposition as UNIFAC so that only molecules that can be decomposed into UNIFAC groups were considered in *data94*. The decomposition of molecules was performed automatically using the software SIMULIS 2.0.40 using the same groups defined for UMR-PRU, with the addition of 4 groups representing small molecules (formaldehyde, formic acid, hydrogen peroxide, and ethenone). In total, 134 molecular groups are present in the group-contribution method for σ -profile present in the *tc*-PR/GC-COSMO-RS EoS.

3.3 Filters applied to the reference database

It is important to clarify that the CompSol databank for binary systems comprises 2,023 molecules (solutes and solvents), but not all of them could be treated for various reasons previously mentioned and reported in Table 1 (see the column “Filters”). The number of molecules present after application of each of these filters are provided.

Table 1 – Filters applied to molecules of the CompSol database

Filters	Number of molecules	Number of datapoints	Name of the corresponding database
CompSol (no filters)	2,023 (100%)	69,302 (100 %)	CompSol
(1) The molecule has known experimental T_c , P_c , and ω in the DIPPR database.	1,969 (97.3%)	69,132 (99.8%)	-
(2) The σ -profiles are available in the database of the COSMOTerm	1,912 (94.5%)	68,372 (98.7%)	-
Combination of filters (1) and (2)	1,911 (94.5%)	68,317 (98.6%)	<i>data99</i>

(3) The decomposition of molecules into UNIFAC functional groups is possible	1,597 (78.9%)	65,120 (94.0%)	-
All 3 filters combined	1,597 (78.9%)	65,120 (94.0%)	<i>data94</i>

In total, it is possible to simulate almost 99 % of the solvation Gibbs energy data of the CompSol database with the *tc*-PR/COSMO-RS model developed here using as input the quantum-calculated σ -profile present in the COSMOtherm database and the (T_c, P_c, ω) datasets extracted from the DIPPR.

For using the predictive version of the EoS (*tc*-PR/GC-COSMO-RS), the decomposition of the binary-system molecules into UNIFAC functional groups must be possible. This justifies the application of a third filter (see Table 1). Including now this filter, 94 % of the solvation Gibbs energy data of the CompSol database can be simulated with the *tc*-PR/GC-COSMO-RS EoS. For the generation of results for both models (*tc*-PR/COSMO-RS and *tc*-PR/GC-COSMO-RS), this value of 94 % (65,120 experimental points involving 1,597 compounds) was kept in order to maintain the same basis for comparison. In any case, this value of 94% is much higher than the one obtained with the UMR-PRU and PSRK EoS, which was around 80 % [17], mainly due to missing group binary interaction parameters in the UNIFAC matrices.

3.4 Model fitting

3.4.1 COSMO-RS parametrization

One of the main advantages of COSMO-RS is the low number of parameters. In the version used in this work (see section 2.3), we have only 4 universal parameters (a_{eff} , α' , c_{hb} , and σ_{hb}), the values of which have been estimated in the open literature at low pressure (and from other quantum packages) [28,60]. Following the conclusions outlined by Piña-Martinez et al., in the paper [46] where the advanced mixing rule used in this study was developed, it was decided to refit all 4 universal parameters of COSMO-RS in order to improve the accuracy of the model. For this purpose, these 4 constants were fitted based on the experimental solvation Gibbs energy data available in the CompSol database. A minimization procedure based on the Nelder-Mead algorithm was applied to the objective function OF expressed as the sum of the quadratic deviations between experimental and calculated data, as shown in equation (22).

$$OF = \sum_{k=1}^N \left[\left(\Delta_{solv} \bar{g}_i \right)_k^{\text{exp}} - \left(\Delta_{solv} \bar{g}_i \right)_k^{\text{calc}} \right]^2 \quad (22)$$

Where $(\Delta\bar{g}_i^\infty)_k^{exp}$ and $(\Delta\bar{g}_i^\infty)_k^{calc}$ are the experimental and calculated values of the Solvation Gibbs energy for the datapoint k , and N is the total number of experimental points. The decision variables of such an optimization were the 4 universal parameters mentioned above:

- The effective contact area (a_{eff}),
- The misfit interaction parameter (α'),
- The hydrogen-bonding interaction parameter (c_{hb}),
- The hydrogen-bonding cut-off charge density (σ_{hb}).

Note that the optimization procedure was carried out in a homemade code written in Matlab[®] considering as input data the σ -profiles of the various solutes and solvents obtained from quantum calculations (GC-based σ -profiles were not considered in the optimization procedure). This is because the group contribution concept represents an alternative to quantum calculations for molecules whose σ -profiles is unknown. In this sense, by maintaining a unique COSMO-RS parameterization, it is possible to use quantum and GC-based σ -profiles in a same calculation framework.

3.4.2 Group contribution method for σ -profiles

The contribution of each group for the calculation of σ -profiles were determined from multilinear regression using, here again, a homemade code in Matlab[®] language. The reference data for such a regression were the quantum-calculated σ -profiles obtained from the COSMO files available in the database of the software COSMOTerm.

4 Results and discussions

The results discussed in the following sections are provided in the supplementary data (Appendix A) for both databases considered in this study (*data99* and *data94*). A database containing 1613 molecules (with UNIFAC decomposition), as well as the group contribution parameters for σ -profiles are also provided in the supplementary data for the use of the predictive version of the model (tc-PR/GC-COSMO-RS).

4.1 Predictions of solvation quantities with tc-PR/COSMO-RS

4.1.1 General results

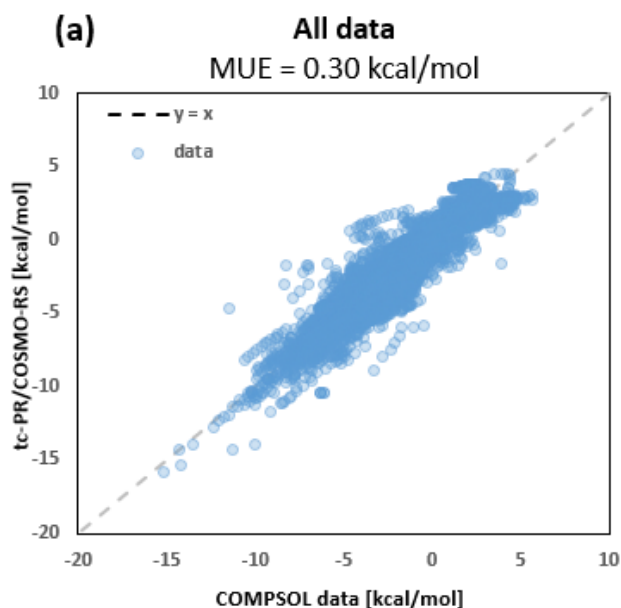
Table 2 gives the parametrization of COSMO-RS determined within this study. This parametrization agrees with others available in the literature [28,60]: indeed, while the universal constants of COSMO-RS were fitted over data at very different pressure conditions, their orders of magnitude are similar, as highlighted by Table 2.

Table 2 – COSMO-RS parameters fitted in this work (BP-TZVPD-FINE).

Parameter	Unit	Initial guess	Final value
a_{eff}	\AA^2	6.15 ^a	4.73
α'	kJ/mol/\AA^2	6635 ^a	10473
c_{hb}	kJ/mol/nm^2	33800 ^a	48321
σ_{hb}	e/\AA^2	0.0084 ^a	0.0109

a. parametrization proposed in [28]

Figure 3 (a) provides statistical information about the quality of the predictions of solvation Gibbs energies using the *tc*-PR/COSMO-RS EoS combined with quantum calculated σ -profiles. Altogether, 64,360 experimental points from the CompSol database (from the so-called *data94* database in this paper), were predicted with the proposed EoS. As it can be seen, the *tc*-PR/COSMO-RS EoS is able to provide reasonable predictions of solvation energies for most of the systems, since the average deviation is only 0.29 kcal/mol for the entire dataset. Regarding the error distribution, Figure 3 (b) points out that errors are normally distributed with a mean of zero indicating that there is no trend for systematic under or overprediction of the experimental data.



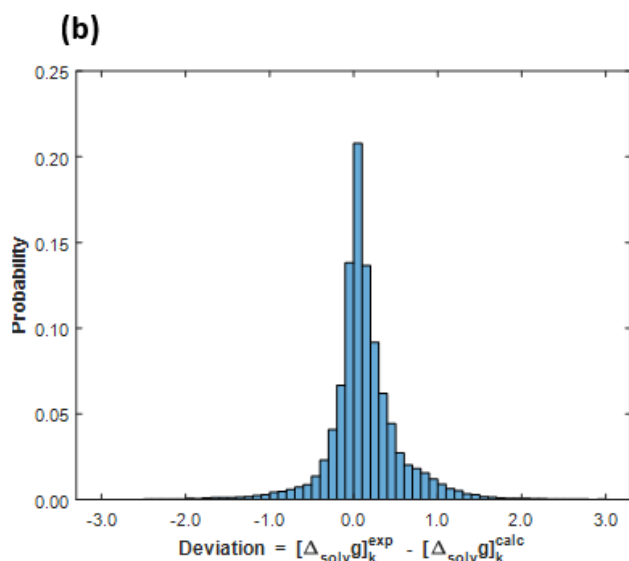


Figure 3 – Parity plot (a) and error distribution histogram (b) concerning the predictions of the solvation Gibbs energies with *tc*-PR/COSMO-RS and quantum calculated σ -profiles. Results obtained for a dataset of 68,317 experimental points (temperature conditions going from 64.83 up to 653.15 K). MUE stands for the mean unsigned error.

The proposed EoS has similar performance to other approaches that were tested on the CompSol database in previous works, as indicated in Table 3.

Table 3 – Mean unsigned error (MUE) and standard deviation (SD) between calculated and experimental data from the CompSol databank involving different models.

Model	Number of data	MUE (kcal/mol)	SD (kcal/mol)	Reference
<i>tc</i> -PR/COSMO-RS	68,317 (<i>data99</i>)	0.30	0.37	This work
<i>tc</i> -PR/COSMO-RS	65,120 (<i>data94</i>)	0.29	0.36	This work
<i>tc</i> -PR/GC-COSMO-RS	65,120 (<i>data94</i>)	0.38	0.45	This work
UMR-PRU	56,603	0.34	0.65	[17]
PSRK	56,603	0.32	0.49	[17]
COSMO-RS ^a	56,603	0.29	0.35	[17]
COSMO-SAC-dsp	56,603	0.35	0.63	[17]
LSER (Abraham method) ^b	-	0.81	-	[11]

a. COSMOTherm implementation of COSMO-RS (BP_TZVPD_FINE_C30_1401)

b. Data at 298 K

Table 3 shows that only 56,603 data points were considered in our previous study [17]. This difference in terms of number of points considered is mainly due to missing group binary interaction parameters in the UNIFAC matrices used with the UMR-PRU and PSRK models.

4.1.2 Effect of the nature of the solvent

Going into detail on the results, we found that the proposed EoS has a good performance for organic solvents. In most of the cases, the average deviation was significantly smaller than that obtained for hydration data. In order to better illustrate this observation, Figure 4 compares different parity diagrams involving several families of organic solvents with respect to the results obtained for the solvation of solutes in water (see figure at bottom right).

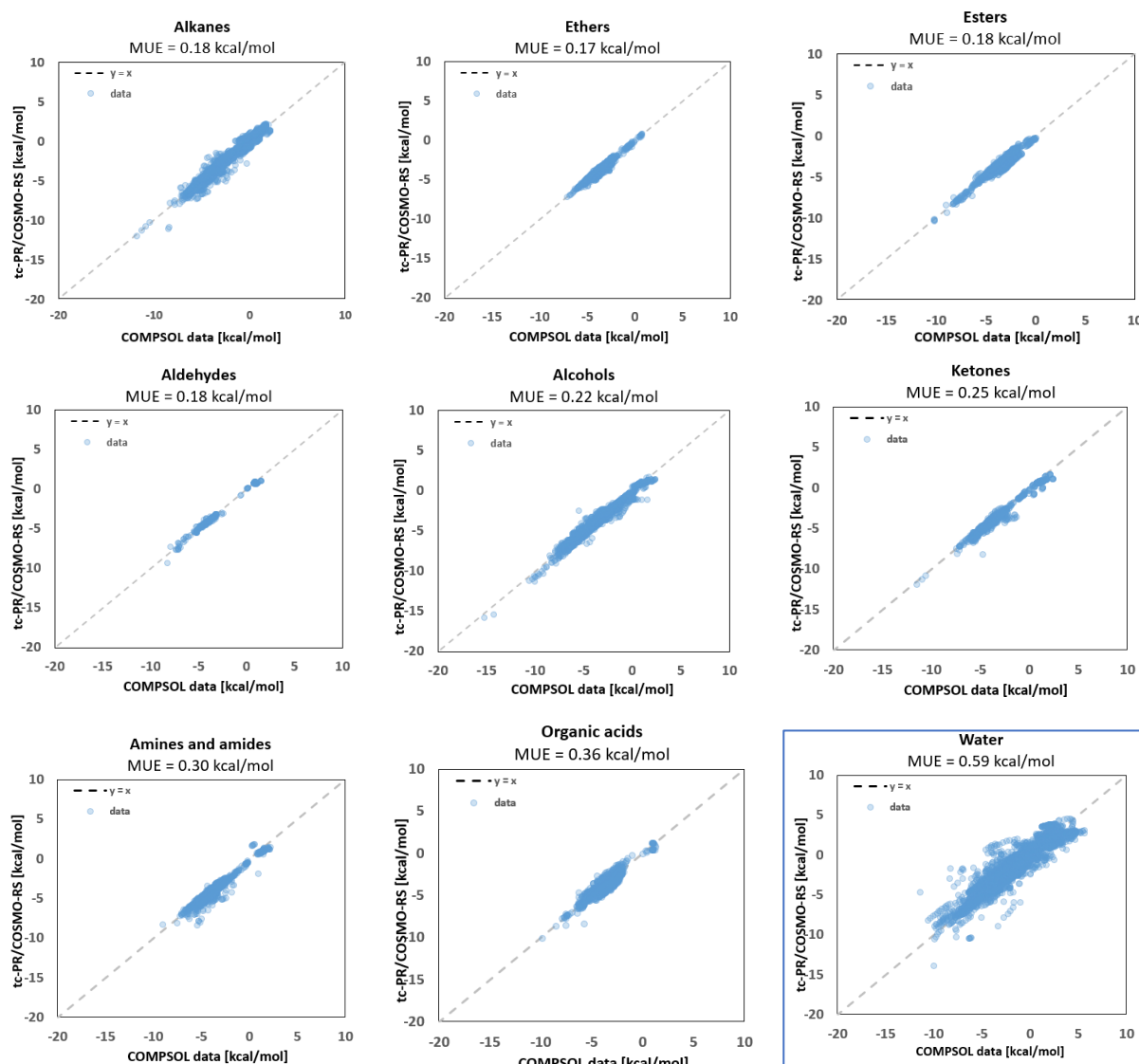


Figure 4 – Parity plots for the prediction of the solvation Gibbs energy with *tc-PR/COSMO-RS (BP-TZVPD-FINE)*, involving different families of organic solvents and water.

Actually, it is always a challenge for any model to predict hydration data, especially for cubic EoS. For instance, when water solvates non-polar solutes, the resulting liquid phase tends to be strongly non-ideal and unstable, with phase-split often observed. Moreover, prediction of hydration data involving associating solutes can be also a difficult task, since water is able to form a complex hydrogen-bonding network that is difficult to correlate with most of thermodynamic models. The same behavior is also observed for some self-associating organic solvents. Figure 4, shows that, like water, strongly associating organic solvents, such as carboxylic acids and amines, have also higher average deviation when compared to the other organic families of solvents.

This discussion puts forward the link between the complexity of the association behavior (ability to form hydrogen bonds) with the quality of the predictions. In order to provide a

more systematic evaluation on this topic, Table 4 shows the deviation results for the predictions of solvation Gibbs energies ordered on the basis of the degree of association of the binary systems. Four types of binary mixtures can be identified [17,46,54,61,62]: (1) Non-associating mixtures; (2) Mixtures in which only self-association takes place; (3) Mixtures in which only cross-association takes place; and (4) mixtures in which both cross and self-association take place. Note that only association by hydrogen bonding is considered in this study. Chemical association by covalent bonding that leads the formation of dimers or trimers is beyond the scope of this study, since this also requires chemical equilibrium calculations [63].

Table 5 and Table 6 present the solvation entropy and enthalpy results following the same classification of binary systems.

Table 4 - Mean unsigned error (MUE) and standard deviation (SD) for the prediction of solvation Gibbs energies ($\Delta_{solv}\bar{g}_i$) with tc-PR/COSMO-RS.

Association behavior	N_{points}	MUE [kcal/mol]	SD [kcal/mol]
Non-associating	34,269	0.21	0.34
Self-association only	16,255	0.39	0.58
Cross-association only	1,148	0.41	0.57
Cross and self-association	16,645	0.37	0.58
All data	68,317	0.30	0.37

Table 5 - Mean unsigned error (MUE) and standard deviation (SD) for the prediction of solvation entropies ($\Delta_{solv}\bar{s}_i$) with tc-PR/COSMO-RS.

Association behavior	N_{points}	MUE [cal/mol/K]	SD [cal/mol/K]
Non-associating	17,643	1.51	2.82
Self-association only	5,693	3.04	5.43
Cross-association only	398	3.63	5.27
Cross and self-association	4,812	3.71	6.38
All data	28,546	2.22	3.62

Table 6 - Mean unsigned error (MUE) and standard deviation (SD) for the prediction of solvation enthalpies ($\Delta_{solv}\bar{h}_i$) with tc-PR/COSMO-RS.

Association behavior	N_{points}	MUE [kcal/mol]	SD [kcal/mol]
Non-associating	17,643	0.56	1.03
Self-association only	5,689	1.05	1.92
Cross-association only	398	1.49	2.12
Cross and self-association	4,812	1.39	2.31
All data	28,542	0.81	1.30

Broadly speaking, non-associating mixtures can be well-predicted by the proposed model (absolute average deviation for $\Delta_{solv}\bar{g}_i$ is only 0.2 kcal/mol for this type of mixture). However, when association takes place, the deviation between experimental and calculated data is, in average, almost twice as bigger. Table 5 also shows that the model maintained the same level of accuracy for all four types of association regarding the prediction of solvation entropies. The quality of the prediction, in this case, depends on how well the model can represent the temperature dependence of the solvation Gibbs energy data. Overall, the temperature dependence is well represented by the proposed EoS, for all types of association behavior. With respect to the solvation enthalpy, Table 6 pinpoints that the model is better able to correlate data relating to non-associating binary mixtures, as was the case with the solvation Gibbs energy. This can be explained by the fact that the solvation enthalpy is calculated indirectly from the solvation Gibbs energy and solvation entropy, so the error on the enthalpy is actually the cumulation of the error obtained for the other two solvation quantities.

4.1.3 Predictions of solvation Gibbs energies in supercritical CO₂

In this section, it is desired to highlight one of the big advantages offered by the coupling of a cubic EoS and a COSMO approach which is the possibility to account for the effect of pressure on fluid phases (liquid, gaseous and supercritical). In that regard, Table 7 and Figure 5 presents solvation Gibbs energy predictions for three pharmaceutical molecules (aspirin, ibuprofen, and carbamazepine) infinitely diluted in supercritical CO₂ at high-pressure conditions.

Note that the group contribution method proposed by Marrero and Gani [64] was used in order to estimate $T_{c,i}$, $P_{c,i}$, and ω_i of carbamazepine, since there are no experimental data available. For aspirin and ibuprofen, $T_{c,i}$, $P_{c,i}$, and ω_i are available in the DIPPR database.

Table 7 - Mean unsigned error (MUE) and standard deviation (SD) for the prediction of the solvation Gibbs energies ($\Delta_{solv}\bar{g}_i$) of aspirin, ibuprofen, and carbamazepine in supercritical CO₂ using tc-PR/COSMO-RS.

compound	N _{points}	MUE (kcal/mol)		SD (kcal/mol)	
		tc-PR/COSMO-RS	COSMO-RS	tc-PR/COSMO-RS	COSMO-RS
aspirin	31	0.58	4.22	0.20	0.92
ibuprofen	29	0.32	3.31	0.26	0.92
carbamazepine	23	0.74	7.30	0.18	0.92
all data	83	0.53	4.75	0.28	1.87

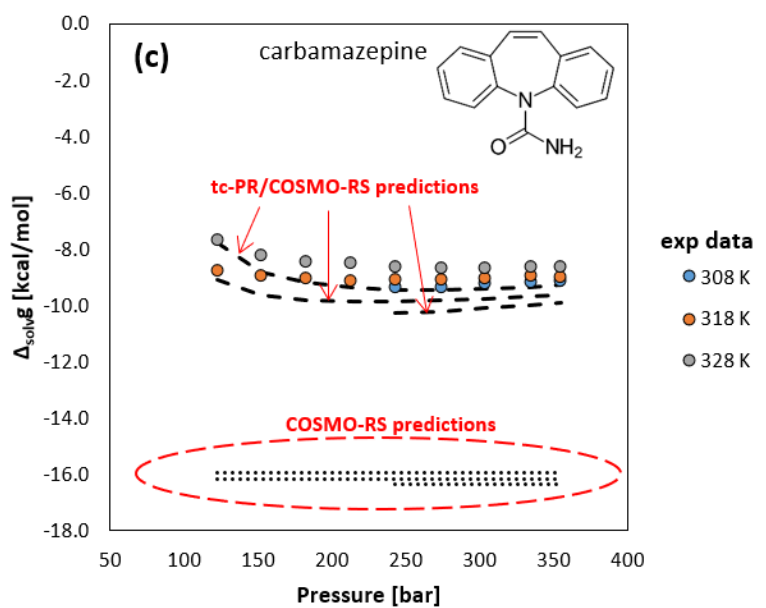
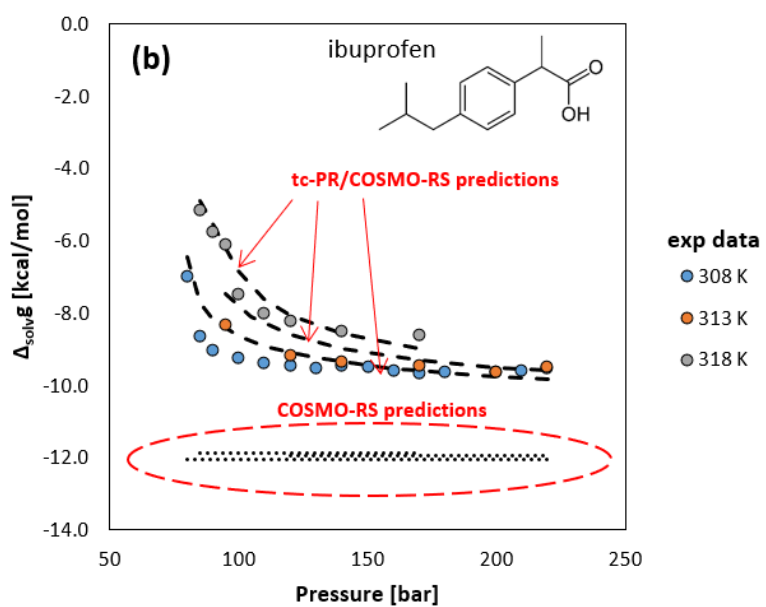
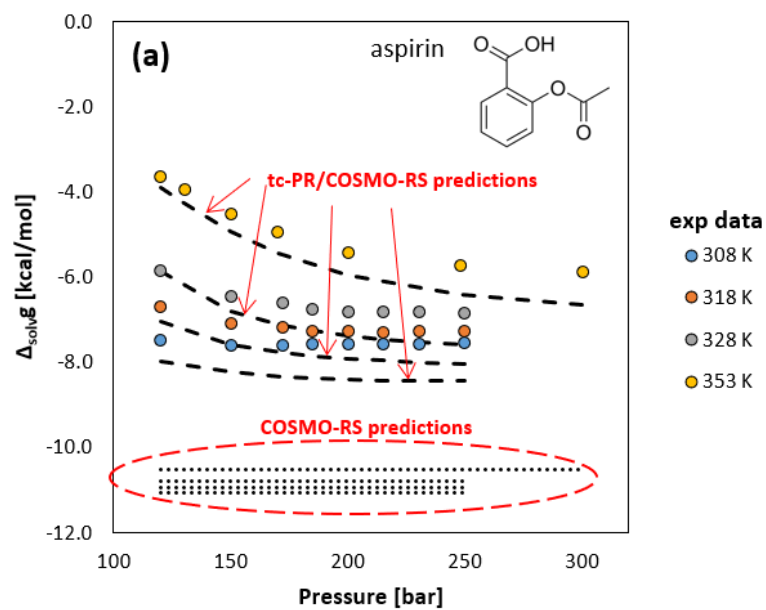


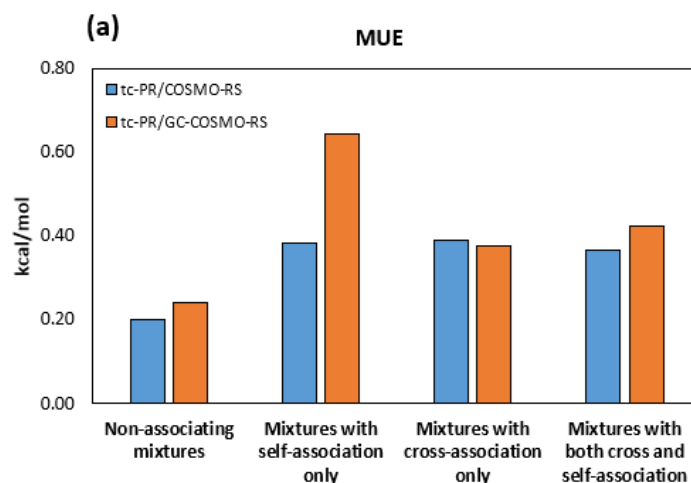
Figure 5 – Solvation Gibbs energy of aspirin (a), ibuprofen (b), and carbamazepine (c) in supercritical CO₂ as a function of pressure. Experimental data extract from [65]. The calculations were performed with quantum-calculated σ -profile (BP-TZVPD-FINE) for both COSMO-RS and *tc*-PR/COSMO-RS. For the COSMO-RS calculation the software COSMOTerm was used (BP_TZVPD_FINE_C30_1401).

It can be seen that predictions in the supercritical region are beyond the scope of a COSMO-based model alone. This is because in such a condition, the solvating phase is no longer an incompressible fluid, so that the solvation Gibbs energy becomes pressure-dependent, as illustrated by the experimental data reported in Figure 5. While the COSMO-RS model cannot account for pressure effects, its coupling with a cubic equation expands the model applicability, allowing reliable predictions in supercritical solvents, even in this example involving the solvation of complex pharmaceutical molecules.

4.2 Performance of the group-contribution version (*tc*-PR/GCM-COSMO-RS) for the prediction of solvation Gibbs energies

Let us look now at the performance of the model to predict solvation Gibbs energies when the group-contribution approach is used to generate the σ -profiles of all molecules (solutes and solvents). In the below results, only the data94 database is used in order to benchmark the *tc*-PR/COSMO-RS and *tc*-PR/GC-COSMO-RS EoS on the same basis.

According to Figure 6 and as expected, a deterioration on the prediction quality is observed. This increase in the mean deviation is mainly due to molecules that experience a degradation of their σ -profile when predicted from the group contribution concept.



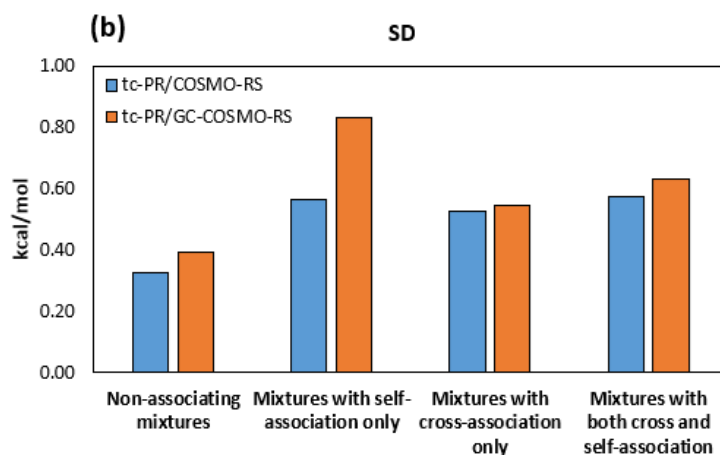
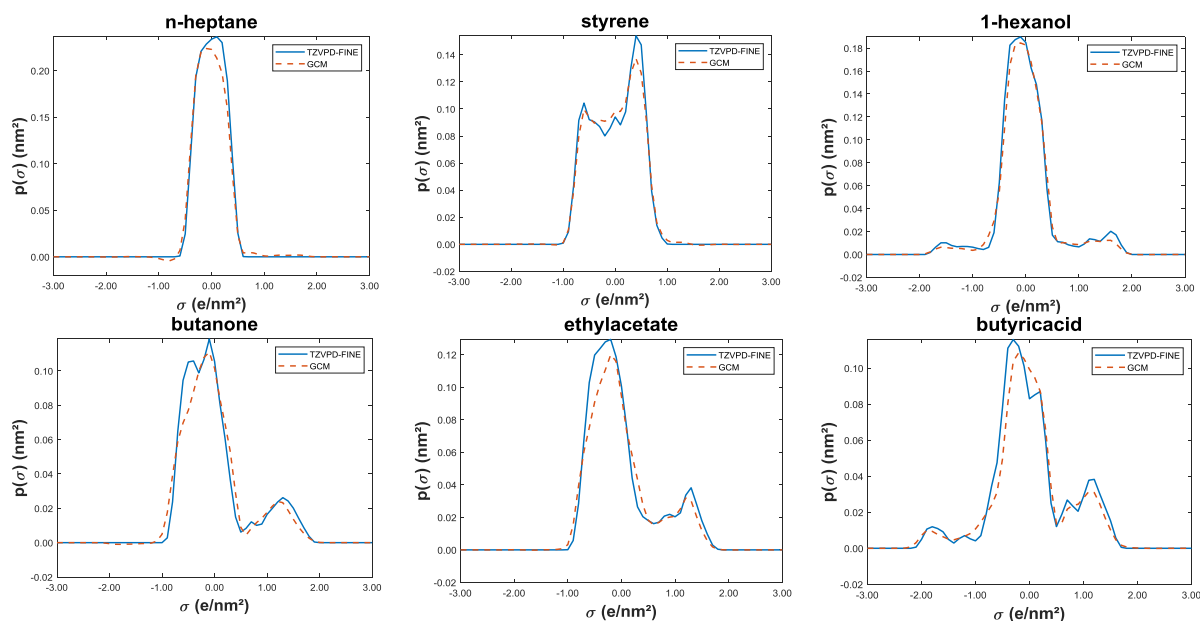


Figure 6 – Mean unsigned error (MUE), in (a), and standard deviation (SD), in (b), for the prediction of solvation Gibbs energies using either quantum-based method (tc-PR/COSMO-RS) or group contribution (tc-PR/GC-COSMO-RS) to obtain the σ -profiles. Results obtained for a dataset of 65,120 experimental points.

In general terms, C/H/O molecules and most of amides and amines are well-represented through this approach, and larger carbon chain molecules especially. As highlighted by Mu and Gmehling [36], “good results are obtained for alkanes, aromatics, alkenes, esters, ethers, ketones, epoxides, anhydrides, carboxylic acids, etc.” On the other hand, poor predictions are achieved for nitrobenzenes and sulfides compounds, for example. To illustrate these comments, Figure 7 presents some σ -profiles calculated using the group contribution method developed here.



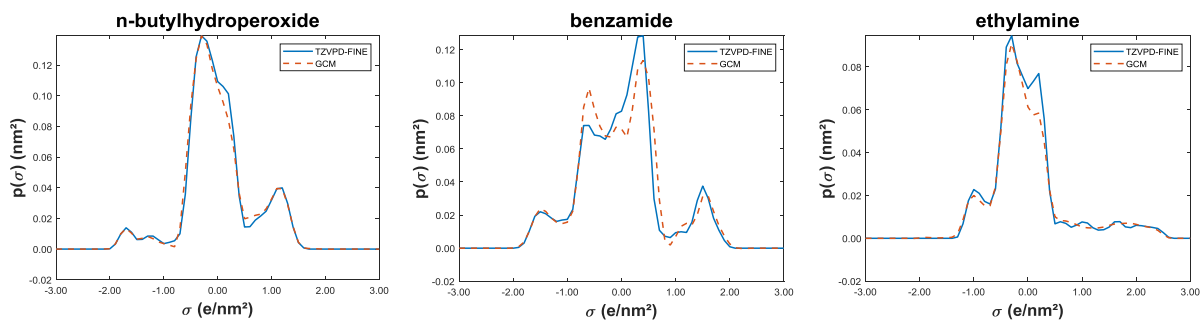
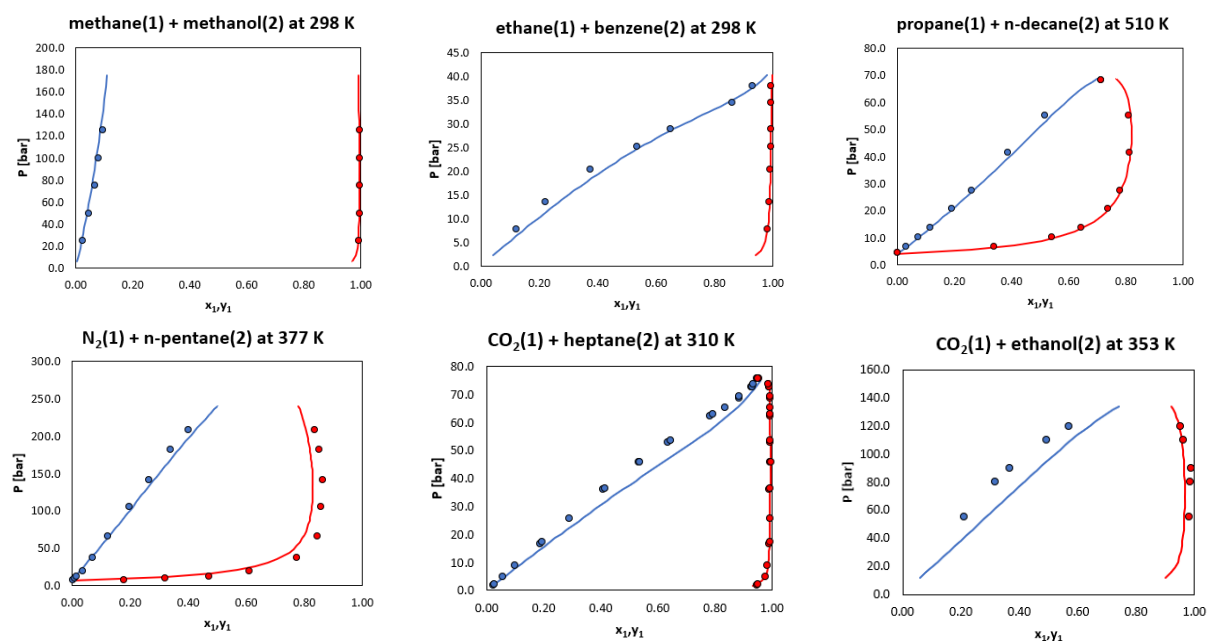


Figure 7 – Example of σ -profiles estimated by group contribution in comparison with the quantum calculated σ -profiles (BP-TZVPD-FINE).

Despite the slight decrease in accuracy, one should keep in mind that the group-contribution method counts for an alternative method that allows the simulation of systems that contain molecules with unknown σ -profile, and for which, quantum calculations would be too costly from a computational viewpoint.

4.3 Evaluating the performance of *tc*-PR/COSMO-RS for phase-equilibrium predictions

The *tc*-PR/COSMO-RS EoS is also able to predict other properties than solvation energies. Indeed, EoS models are used in many applications for evaluating the energy balance and producing phase-equilibrium diagrams. As far as the latter is concerned, Figure 8 reports some phase-equilibrium diagrams at constant temperature (isothermal pressure versus phase compositions diagrams, denoted P_{xy}) calculated with the *tc*-PR/COSMO-RS model, using quantum-calculated σ -profiles.



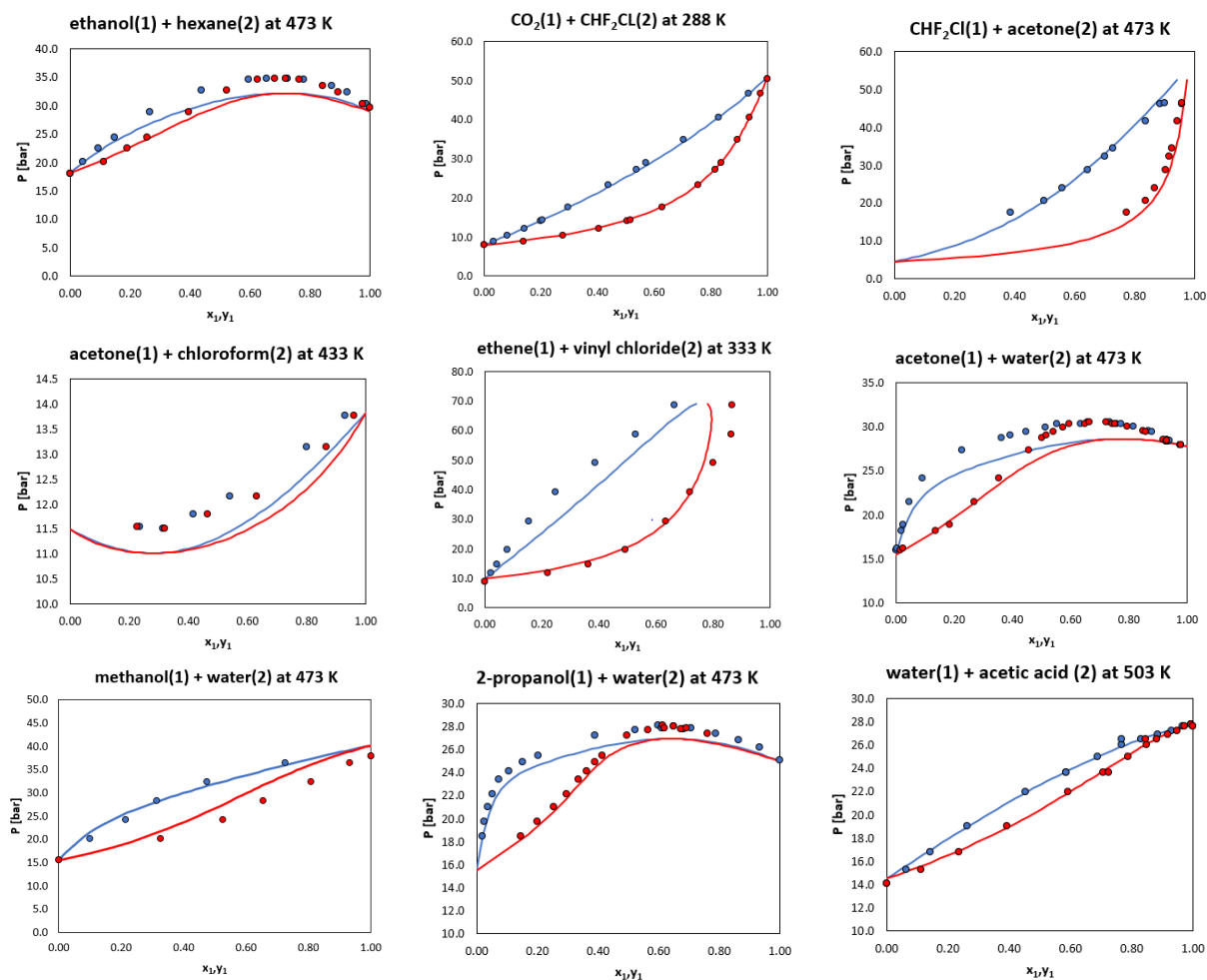


Figure 8 – Examples of phase equilibria predictions with *tc*-PR/COSMO-RS at constant temperature and moderate to high-pressure (*Pxy* diagrams)

As it can be observed, the model is still able to provide reasonable predictions of phase equilibria, at least from a qualitative point of view and often, from a quantitative point of view as well. Even for the cases where we observe a less accurate correlation between experimental and calculated data (see the diagram involving acetone and chloroform, for example) the trends of the equilibrium curves are still coherent, and not so far from the experimental data. At this step of the discussion, it must be recalled that the mixing rule used here is purely predictive, i.e, it does not require any fitted binary interaction parameters (only pure compound σ -profiles are used as input information to capture binary interactions).

In order to improve the results regarding the calculation of phase equilibrium data, the following actions could be made:

- Fit the universal constants of COSMO-RS using phase equilibrium data in the objective function represented by equation (22). Here, we only used solvation Gibbs

energy at infinite dilution condition, since the main aim of this work is to predict this quantity in particular. Thus, the model tends to work better at low concentrations.

- Add dispersive interactions to the contact energy term in equation (13) in the condensed phase, since only electrostatic and hydrogen-bond interactions were considered in order to maintain a small number of parameters and inputs for COSMO-RS calculations. For molecules for which the input COSMO file is available, the addition of dispersive interactions can be done using element-specific parameters (see equations (1-3) of [28], and the approach introduced in [66]). If it is not the case, a simplified group-contribution method dedicated to these dispersive interactions can be implemented, as in the F-SAC-Phi model [67].
- Change for a more adequate hydrogen-bond description. In this study, for the sake of simplicity, we have opted for a reduced description of hydrogen-bonding. More recent versions of the COSMO-SAC model, for example [66,68,69], use more complete descriptions of the association phenomenon.

All in all, coupling COSMO-RS with a cubic EoS also enlarges the application range of COSMO-type models to the prediction of VLE data of pressurized systems, including compounds at supercritical condition. Taken together, the results confirm that such an approach can be an alternative to the activity coefficient models that are usually applied in the advanced mixing rules.

5 Conclusions

In this paper, a new model based on an EoS/ g^E approach is introduced to predict solvation energies and phase diagrams of multicomponent (binary and beyond) systems in sub and supercritical states. It has been benchmarked over a large database containing solvation data of binary systems at infinite dilution conditions and some series of isothermal VLE data. This model combines the *tc*-PR EoS with COSMO-RS, thus resulting in a predictive model that requires only pure component information as input (T_c , P_c , ω and the σ -profiles).

The main advantages offered by this model with respect to the existing EoS are:

- The possibility to describe any multicomponent system provided the (T_c , P_c , ω) dataset is known and the σ -profiles of species can be generated or are available in a database. In the present study, switching from classical approaches (UMR-PRU, PSRK EoS) to the one proposed here enables to cover 19%pt. more data of the CompSol database.

- No binary interaction parameter is needed. Only pure-component σ -profiles are required to generate the σ -profile of any mixture. This is a huge advantage with respect to all other EoS-based approaches.

With respect to the COSMO-based approaches, the *tc*-PR/COSMO-RS model is applicable in the complete fluid phase region (including the supercritical domain) and is capable of accounting for pressure effects, regardless of the nature of the studied fluid (liquid, gaseous or supercritical).

The results of this study indicate that this method can provide accurate predictions of Gibbs solvation energies ($MUE = 0.30$ kcal/mol); the obtained accuracy is comparable (and even better) to those of widely used predictive cubic EoS such as PSRK and UMR-PRU. It offers a prediction capacity quite similar to the one offered by the COSMO-RS version implemented in COSMOTerm software while extending its range of application to pressurized states, including the supercritical region.

A group-contribution approach was also developed for a fast estimation of σ -profiles. This allows to avoid the time-consuming quantum calculations necessary to generate these σ -profiles. This issue concerns more particularly the biggest molecules (proteins, pharmaceuticals ...) for which quantum calculations are excessively long.

As a final comment, this approach opens the way to the thermodynamic description of *unusual* molecules such as radicals or molecular structures involved in transition states. In a coming article, this method will be extended to the prediction of solvation properties for radical species with the aim to be able, in fine, to model the kinetics of liquid systems in which oxidation takes place.

Acknowledgements

We are thankful to the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 101003318) for the financial support and scholarships.

This work was granted access to the HPC resources of IDRIS under the allocation AD010812434R1 made by GENCI.

Appendix A. Supplementary data

The following information is provided in a supplementary material file:

- The experimental solvation data associated with the *data99* and *data94* databases, as well as the corresponding predictions from the tc-PR/COSMO-RS and tc-PR/GC-COSMO-RS models and the resulting deviations,
- The group contribution parameters for the prediction of σ -profiles are also provided as supplementary data,
- The (T_c, P_c, ω) and (L, M, N) [i.e., the parameters of the alpha function of the tc-PR EoS] datasets for all the pure compounds involved in this study as well as their UNIFAC decomposition.

References

- [1] A. Ben-Naim, Solvation thermodynamics, Plenum Press, New York, 1987.
- [2] N.D. Austin, N.V. Sahinidis, I.A. Konstantinov, D.W. Trahan, COSMO-based computer-aided molecular/mixture design: A focus on reaction solvents, *AIChE J.* 64 (2018) 104–122. <https://doi.org/10.1002/aic.15871>.
- [3] M.L. Verteramo, O. Stenström, M.M. Ignjatović, O. Caldaru, M.A. Olsson, F. Manzoni, H. Leffler, E. Oksanen, D.T. Logan, U.J. Nilsson, U. Ryde, M. Akke, Interplay between Conformational Entropy and Solvation Entropy in Protein–Ligand Binding, *J. Am. Chem. Soc.* 141 (2019) 2012–2026. <https://doi.org/10.1021/jacs.8b11099>.
- [4] Y. Zhao, W. Liu, J. Zhu, H. Zhang, X. Pei, Solvent affinity and its applications in the prediction of mutual solubility, *J. Mol. Liq.* 343 (2021) 117700. <https://doi.org/10.1016/j.molliq.2021.117700>.
- [5] A. Jalan, R.H. West, W.H. Green, An extensible framework for capturing solvent effects in computer generated kinetic models, *J. Phys. Chem. B.* 117 (2013) 2955–2970. <https://doi.org/10.1021/jp310824h>.
- [6] B.L. Slakman, R.H. West, Kinetic solvent effects in organic reactions, *J. Phys. Org. Chem.* 32 (2019). <https://doi.org/10.1002/poc.3904>.
- [7] M.D. Le, V. Warth, L. Giarracca, E. Moine, R. Bounaceur, R. Privat, J.-N. Jaubert, R. Fournet, P.-A. Glaude, B. Sirjean, Development of a detailed kinetic model for the oxidation of *n*-butane in the liquid phase, *J. Phys. Chem. B.* 125 (2021) 6955–6967. <https://doi.org/10.1021/acs.jpcc.1c02988>.
- [8] M.D. Le, Z. El Sayah, R. Benrabah, V. Warth, P.-A. Glaude, R. Privat, R. Fournet, B. Sirjean, An experimental and detailed kinetic modeling of the thermal oxidation stability of *n*-decane as a jet fuel surrogate component, *Fuel.* 342 (2023) 127754. <https://doi.org/10.1016/j.fuel.2023.127754>.
- [9] A. Jalan, R.W. Ashcraft, R.H. West, W.H. Green, Predicting solvation energies for kinetic modeling, *Annu. Rep. Sect. C Phys. Chem.* 106 (2010) 211. <https://doi.org/10.1039/b811056p>.
- [10] Y. Chung, F.H. Vermeire, H. Wu, P.J. Walker, M.H. Abraham, W.H. Green, Group Contribution and Machine Learning Approaches to Predict Abraham Solute Parameters, Solvation Free Energy, and Solvation Enthalpy, *J. Chem. Inf. Model.* 62 (2022) 433–446. <https://doi.org/10.1021/acs.jcim.1c01103>.
- [11] C. Nait Saidi, D.C. Mielczarek, P. Paricaud, Predictions of solvation Gibbs free energies with COSMO-SAC approaches, *Fluid Phase Equilibria.* 517 (2020) 112614. <https://doi.org/10.1016/j.fluid.2020.112614>.
- [12] F.H. Vermeire, W.H. Green, Transfer learning for solvation free energies: From quantum chemistry to experiments, *Chem. Eng. J.* 418 (2021) 129307. <https://doi.org/10.1016/j.cej.2021.129307>.
- [13] Y. Chung, R.J. Gillis, W.H. Green, Temperature- dependent vapor–liquid equilibria and solvation free energy estimation from minimal data, *AIChE J.* 66 (2020). <https://doi.org/10.1002/aic.16976>.
- [14] C.J. Cramer, D.G. Truhlar, A Universal Approach to Solvation Modeling, *Acc. Chem. Res.* 41 (2008) 760–768. <https://doi.org/10.1021/ar800019z>.
- [15] A.V. Marenich, C.J. Cramer, D.G. Truhlar, Generalized Born Solvation Model SM12, *J. Chem. Theory Comput.* 9 (2013) 609–620. <https://doi.org/10.1021/ct300900e>.
- [16] A. Klamt, M. Diedenhofen, Calculation of Solvation Free Energies with DCOSMO-RS, *J. Phys. Chem. A.* 119 (2015) 5439–5445. <https://doi.org/10.1021/jp511158y>.

- [17] F.C. Paes, R. Privat, J.-N. Jaubert, B. Sirjean, A comparative study of COSMO-based and equation-of-state approaches for the prediction of solvation energies based on the compsol databank, *Fluid Phase Equilibria*. 561 (2022) 113540. <https://doi.org/10.1016/j.fluid.2022.113540>.
- [18] J.B. Abrams, L. Rosso, M.E. Tuckerman, Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics, *J. Chem. Phys.* 125 (2006) 074115. <https://doi.org/10.1063/1.2232082>.
- [19] D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley, W. Sherman, Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field, *J. Chem. Theory Comput.* 6 (2010) 1509–1519. <https://doi.org/10.1021/ct900587b>.
- [20] E. Moine, R. Privat, B. Sirjean, J.-N. Jaubert, Estimation of Solvation Quantities from Experimental Thermodynamic Data: Development of the Comprehensive CompSol Databank for Pure and Mixed Solutes, *J. Phys. Chem. Ref. Data*. 46 (2017) 033102. <https://doi.org/10.1063/1.5000910>.
- [21] E. Moine, R. Privat, J.-N. Jaubert, B. Sirjean, N. Novak, E. Voutsas, C. Boukouvalas, Can we safely predict solvation Gibbs energies of pure and mixed solutes with a cubic equation of state?, *Pure Appl. Chem.* 91 (2019) 1295–1307. <https://doi.org/10.1515/pac-2018-1112>.
- [22] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.* 21 (1975) 1086–1099. <https://doi.org/10.1002/aic.690210607>.
- [23] H.V. Kehiaian, Group contribution methods for liquid mixtures: A critical review, *Fluid Phase Equilibria*. 13 (1983) 243–252. [https://doi.org/10.1016/0378-3812\(83\)80098-3](https://doi.org/10.1016/0378-3812(83)80098-3).
- [24] H.S. Wu, S.I. Sandler, Use of ab initio quantum mechanics calculations in group contribution methods. 1. Theory and the basis for group identifications, *Ind. Eng. Chem. Res.* 30 (1991) 881–889. <https://doi.org/10.1021/ie00053a010>.
- [25] F. Jirasek, R.A.S. Alves, J. Damay, R.A. Vandermeulen, R. Bamler, M. Bortz, S. Mandt, M. Kloft, H. Hasse, Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion, *J. Phys. Chem. Lett.* 11 (2020) 981–985. <https://doi.org/10.1021/acs.jpcl.9b03657>.
- [26] E.C. Voutsas, D.P. Tassios, Analysis of the UNIFAC-Type Group-Contribution Models at the Highly Dilute Region. 1. Limitations of the Combinatorial and Residual Expressions, *Ind. Eng. Chem. Res.* 36 (1997) 4965–4972. <https://doi.org/10.1021/ie960770c>.
- [27] D. Constantinescu, A. Klamt, D. Geană, Vapor–liquid equilibrium prediction at high pressures using activity coefficients at infinite dilution from COSMO-type methods, *Fluid Phase Equilibria*. 231 (2005) 231–238. <https://doi.org/10.1016/j.fluid.2005.01.014>.
- [28] A. Klamt, F. Eckert, COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids, *Fluid Phase Equilibria*. 172 (2000) 43–72. [https://doi.org/10.1016/S0378-3812\(00\)00357-5](https://doi.org/10.1016/S0378-3812(00)00357-5).
- [29] A. Klamt, F. Eckert, W. Arlt, COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures, *Annu. Rev. Chem. Biomol. Eng.* 1 (2010) 101–122. <https://doi.org/10.1146/annurev-chembioeng-073009-100903>.
- [30] Y. Shimoyama, Y. Iwai, S. Takada, Y. Arai, T. Tsuji, T. Hiaki, Prediction of phase equilibria for mixtures containing water, hydrocarbons and alcohols at high temperatures and pressures by cubic equation of state with GE type mixing rule based on COSMO-RS, *Fluid Phase Equilibria*. 243 (2006) 183–192. <https://doi.org/10.1016/j.fluid.2006.03.007>.

- [31] K. Leonhard, J. Veverka, K. Lucas, A comparison of mixing rules for the combination of COSMO-RS and the Peng–Robinson equation of state, *Fluid Phase Equilibria*. 275 (2009) 105–115. <https://doi.org/10.1016/j.fluid.2008.09.016>.
- [32] L.F.K. Possani, P.B. Staudt, R. de P. Soares, Prediction of water solubilities in hydrocarbons and oils using F-SAC coupled with SRK–EoS, *Fluid Phase Equilibria*. 427 (2016) 394–405. <https://doi.org/10.1016/j.fluid.2016.08.001>.
- [33] Y. Le Guennec, R. Privat, J.-N. Jaubert, Development of the translated-consistent *tc*-PR and *tc*-RK cubic equations of state for a safe and accurate prediction of volumetric, energetic and saturation properties of pure compounds in the sub- and super-critical domains, *Fluid Phase Equilibria*. 429 (2016) 301–312. <https://doi.org/10.1016/j.fluid.2016.09.003>.
- [34] A. Pina-Martinez, R. Privat, J. Jaubert, Use of 300,000 pseudo- experimental data over 1800 pure fluids to assess the performance of four cubic equations of state: SRK, PR, *tc*- RK, and *tc*- PR, *AIChE J.* 68 (2022) e17518. <https://doi.org/10.1002/aic.17518>.
- [35] E. Mullins, R. Oldland, Y.A. Liu, S. Wang, S.I. Sandler, C.-C. Chen, M. Zwolak, K.C. Seavey, Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods, *Ind. Eng. Chem. Res.* 45 (2006) 4389–4415. <https://doi.org/10.1021/ie060370h>.
- [36] T. Mu, J. Rarey, J. Gmehling, Group contribution prediction of surface charge density profiles for COSMO-RS(OI), *AIChE J.* 53 (2007) 3231–3240. <https://doi.org/10.1002/aic.11338>.
- [37] T. Mu, J. Rarey, J. Gmehling, Group contribution prediction of surface charge density distribution of molecules for COSMO-SAC, *AIChE J.* 55 (2009) 3298–3300. <https://doi.org/10.1002/aic.11933>.
- [38] J.-N. Jaubert, R. Privat, Y. Le Guennec, L. Coniglio, Note on the properties altered by application of a Pénélox–type volume translation to an equation of state, *Fluid Phase Equilibria*. 419 (2016) 88–95. <https://doi.org/10.1016/j.fluid.2016.03.012>.
- [39] R. Privat, J.-N. Jaubert, Y. Le Guennec, Incorporation of a volume translation in an equation of state for fluid mixtures: which combining rule? Which effect on properties of mixing?, *Fluid Phase Equilibria*. 427 (2016) 414–420. <https://doi.org/10.1016/j.fluid.2016.07.035>.
- [40] Y. Le Guennec, S. Lasala, R. Privat, J.-N. Jaubert, A consistency test for α -functions of cubic equations of state, *Fluid Phase Equilibria*. 427 (2016) 513–538. <https://doi.org/10.1016/j.fluid.2016.07.026>.
- [41] Y. Le Guennec, R. Privat, S. Lasala, J.-N. Jaubert, On the imperative need to use a consistent α -function for the prediction of pure-compound supercritical properties with a cubic equation of state, *Fluid Phase Equilibria*. 445 (2017) 45–53. <https://doi.org/10.1016/j.fluid.2017.04.015>.
- [42] C.H. Twu, D. Bluck, J.R. Cunningham, J.E. Coon, A cubic equation of state with a new alpha function and a new mixing rule, *Fluid Phase Equilibria*. 69 (1991) 33–50. [https://doi.org/10.1016/0378-3812\(91\)90024-2](https://doi.org/10.1016/0378-3812(91)90024-2).
- [43] G. Soave, Equilibrium constants from a modified Redlich-Kwong equation of state, *Chem. Eng. Sci.* 27 (1972) 1197–1203. [https://doi.org/10.1016/0009-2509\(72\)80096-4](https://doi.org/10.1016/0009-2509(72)80096-4).
- [44] A. Pina-Martinez, R. Privat, J.-N. Jaubert, D.-Y. Peng, Updated versions of the generalized Soave α -function suitable for the Redlich-Kwong and Peng-Robinson equations of state, *Fluid Phase Equilibria*. 485 (2019) 264–269. <https://doi.org/10.1016/j.fluid.2018.12.007>.
- [45] K. Magoulas, D. Tassios, Thermophysical properties of n-Alkanes from C1 to C20 and their prediction for higher ones, *Fluid Phase Equilibria*. 56 (1990) 119–140. [https://doi.org/10.1016/0378-3812\(90\)85098-U](https://doi.org/10.1016/0378-3812(90)85098-U).

- [46] A. Piña-Martinez, R. Privat, I.K. Nikolaidis, I.G. Economou, J.-N. Jaubert, What Is the Optimal Activity Coefficient Model To Be Combined with the *translated – consistent* Peng–Robinson Equation of State through Advanced Mixing Rules? Cross-Comparison and Grading of the Wilson, UNIQUAC, and NRTL a^E Models against a Benchmark Database Involving 200 Binary Systems, *Ind. Eng. Chem. Res.* 60 (2021) 17228–17247. <https://doi.org/10.1021/acs.iecr.1c03003>.
- [47] M.-J. Huron, J. Vidal, New mixing rules in simple equations of state for representing vapour-liquid equilibria of strongly non-ideal mixtures, *Fluid Phase Equilibria.* 3 (1979) 255–271. [https://doi.org/10.1016/0378-3812\(79\)80001-1](https://doi.org/10.1016/0378-3812(79)80001-1).
- [48] M.L. Michelsen, A modified Huron-Vidal mixing rule for cubic equations of state, *Fluid Phase Equilibria.* 60 (1990) 213–219. [https://doi.org/10.1016/0378-3812\(90\)85053-D](https://doi.org/10.1016/0378-3812(90)85053-D).
- [49] M.L. Michelsen, A method for incorporating excess Gibbs energy models in equations of state, *Fluid Phase Equilibria.* 60 (1990) 47–58. [https://doi.org/10.1016/0378-3812\(90\)85042-9](https://doi.org/10.1016/0378-3812(90)85042-9).
- [50] S. Dahl, M.L. Michelsen, High-pressure vapor-liquid equilibrium with a UNIFAC-based equation of state, *AIChE J.* 36 (1990) 1829–1836. <https://doi.org/10.1002/aic.690361207>.
- [51] G.M. Wilson, Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing, *J. Am. Chem. Soc.* 86 (1964) 127–130. <https://doi.org/10.1021/ja01056a002>.
- [52] H. Renon, J.M. Prausnitz, Local compositions in thermodynamic excess functions for liquid mixtures, *AIChE J.* 14 (1968) 135–144. <https://doi.org/10.1002/aic.690140124>.
- [53] D.S. Abrams, J.M. Prausnitz, Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems, *AIChE J.* 21 (1975) 116–128. <https://doi.org/10.1002/aic.690210115>.
- [54] J.-N. Jaubert, Y. Le Guennec, A. Pina-Martinez, N. Ramirez-Velez, S. Lasala, B. Schmid, I.K. Nikolaidis, I.G. Economou, R. Privat, Benchmark database containing binary-system-high-quality-certified data for cross-comparing thermodynamic models and assessing their accuracy, *Ind. Eng. Chem. Res.* 59 (2020) 14981–15027. <https://doi.org/10.1021/acs.iecr.0c01734>.
- [55] J.-W. Qian, R. Privat, J.-N. Jaubert, P. Duchet-Suchaux, Enthalpy and heat capacity changes on mixing: fundamental aspects and prediction by means of the PPR78 cubic equation of state, *Energy Fuels.* 27 (2013) 7150–7178. <https://doi.org/10.1021/ef401605c>.
- [56] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J Chem Soc Perkin Trans 2.* (1993) 799–805. <https://doi.org/10.1039/P29930000799>.
- [57] A. Klamt, COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design, 1st ed, Elsevier, Amsterdam Boston, 2005.
- [58] I.H. Bell, E. Mickoleit, C.-M. Hsieh, S.-T. Lin, J. Vrabec, C. Breitkopf, A. Jäger, A Benchmark Open-Source Implementation of COSMO-SAC, *J. Chem. Theory Comput.* 16 (2020) 2635–2646. <https://doi.org/10.1021/acs.jctc.9b01016>.
- [59] A. Klamt, M. Diederhofen, A refined cavity construction algorithm for the conductor-like screening model, *J. Comput. Chem.* 39 (2018) 1648–1655. <https://doi.org/10.1002/jcc.25342>.
- [60] T. Gerlach, S. Müller, A.G. de Castilla, I. Smirnova, An open source COSMO-RS implementation and parameterization supporting the efficient implementation of multiple segment descriptors, *Fluid Phase Equilibria.* 560 (2022) 113472. <https://doi.org/10.1016/j.fluid.2022.113472>.

- [61] I.K. Nikolaidis, R. Privat, J.-N. Jaubert, I.G. Economou, Assessment of the perturbed chain-statistical associating fluid theory equation of state against a benchmark database of high-quality binary-system data, *Ind. Eng. Chem. Res.* 60 (2021) 8935–8946. <https://doi.org/10.1021/acs.iecr.1c01234>.
- [62] I.K. Nikolaidis, R. Privat, J.-N. Jaubert, I.G. Economou, The effect of binary interaction parameters and induced association on the performance of the PC-SAFT equation of state: validation against a benchmark database of high-quality binary-system data, *J. Chem. Eng. Data.* (2023). <https://doi.org/10.1021/acs.jced.2c00781>.
- [63] W.-L. Chen, C.-C. Hsu, S.-T. Lin, Prediction of phase behaviors of acetic acid containing fluids, *Fluid Phase Equilibria.* 353 (2013) 61–68. <https://doi.org/10.1016/j.fluid.2013.05.039>.
- [64] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, *Fluid Phase Equilibria.* 183–184 (2001) 183–208. [https://doi.org/10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9).
- [65] N.N. Kalikin, Y.A. Budkov, A.L. Kolesnikov, D.V. Ivlev, M.A. Krestyaninov, M.G. Kiselev, Computation of drug solvation free energy in supercritical CO₂: Alternatives to all-atom computer simulations, *Fluid Phase Equilibria.* 544–545 (2021) 113096. <https://doi.org/10.1016/j.fluid.2021.113096>.
- [66] C.-M. Hsieh, S.-T. Lin, J. Vrabec, Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior, *Fluid Phase Equilibria.* 367 (2014) 109–116. <https://doi.org/10.1016/j.fluid.2014.01.032>.
- [67] L.F. Baladão, P. B. Staudt, R. de P. Soares, Functional-Segment Activity Coefficient Equation of State: F-SAC-Phi, *Ind. Eng. Chem. Res.* 58 (2019) 16934–16944. <https://doi.org/10.1021/acs.iecr.9b02190>.
- [68] C.-M. Hsieh, S.I. Sandler, S.-T. Lin, Improvements of COSMO-SAC for vapor–liquid and liquid–liquid equilibrium predictions, *Fluid Phase Equilibria.* 297 (2010) 90–97. <https://doi.org/10.1016/j.fluid.2010.06.011>.
- [69] C.-K. Chang, W.-L. Chen, D.T. Wu, S.-T. Lin, Improved Directional Hydrogen Bonding Interactions for the Prediction of Activity Coefficients with COSMO-SAC, *Ind. Eng. Chem. Res.* 57 (2018) 11229–11238. <https://doi.org/10.1021/acs.iecr.8b02493>.