A comparative study of COSMO-based and equation-of-state approaches for the prediction of solvation energies based on the compsol databank

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

▶ **To cite this version:**

Francisco Carlos Paes, Romain Privat, Jean-Noël Jaubert, Baptiste 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, 2022, 561, pp.113540. 10.1016/j.fluid.2022.113540 . hal-03763390

HAL Id: hal-03763390
https://hal.science/hal-03763390

Submitted on 29 Aug 2022

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

A COMPARATIVE STUDY OF COSMO-BASED AND EQUATION-OF-STATE APPROACHES FOR THE PREDICTION OF SOLVATION ENERGIES BASED ON THE COMPSOL DATABANK

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

Université de Lorraine, École Nationale Supérieure des Industries Chimiques, Laboratoire Réactions et Génie des Procédés (UMR CNRS 7274), 1 rue Grandville, 54000 Nancy, France.

* Corresponding author. Tel.: +33 03 72 74 37 73

E-mail address: romain.privat@univ-lorraine.fr

ABSTRACT

The computation of solvation energies has many uses in several fields, such as design of separation processes, pharmacology and drug-design, and kinetic modeling. These applications require thermodynamic models capable of accurately predicting solvation energies, accounting for the temperature dependency of this property, and which are fast and robust. Within this framework, we compared two COSMO-based continuum solvation models (COSMO-RS and COSMO-SAC-dsp) with two versions of predictive cubic equations of state well-acknowledged for their efficiency (PSRK and UMR-PRU). For this purpose, a large experimental set of 65,000 datapoints extracted from the COMPSOL databank was considered. Comparisons between computed and experimental data were performed for Gibbs solvation energy and, for the first time in the literature, for both entropy and enthalpy of solvation simultaneously. For simpler binary mixtures, in which hydrogen bonding does not take place, all models were capable of providing accurate predictions, with average absolute deviations below 0.3 kcal/mol regarding the solvation Gibbs energy. For more complex associating mixtures, COSMO-RS showed the best correlation between experimental and calculated data, especially for aqueous systems; among EoS, it is observed that the PSRK model offers the best accuracy.

1 INTRODUCTION

The estimation of solvation Gibbs energy (denoted $\Delta_{\mathrm{solv}}\bar{g}_i$ for a given component i) is an important issue in process and product design. Given that this thermodynamic quantity is related to the degree of affinity between different chemical species, it can be very useful, for example, in the selection of an adequate solvent for a separation process or a chemical reaction. The computation of solvation quantities is also a critical matter in pharmacology and drug-design, since the solvation phenomena directly affects the affinity of a given biologically active substance for target proteins, as well as its solubility and chemical stability [1–4].

Solvation energies are also required for liquid-phase detailed chemical kinetic models, that involves a large number of reactions and chemical species, and are based on gas-phase models developed to simulate pyrolysis, combustion or atmospheric oxidation phenomena [5–7]. These examples highlight the need for predictive and accurate methods that can predict solvation energies and at the same time are fast and robust and can be applied to a wide range of chemical compounds. As a noticeable feature, these applications cover large ranges of temperature and therefore, models used for solvation energies must include a temperature dependency as a prerequisite.

With regard to the accuracy, a high degree of detail can be achieved by using the so-called discrete solvation models. In such models, both solute and solvent are treated explicitly, i.e., the solute is considered to be surrounded by a large number of solvent molecules. Solute and solvent molecules can be described by quantum mechanics (QM), and the sampling of their possible arrangements is usually carried out by Monte-Carlo (MC) or Molecular Dynamics (MD) algorithms [8]. An alternative to simplify the calculation is to use classical molecular mechanics (MM) for the representation of the surrounding solvent molecules. The accuracy of such simulations depends on the quality of the force-field potentials and the accuracy of the simulation algorithm used. This type of discrete method has been widely used for the computation of solvation quantities in pharmacology and drug-design [9–11]. Although accurate, such approaches require high computational efforts, which limits their application as a high-throughput method.

Computational costs can be further reduced by using continuum solvation models, in which a solute molecule is described at a homogeneous QM level and then placed into a void cavity within a continuous dielectric medium that represents the ensemble of solvent molecules [12]. This category includes the solvation models (SMx) derived from generalized Born theory

(GB), as well as the continuum solvation models based upon apparent surface charge methods (ACS), such as polarizable continuum models (PCM) and conductor-like screening models (COSMO) [8]. Due to the good trade-off between accuracy and computational efforts provided by continuum solvation models they have been widely used for the computation of solvation quantities [13–16]. This approach can be adapted to accurate / high-throughput calculations if the results of the electronic structure calculation (QM) are tabulated or can be approximated without loss of accuracy.

Strictly speaking, the chemical potential of solvation of a species i, often simply called "solvation energy", is defined as the change in the chemical potential of i when a molecule, devoid of kinetic energy, is transferred from a perfect-gas phase into a condensed phase [17]. Denoting $\bar{g}_i^{\emptyset}$, the pseudo-chemical potential of component i, i.e., the chemical potential of species i in a given mixture, when i is devoid of kinetic energy, the solvation chemical potential is defined as:

$$\Delta_{\text{solv}}\bar{\text{g}}_{\text{i}}(\text{T},\text{P},\mathbf{n}) = \bar{\text{g}}_{\text{i,liq}}^{\varnothing}(\text{T},\text{P},\mathbf{n}) - \bar{\text{g}}_{\text{i}}^{\varnothing,\bullet}(\text{T},\text{P}) \tag{1}$$

Where symbol • refers to perfect-gas properties and $\boldsymbol{n}$ denotes the mole number vector of the mixture considered. This is the strict definition of solvation chemical potential following Ben-Naim seminal proposal [17] that will be used all along this study.

Assuming infinite dilution of a solute in a solvent, one can demonstrate that this quantity can be calculated by a combined estimation of pure solvent density (denoted $\rho_{\text{j,liq}}^{\text{sat}}$ for liquid solvent *j*) and the fugacity coefficient of the infinitely-diluted solute (denoted $\varphi_{\text{i,liq}}^{\infty}$ for solute i) at a given condition of temperature (T) and pressure (P), as highlighted in Eq. (2) [17,18]. This equation is well adapted for the implementation of an approach based on the use of an equation of state (EoS) which is a promising approach as highlighted previously by a common study between our research group and authors of the UMR-PRU EoS [19].

$$\Delta_{\text{solv}}\bar{\text{g}}_{\text{i}}^{\infty}(\text{T},\text{P}) = \text{RT}\ln\left[\frac{\text{P}\varphi_{\text{i,liq}}^{\infty}(\text{T},\text{P})}{\text{RT}\rho_{\text{j,liq}}^{\text{sat}}(\text{T},\text{P})}\right] \tag{2}$$

Regardless of the approach adopted, it is possible to predict other solvation quantities from the calculation of $\Delta_{\text{solv}}\bar{g}_i$ for instance, solvation entropy ($\Delta_{\text{solv}}\bar{s}_i$) and enthalpy ($\Delta_{\text{solv}}\bar{h}_i$) can be calculated by Eqs. (3) and (4), respectively. The calculation of such properties are especially interesting for systems in which entropy ($-T\cdot\Delta_{\text{solv}}\bar{s}_i$) and enthalpy ($\Delta_{\text{solv}}\bar{h}_i$) contributions compensate each other, so that the solvation Gibbs energy remains small [20].

$$\Delta_{solv}\bar{s}_i = -\left(\frac{\partial \Delta_{solv}\bar{g}_i}{\partial T}\right)_{P,\mathbf{x}} \quad (3)$$

$$\Delta_{solv}\bar{h}_i = \Delta_{solv}\bar{g}_i + T\Delta_{solv}\bar{s}_i \quad (4)$$

With this in mind, we undertook this study to compare the prediction capability of COSMO-based approaches (COSMO-RS and COSMO-SAC-dsp) with EoS-based approaches, using two predictive versions of cubic EoS: the PSRK and UMR-PRU models. These EoS were selected mainly for two reasons: (i) they are among the most efficient predictive EoS currently available in the open literature, and (ii) they enable to model simple to complex mixtures including the ones were chemical association (e.g., hydrogen bonding) takes place.

To compare these models on a fair basis, the extensive COMPSOL databank [18] was used as reference data for the computation of the solvation quantities of interest (Gibbs energy, entropy and enthalpy). In order to provide an instructive comparison of the different models, a systematic classification of the binary mixtures was carried out based on the associative character of their compounds. It is important to make it clear that this benchmarking study focused only on neutral molecules.

2 METHODS

2.1 Thermodynamic models

2.1.1 Conductor-like screening models (COSMO)

In short, the prediction of thermodynamic properties with COSMO-type models is based on molecular quantum calculation in gas phase, mostly DFT methods, followed by a conductor-like screening calculation (COSMO) [21]. This calculation generates the so-called σ-profile, which basically stands for a distribution of the local screening charges [22]. The screening charge density (σ) is then one of the descriptors used in the further calculation of thermodynamic quantities, along with size and shape descriptors for the calculation of combinatorial contributions, and element specific coefficients used to estimate the dispersive interactions [23]. In COSMO-RS (realistic solvation), such calculation is based on the estimation of the chemical potential of an average molecular contact area of the ensemble that represents the condensed phase. This information is afterwards used to calculate the chemical potential of a given solute in this ensemble [24].

On the other hand, in COSMO-SAC (segment activity coefficient), the activity coefficient of the surface segments is usually calculated instead. Indeed, COSMO-SAC is still mathematically equivalent to COSMO-RS, with few differences. For instance, the formulation

of the combinatorial term and the hydrogen-bonding interaction energy contribution are different from one model to another [25,26]. In this study, we used the COSMO-SAC-dsp version, which introduces a supplementary dispersion term calculated from the one-parameter Margules equation [27].

2.1.2 Equation of state (EoS)

Regarding the EoS approach, the predictive Soave–Redlich–Kwong (PSRK) and the Universal Mixing Rule-Peng Robinson UNIFAC (UMR-PRU) models were evaluated. Both models combine a cubic EoS with an activity-coefficient (or, excess Gibbs energy, g^E) model.

In the case of the PSRK model, the EoS adopted is the Soave–Redlich–Kwong (SRK), and the mixing rule is MHV1 combined with a modified UNIFAC g^E model proposed by J. Gmehling [28–32]. The UMR-PRU combines the volume-translated Peng-Robinson (VTPR) EoS with a MHV-based mixing rule that eliminates the double combinatorial term that appears in the original formulation of MHV mixing rules [33]. In this case, the g^E model used in the mixing rule is a modified version of the UNIFAC model previously proposed by Hansen and co-workers [34].

Due to the use of a predictive g^E model, for both EoS, three to four pure-compound parameters are required as input parameters for each component in the mixture: their critical temperature, critical pressure, and acentric factor. An additional volumetric translation parameter is present in the UMR-PRU model.

2.2 Reference dataset (COMPSOL databank)

The benchmark of the different models was based on experimental data extracted from the COMPSOL databank [18] (where COMPSOL stands for *Comprehensive solvation*). The advantage of using this database is the large variety of chemical species available, as well as the wide range of pressure and temperature conditions. It is certainly the widest database of solvation properties currently available. Indeed, the original COMPSOL dataset comprises 14,202 binary mixtures involving 865 different solvents and 775 different solutes. Altogether, 70,062 experimental values for solvation energy and 29,344 experimental values for solvation entropy and enthalpy data, at infinite dilution, are reported in this database [18].

It was not possible to make use of the entire COMPSOL data. In fact, due to some limitations related to the models considered in this study, certain filters had to be applied in order to ensure that the various approaches were being used within their respective scopes of

application, or even that their application was actually possible. Thus, we considered only experimental data for which the following conditions were satisfied:

- The solvent is liquid. Therefore, the experimental pressure and temperature must be lower than the critical coordinates of the solvent ($P_{exp} < P_{c,solvent}$ and $T_{exp} < T_{c,solvent}$).
- The solvent remains incompressible, which is a starting assumption for the use of COSMO-based models. This assumption is considered valid provided $P_{exp} < 30$ bar.
- The solute and solvent must have COSMO files (i.e., the files containing input information for use of COSMO based models) present in both COSMOTherm and Simulis databases (see section 2.4), in order to enable the calculation with COSMO-RS and COSMO-SAC-dsp, respectively, without requiring an additional *ab initio* calculation step.
- For the COSMO-SAC-dsp model, an additional assumption must be considered: only data at temperature T such that $P_i^{sat}(T) < 5$ bar (see section 2.4).
- The solute and solvent can be decomposed into groups for the use of the group contribution methods of each predictive cubic EoS considered (PSRK and UMR-PRU). In this case, we considered only binary mixtures in which all interaction parameters between groups were available.

In the case of the EoS-approach, a complementary verification was added to ensure that the calculations were actually performed in liquid phase. Indeed, although the solvent is liquid in experimental conditions of T and P, the EoS may not predict the correct phase. Consequently, as a preliminary step, the vapor pressure ($P_{solvent}^{sat}$) was calculated and then compared with the experimental pressure (P_{exp}). The higher value was then retained for the subsequent calculations of the solvation quantities: $P = \max\{P_{exp}; P_{solvent}^{sat}\}$.

Due to these filters, a total of 81% of the data published in the COMPSOL database was actually taken into account in the present work.

2.3 Binary association codes (BAC)

As stated in the introduction section, one of the aims of this work is to perform a systematic analysis of the results obtained for the different binary systems available in the extensive COMPSOL database. To discuss the results, binary mixtures will be sorted according to their binary association code, a descriptor revealing the presence and strength of molecular association, through hydrogen bonding mainly. Note that the degree of association in a mixture has a significant impact on the solvation properties. For a detailed discussion on this topic see [35].

In general terms, the binary association code typifies the hydrogen-bonding capability of both compounds in the binary mixture. At first, this capability is evaluated for each pure compound based on their σ-profiles obtained from DFT and COSMO calculation, as described in [36]. Basically, a pure component, depending on its ability to donate and/or accept hydrogen, can be classified as:

- NA = Nonassociating, nonpolar (alkanes)
- HA = Hydrogen-Acceptor, polar but nonassociating (ketones, aldehydes, and ethers)
- HD = Hydrogen-Donor, polar but nonassociating (di- or trihalogenated compounds)
- SA = Self-Associating, polar and associating (water, alcohols, and carboxylic acids)

As illustrated in Table 1, by performing all of the possible combinations between the four types (NA, HA, HD and SA) of pure components, it was possible to define 9 categories of binary systems, identified each by the so-called binary association code (BAC).

Table 1 – Binary association code (BAC).

		Solute			
		NA	**HA**	**HD**	**SA**
Solvent	**NA**	NA-NA (1)	-	-	-
	HA	HA-NA (2)	HA-HA (4)	-	-
	HD	HD-NA (3)	HD-HA (6)	HD-HD (4)	-
	SA	SA-NA (5)	SA-HA (8)	SA-HD (7)	SA-SA (9)

By grouping together some of these 9 categories, it was possible to further classify binary systems into four main types based on the level of association they exhibit:

- Type 1 (BAC =1-4): mixtures without association.
- Type 2 (BAC = 5): mixtures in which only self-association takes place, but tends to be broken.
- Type 3 (BAC = 6): mixtures in which only cross-association takes place. The two components do not exhibit association when they are pure, but their mixture exhibits hydrogen bonding.
- Type 4 (BAC = 7-9): mixtures in which both cross-association and self-association take place.

2.4 Technical details

The predictions of solvation Gibbs energy using COSMO-RS were performed in the commercially available software COSMOTherm. The quantum chemical level considered was BP/TZVPD-FINE and the COSMO-RS parametrization was BP_TZVPD_FINE_C30_1401. All the COSMO files were already available in the COSMOTherm database. In COSMO-RS, the solvation Gibbs energy is directly calculated from the solute chemical potential in liquid and gas phase, as showed in Eq.(5). In this equation, the chemical potential of the solute molecule infinitely diluted in the solvent ($\mu_{i,liq}^{\infty}$) is obtained from statistical thermodynamic calculation that considers the σ-profiles and the σ-potential, whereas the solute's chemical potential in the gas phase ($\mu_{i,gas}$) is estimated from the DFT and COSMO calculation outputs, as denoted in Eq.(6).

$$\Delta_{solv}\bar{g}_i^{\infty}(T) = \mu_{i,liq}^{\infty}(T) - \mu_{i,gas}(T) \tag{5}$$

$$\mu_{i,gas}(T) = E_{i,gas} - E_{i,cosmo} - \omega_{ring} n_{i,ring} + \eta_{gas}(T) \tag{6}$$

In the equation above, $E_{i,gas}$ and $E_{i,cosmo}$ are the quantum chemical total energies of the solute in the gas phase and in the COSMO conductor, respectively. The adjustable parameter ω_{ring} is a correction for ring compounds, and $n_{i,ring}$ is the number of ring atoms in the molecule. Finally, η_{gas} provides the link between the reference states of the system in the gas and in the liquid phase.

Meanwhile, the calculation with COSMO-SAC-dsp were carried out using Simulis Thermodynamics v.2.0.38, with the original parameterization of COSMO-SAC-dsp published in [27]. In this case, the σ-profile of the different molecules was generated by GGA-WVNBP/DNP ab initio method. Again, only the COSMO files already available in the software database were taken in account. When COSMO-SAC-dsp is used, Simulis Thermodynamics returns the activity coefficient of the solute ($\gamma_{i,liq}^{\infty}$). This result can be related to $\Delta_{solv}\bar{g}_i$ by means of Eq. (7) below (derivation leading to this equation can be found in ref. [18]), which actually is an equivalent representation of Eq. (2). Both solute saturation pressure (P_i^{sat}) and solvent liquid density ($\rho_{j,liq}^{sat}$) were obtained from DIPPR database.

$$\Delta_{solv}\bar{g}_i^{\infty}(T,P) = RT\ln\left[\frac{P_i^{sat}(T)\cdot\gamma_{i,liq}^{\infty}(T)\cdot\varphi_i^{sat}(T)}{RT\rho_{j,liq}^{sat}(T)\cdot\exp\left[\frac{P_i^{sat}(T)-P}{RT\rho_{j,liq}^{sat}(T)}\right]}\right] \tag{7}$$

In the present case, assuming that the vapor phase of pure i in vapor-liquid equilibrium at T behaves as a perfect gas, the fugacity coefficient $\varphi_i^{sat}(T)$ was set to 1. As mentioned before, to make this assumption acceptable, only data at temperature T such that $P_i^{sat}(T) < 5$ bar were considered.

We also used the same version of Simulis Thermodynamics for the calculation of solvation energies from the cubic EoS (PSRK and UMR-PRU), which is done by Eq. (2). In this case, saturation pressure (P_i^{sat}) and solvent liquid density ($\rho_{j,liq}^{sat}$) were directly calculated by the EoS (not obtained from DIPPR correlations).

For all models, the solvation entropy was calculated by Eq. (3), via the finite difference method (central scheme), and the solvation enthalpy was then calculated by Eq.(4).

3 RESULTS AND DISCUSSIONS

The comparison of the different models is done on the basis of the average absolute deviations, which can be consulted in detail in the appendix 1, along with the standard deviation results, for the three solvation quantities considered in this study (solvation Gibbs energy, solvation entropy, and solvation enthalpy).

3.1 Solvation Gibbs Energy

Concerning the prediction of the solvation Gibbs energy, Figure 1 shows the performance of all models for aqueous and non-aqueous systems. It can be seen that, in general terms, all models are able to provide a satisfactory prediction of $\Delta_{solv}\bar{g}_i^{\infty}$ in non-aqueous media, since the absolute average deviations did not exceed 0.4 kcal/mol. As could be expected, Figure 1 pinpoints a significant deviation increase in the case where water is the solvent. Indeed, water has a complex association scheme with four possible hydrogen bonding sites, which normally leads to a strongly non-ideal behavior. In that respect, COSMO-RS clearly shows better performances than the other models evaluated in this study.

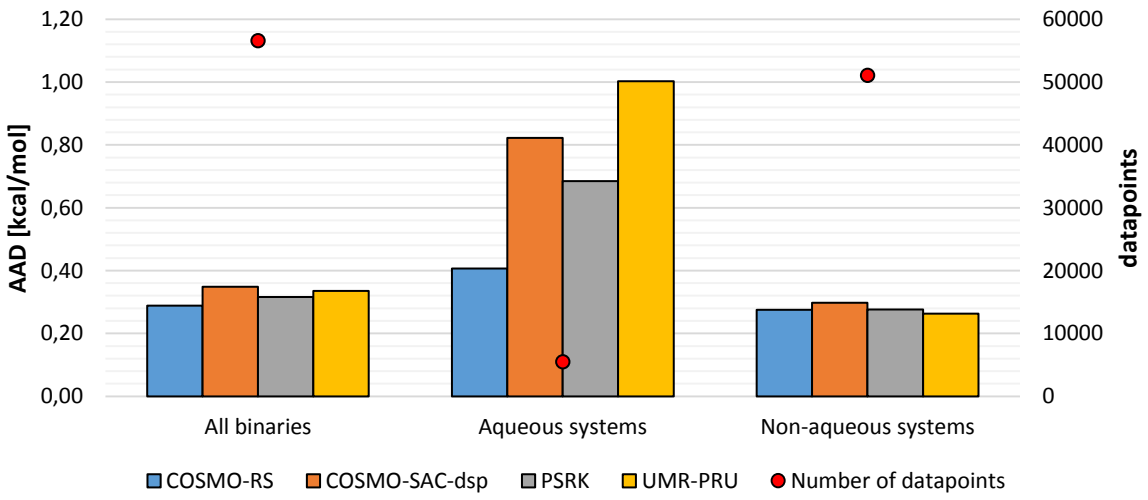

Figure 1 –Average absolute deviations between experimental solvation Gibbs energy (or equivalently, chemical potential of solvation $\Delta_{solv}\bar{g}_i$) data and their prediction from 4 models for aqueous and non-aqueous systems.

The hydration results put forward the role of hydrogen bonding in the quality of the prediction. For that matter, Figure 2 and Figure 3 illustrate the variation of the average absolute deviation according to the association character of the binary mixtures. As expected, a good agreement between experimental and calculated data was found for all models regarding the non-associating binary mixtures (BAC=1-4), with a slight advantage to COSMO-SAC and UMR-PRU models. In fact, this sort of system is generally well correlated by most of sophisticated-enough thermodynamic models. Small deviations were also observed for binary mixtures in which only cross-association takes place (BAC=6). In such systems, solute and solvent molecules form up hydrogen bonds with each other and this stabilizes the liquid phase. Note that the good performances of cubic models for the description of cross-associating systems were highlighted already by past studies (e.g., see Table 15 of [35]). From Figure 2 and Figure 3 we can observe that all models were able to describe such a behavior satisfactorily.

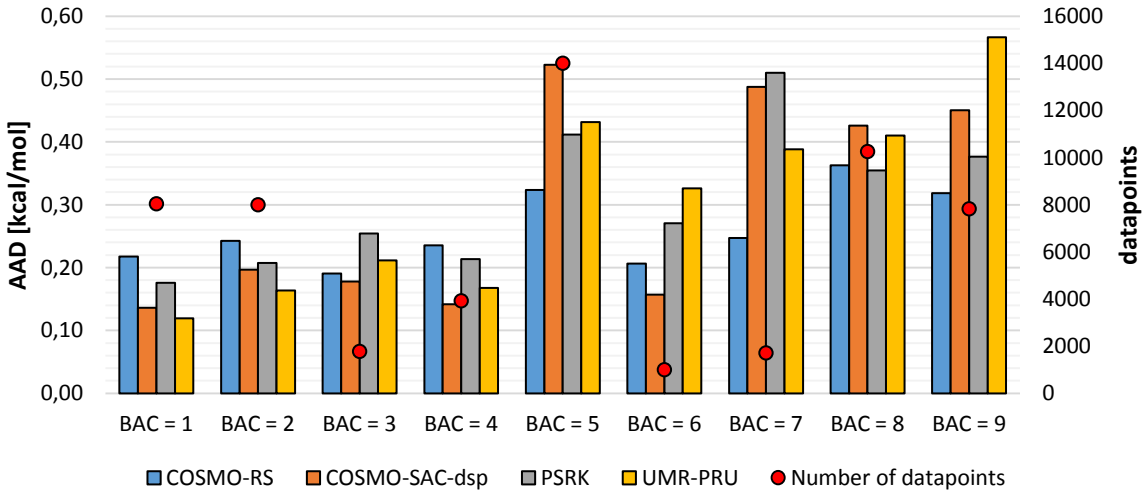

Figure 2 – Average absolute deviations between experimental solvation Gibbs energy ($\Delta_{solv}\bar{g}_i$) data and their prediction from 4 different models calculated for each binary association code (BAC).

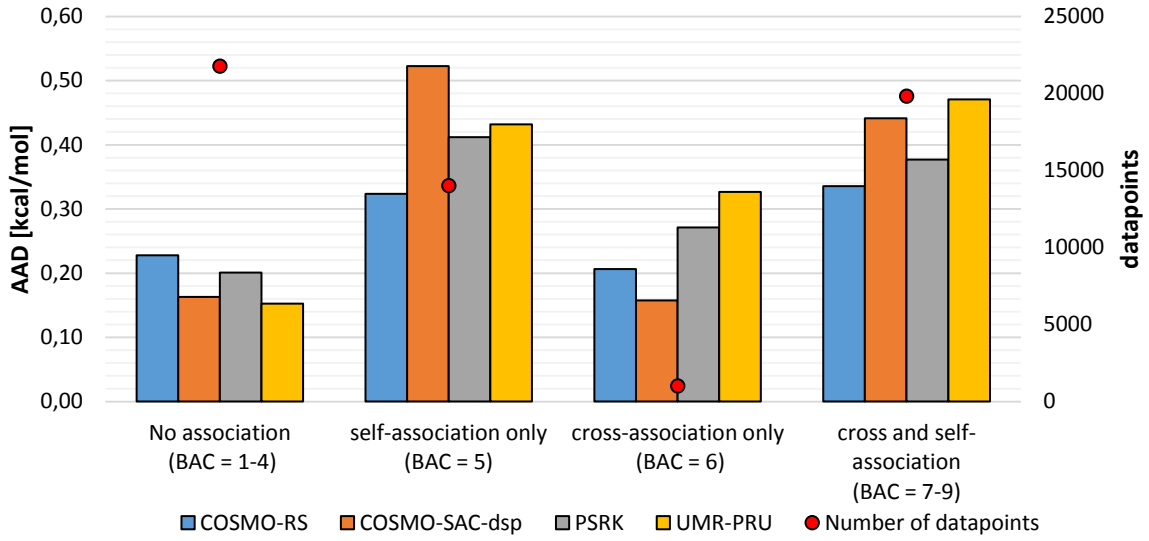

Figure 3 – Average absolute deviations between experimental solvation Gibbs energy ($\Delta_{solv}\bar{g}_i$) data and their prediction from 4 different models calculated for each type of association.

On the other hand, when cross and self-association happen simultaneously (BAC=7, 8, 9), a more complex phenomenon takes place. Due to dilution effects, the hydrogen bonds between self-associating molecules tend to be broken while new hydrogen bonds are formed by cross-association. For these cases, an increase in the deviation between experimental and calculated data was observed. Attention must be paid to mixtures containing self-associating and hydrogen-donor molecules (BAC 7), since a considerable loss of accuracy was reported for COSMO-SAC and PSRK. This category contains, for example, binary mixtures made up of

alcohols or water with halogen compounds. A significant increase in deviation was also observed for binary mixtures of self-associating molecules (BAC 9), such as alcohols, using UMR-PRU EoS. COSMO-RS presented the best level of accuracy for this sort of systems.

Moreover, hydrogen bonds between self-associating molecules are also broken when they are mixed with non-associating molecules (BAC=5), but in this case there is no cross-association to stabilize the liquid-phase and, thus, phase splitting is often observed. This kind of mixture is naturally more difficult to be represented by a thermodynamic model, which explains the increase in the deviations results observed in all models. The COSMO-RS approach provided the most accurate results for this case as well.

Broadly speaking, COSMO-RS was found to be the most accurate model, especially mixtures showing high degree of association (i.e., with a binary association code greater than 4). Both COSMO-RS and COSMO-SAC-dsp models have a dedicated hydrogen bonding term for calculating the interacting energy between the surface segments. It seems that the temperature-dependent formulation implemented in COSMO-RS tends to be more accurate than COSMO-SAC-dsp for the calculation of the solvation quantities of associating mixtures at infinite dilution. On the other hand, PSRK and UMR-PRU EoS do not have a specific term to predict the association phenomena; it must however be noticed that these complex interactions are addressed partially through the use of advanced EoS/g^E mixing rules.

Surprisingly, PSRK provided better results than UMR-PRU for most associating binary mixtures. One could say that it is in contrast to expectations since the addition of a volume-translation parameter in the latter model is supposed to lead to better predictions of liquid-phase density [37]. Moreover, unlike PSRK, the mixing rules of UMR-PRU do not suffer from the double combinatorial term issue [38,39]. Therefore, better performances of UMR-PRU over PSRK would be expected, especially for asymmetric systems. However, when there is no hydrogen bond, UMR-PRU shows better performance.

3.2 Solvation entropy and enthalpy

Let us now turn to the prediction of other solvation quantities. Figures 4-6 show the deviation results between experimental and calculated data for solvation entropy, whereas Figures 7-9 show the deviation results for solvation enthalpy.

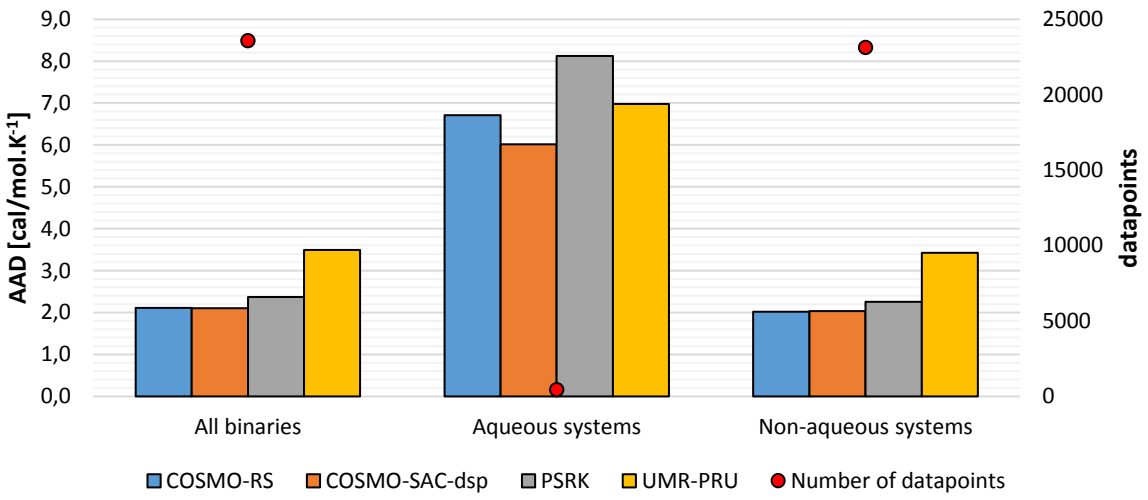

Figure 4 – Average absolute deviations between experimental solvation entropy (or equivalently, partial molar entropy of solvation $\Delta_{solv}\bar{s}_i$) data and their prediction from 4 models for aqueous and non-aqueous systems.

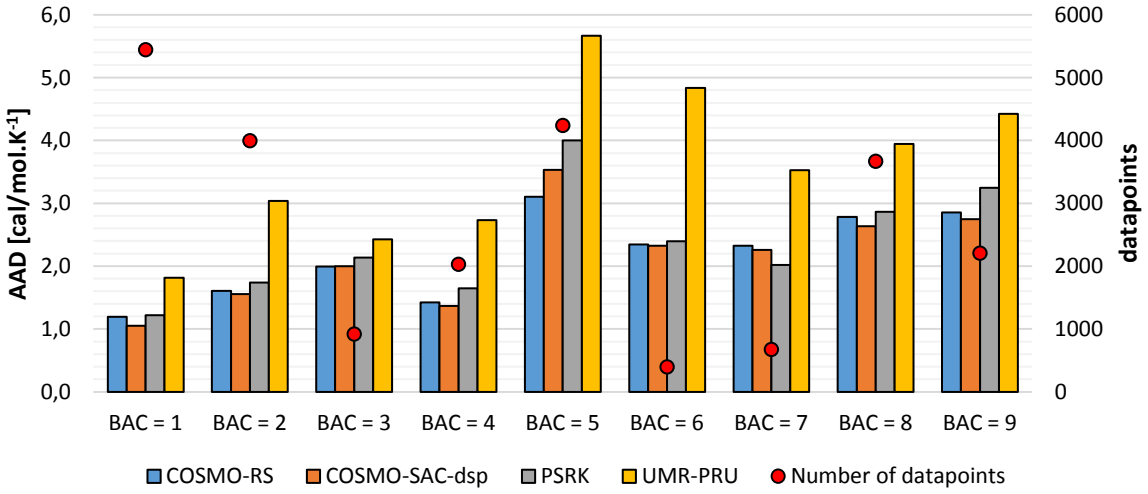

Figure 5 – Average absolute deviations between experimental solvation entropy ($\Delta_{solv}\bar{s}_i$) data and their prediction from 4 different models calculated for each binary association code (BAC).

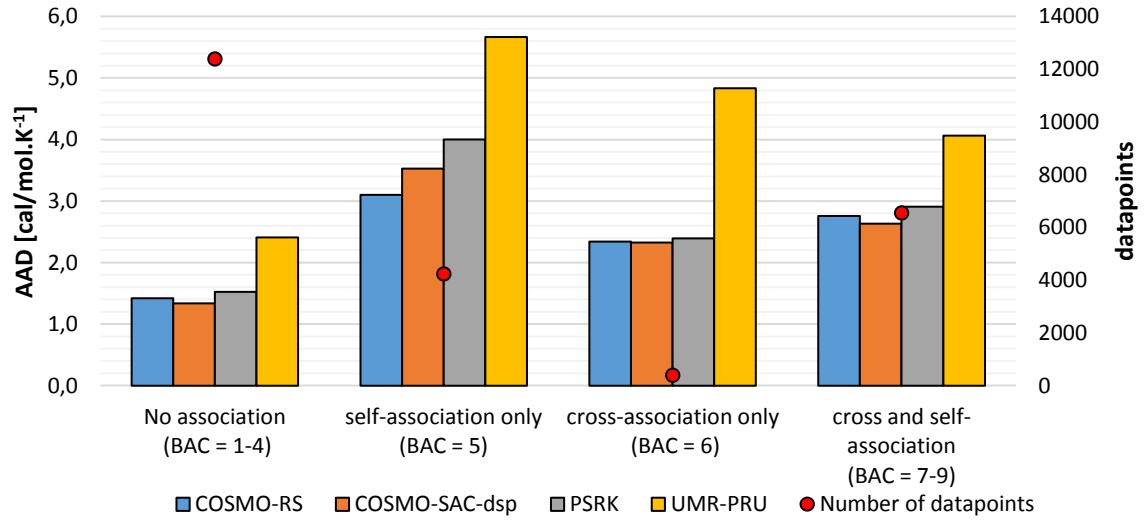

Figure 6 – Average absolute deviations between experimental solvation entropy ($\Delta_{solv}\bar{s}_i$) data and their prediction from 4 different models calculated for each type of association.

Similarly, an increasing tendency in the average absolute deviation was observed for mixtures in which hydrogen-bonding takes place, especially for mixtures involving a self-associating and a non-associating molecule (those of BAC 5). Whatever the binary association code, COSMO-RS, COSMO-SAC-dsp and PSRK correlated the experimental data with similar accuracy, while a significant increase in the deviation results was observed for the UMR-PRU EoS. Again, a loss of accuracy was reported concerning the hydration data. For such a case, COSMO-SAC-dsp provided better results.

It is important to note that the prediction of solvation enthalpy turned out to be less accurate than the prediction of solvation Gibbs energy, which agrees with the results obtained by Nait Saidi et al. [16]. This can be justified by the way used to estimate the $\Delta_{\mathrm{solv}}\bar{h}_i$ quantity which is deduced from preliminary calculations of $\Delta_{\mathrm{solv}}\bar{g}_i$ and $\Delta_{\mathrm{solv}}\bar{s}_i$ (see Eqs. (2) to (4) in the introduction); therefore, errors on the estimation of $\Delta_{\mathrm{solv}}\bar{h}_i$ cumulate those on $\Delta_{\mathrm{solv}}\bar{g}_i$ and $\Delta_{\mathrm{solv}}\bar{s}_i$.

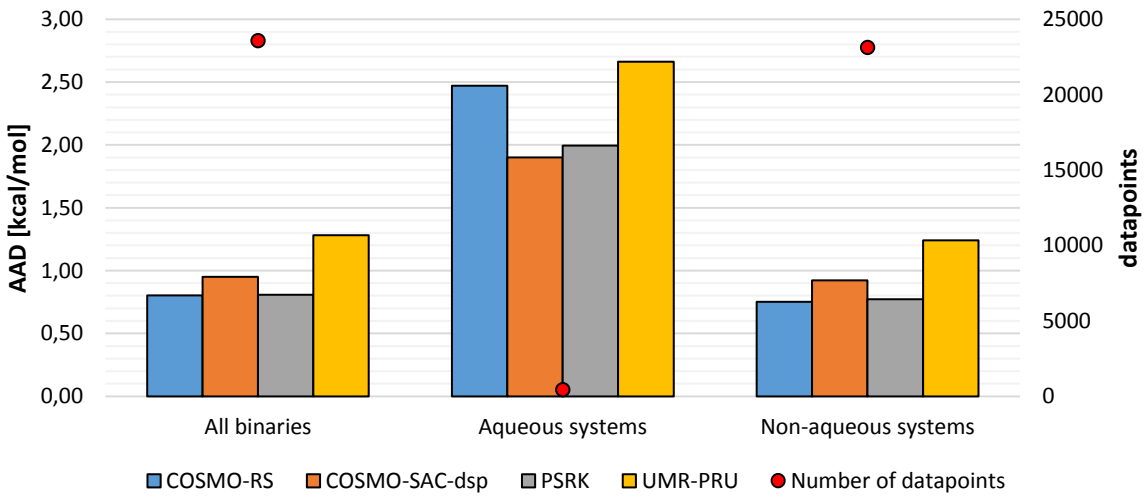

Figure 7 – Average absolute deviations between experimental solvation enthalpy (or equivalently, partial molar enthalpy of solvation $\Delta_{solv}\bar{h}_i$) data and their prediction from 4 models for aqueous and non-aqueous systems.

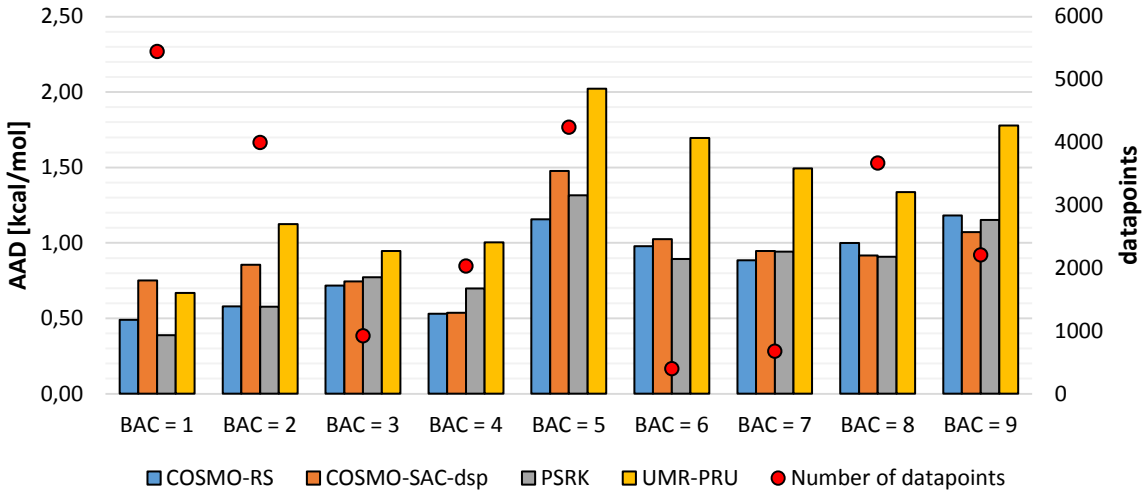

Figure 8 – Average absolute deviations between experimental solvation enthalpy ($\Delta_{solv}\bar{h}_i$) data and their prediction from 4 different models calculated for each binary association code (BAC).

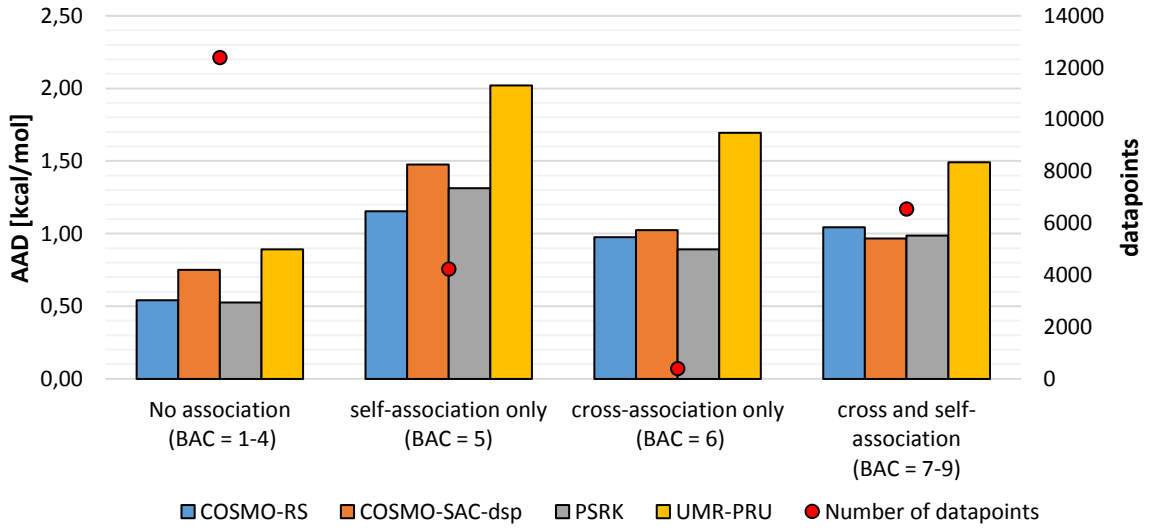

Figure 9 – Average absolute deviations between experimental solvation enthalpy ($\Delta_{solv}\bar{h}_i$) data and their prediction from 4 different models calculated for each type of association.

3.3 Probability distribution of deviation results

In order to verify if any of the models was providing a systematic under or overestimation, we also calculated the probability distribution of the signed deviations regarding the prediction of $\Delta_{\mathrm{solv}}\bar{g}_i$, $\Delta_{\mathrm{solv}}\bar{s}_i$, and $\Delta_{\mathrm{solv}}\bar{h}_i$, as shown in Figure 10.

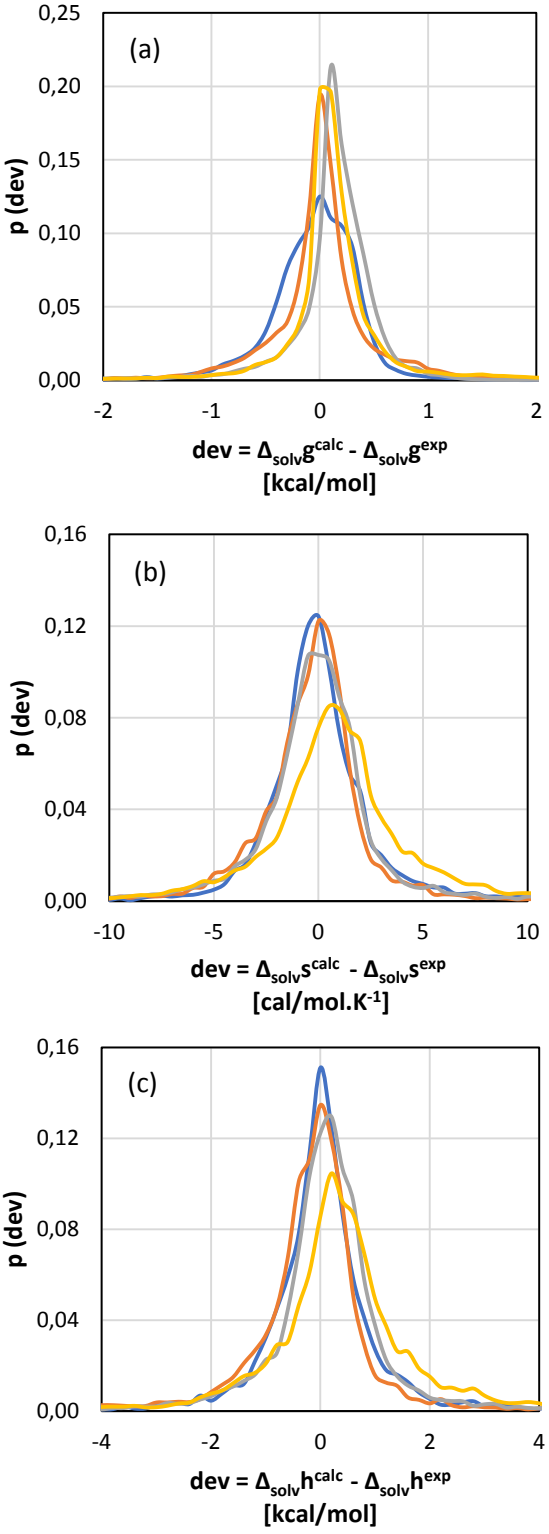

Figure 10 – Probability distribution of signed deviations between the calculated and experimental data of solvation Gibbs energy (a), solvation entropy (b) and solvation enthalpy (c) from 4 models (—COSMO-RS, —COSMO-SAC-dsp, —PSRK, —UMR-PRU).

For all models, the deviation distribution showed a gaussian behavior with mean close to zero. Only for the distribution of entropy and enthalpy deviations calculated by UMR-PRU, the mean values are slightly displaced from the origin of the abscissa axis.

4 CONCLUSIONS

In this paper, we have compared the accuracy of different methods (COSMO-based approaches and EoS) in the prediction of solvation energies of binary mixtures at infinite dilution. For this purpose, around 65,000 experimental points from the COMPSOL database were used as reference data. To analyze the results in more detail, the binary mixtures were sorted according to their degree of association (through hydrogen bonding).

As expected, all models (COSMO-RS, COSMO-SAC, PSRK and UMR-PRU) provided satisfactory predictions for non-associating binary mixtures. In such cases, the average absolute deviation did not exceed 0.30 kcal/mol for the calculation of the solvation Gibbs energy. For binary mixture with more complex association schemes, COSMO-RS and PSRK showed better performance than COSMO-SAC and UMR-PRU. A similar conclusion can be drawn with respect to entropy and enthalpy of solvation.

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.

REFERENCES

[1] M.L. Verteramo, O. Stenström, M.M. Ignjatović, O. Caldararu, 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.

[2] T. Hüfner-Wulsdorf, G. Klebe, Mapping Water Thermodynamics on Drug Candidates *via* Molecular Building Blocks: a Strategy to Improve Ligand Design and Rationalize SAR, J. Med. Chem. 64 (2021) 4662–4676. https://doi.org/10.1021/acs.jmedchem.0c02115.

[3] 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.

[4] H. Douroudgari, M. Vahedpour, A computer-aided method for controlling chemical resistance of drugs using RRKM theory in the liquid phase, Sci. Rep. 11 (2021) 22971. https://doi.org/10.1038/s41598-021-01751-z.

[5] C.W. Gao, J.W. Allen, W.H. Green, R.H. West, Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms, Comput. Phys. Commun. 203 (2016) 212–225. https://doi.org/10.1016/j.cpc.2016.02.013.

[6] 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.jpcb.1c02988.

[7] R. Van de Vijver, N.M. Vandewiele, P.L. Bhoorasingh, B.L. Slakman, F. Seyedzadeh Khanshan, H.-H. Carstensen, M.-F. Reyniers, G.B. Marin, R.H. West, K.M. Van Geem, Automatic Mechanism and Kinetic Model Generation for Gas- and Solution-Phase Processes: A Perspective on Best Practices, Recent Advances, and Future Challenges: AUTOMATIC MECHANISM GENERATION FOR GAS- AND SOLUTION-PHASE PROCESSES, Int. J. Chem. Kinet. 47 (2015) 199–231. https://doi.org/10.1002/kin.20902.

[8] 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.

[9] G. Duarte Ramos Matos, D.Y. Kyu, H.H. Loeffler, J.D. Chodera, M.R. Shirts, D.L. Mobley, Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database, J. Chem. Eng. Data. 62 (2017) 1559–1569. https://doi.org/10.1021/acs.jced.7b00104.

[10] 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.

[11] Y. Deng, B. Roux, Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations, J. Chem. Phys. 128 (2008) 115103. https://doi.org/10.1063/1.2842080.

[12] J. Tomasi, B. Mennucci, R. Cammi, Quantum Mechanical Continuum Solvation Models, Chem. Rev. 105 (2005) 2999–3094. https://doi.org/10.1021/cr9904009.

[13] J. Zhang, H. Zhang, T. Wu, Q. Wang, D. van der Spoel, Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic

Solvents, J. Chem. Theory Comput. 13 (2017) 1034–1043. https://doi.org/10.1021/acs.jctc.7b00169.
[14] 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.
[15] 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.
[16] 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.
[17] A. Ben-Naim, Solvation thermodynamics, Plenum Press, New York, 1987.
[18] 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.
[19] 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.
[20] A.I. Dragan, C.M. Read, C. Crane-Robinson, Enthalpy–entropy compensation: the role of solvation, Eur. Biophys. J. 46 (2017) 301–308. https://doi.org/10.1007/s00249-016-1182-6.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] S.-T. Lin, S.I. Sandler, A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model, Ind. Eng. Chem. Res. 41 (2002) 899–913. https://doi.org/10.1021/ie001047w.
[26] 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.
[27] 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.
[28] T. Holderbaum, J. Gmehling, PSRK: A Group Contribution Equation of State Based on UNIFAC, Fluid Phase Equilibria. 70 (1991) 251–265. https://doi.org/10.1016/0378-3812(91)85038-V.
[29] J. Gmehling, From UNIFAC to modified UNIFAC to PSRK with the help of DDB, Fluid Phase Equilibria. 107 (1995) 1–29. https://doi.org/10.1016/0378-3812(95)02720-Y.
[30] K. Fischer, J. Gmehling, Further development, status and results of the PSRK method for the prediction of vapor-liquid equilibria and gas solubilities, Fluid Phase Equilibria. 121 (1996) 185–206. https://doi.org/10.1016/0378-3812(95)02792-0.

[31] J. Chen, K. Fischer, J. Gmehling, Modification of PSRK mixing rules and results for vapor–liquid equilibria, enthalpy of mixing and activity coefficients at infinite dilution, Fluid Phase Equilibria. 200 (2002) 411–429. https://doi.org/10.1016/S0378-3812(02)00048-1.
[32] S. Horstmann, A. Jabłoniec, J. Krafczyk, K. Fischer, J. Gmehling, PSRK group contribution equation of state: comprehensive revision and extension IV, including critical constants and α-function parameters for 1000 components, Fluid Phase Equilibria. 227 (2005) 157–164. https://doi.org/10.1016/j.fluid.2004.11.002.
[33] E. Voutsas, K. Magoulas, D. Tassios, Universal Mixing Rule for Cubic Equations of State Applicable to Symmetric and Asymmetric Systems: Results with the Peng−Robinson Equation of State, Ind. Eng. Chem. Res. 43 (2004) 6238–6246. https://doi.org/10.1021/ie049580p.
[34] H.K. Hansen, B. Cota, B. Kuhlmann, UNIFAC with Linearly Temperature-Dependent Group-Interaction Parameters, (1992) Institut for kemiteknik, DTU, Lyngby, Denmark, SEP 9212.
[35] 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.
[36] I. Khan, K.A. Kurnia, F. Mutelet, S.P. Pinho, J.A.P. Coutinho, Probing the Interactions between Ionic Liquids and Water: Experimental and Quantum Chemical Approach, J. Phys. Chem. B. 118 (2014) 1848–1860. https://doi.org/10.1021/jp4113552.
[37] A.F. Young, F.L.P. Pessoa, V.R.R. Ahón, Comparison of volume translation and co-volume functions applied in the Peng-Robinson EoS for volumetric corrections, Fluid Phase Equilibria. 435 (2017) 73–87. https://doi.org/10.1016/j.fluid.2016.12.016.
[38] G.M. Kontogeorgis, G.K. Folas, Thermodynamic Models for Industrial Applications, John Wiley & Sons, Ltd, Chichester, UK, 2010. https://doi.org/10.1002/9780470747537.
[39] G.M. Kontogeorgis, P.M. Vlamos, An interpretation of the behavior of EoS/GE models for asymmetric systems, Chem. Eng. Sci. 55 (2000) 2351–2358. https://doi.org/10.1016/S0009-2509(99)00472-8.

APPENDIX 1

Table 2 – Average absolute deviations (AAD) and standard deviations (SD) between experimental solvation Gibbs energy (or, more precisely, chemical potential of solvation) data and their prediction with COSMO-based and equation-of-state approaches.

Deviations results $\Delta_{solv}\bar{g}_i^{\infty}$	N_{points}	COSMO-RS		COSMO-SAC-dsp		PSRK		UMR-PRU	
		AAD [kcal/mol]	SD [kcal/mol]	AAD [kcal/mol]	SD [kcal/mol]	AAD [kcal/mol]	SD [kcal/mol]	AAD [kcal/mol]	SD [kcal/mol]
All binaries	56603	0.29	0.30	0.35	0.63	0.32	0.49	0.34	0.65
Aqueous systems	5511	0.41	0.50	0.82	1.34	0.68	0.83	1.00	1.02
Non-aqueous systems	51092	0.28	0.27	0.30	0.46	0.28	0.42	0.26	0.55
BAC = 1	8047	0.22	0.18	0.14	0.18	0.18	0.20	0.12	0.41
BAC = 2	8005	0.24	0.19	0.20	0.27	0.21	0.28	0.16	0.27
BAC = 3	1784	0.19	0.15	0.18	1.04	0.25	0.45	0.21	0.50
BAC = 4	3930	0.24	0.20	0.14	0.21	0.21	0.21	0.17	0.41
BAC = 5	14011	0.32	0.34	0.52	0.61	0.41	0.70	0.43	0.73
BAC = 6	1007	0.21	0.16	0.16	0.17	0.27	0.37	0.33	1.08
BAC = 7	1720	0.25	0.25	0.49	0.63	0.51	0.80	0.39	0.57
BAC = 8	10265	0.36	0.40	0.43	0.84	0.35	0.44	0.41	0.72
BAC = 9	7834	0.32	0.33	0.45	0.77	0.38	0.43	0.57	0.77
No association (BAC = 1-4)	21766	0.23	0.19	0.16	0.37	0.20	0.26	0.15	0.37
self-association only (BAC = 5)	14011	0.32	0.34	0.52	0.61	0.41	0.70	0.43	0.73
cross-association only (BAC = 6)	1007	0.21	0.16	0.16	0.17	0.27	0.37	0.33	1.08
cross and self-association (BAC = 7-9)	19819	0.34	0.36	0.44	0.79	0.38	0.48	0.47	0.73

Table 3 – Average absolute deviations (AAD) and standard deviations (SD) between experimental solvation entropy (or, more precisely, partial molar entropy of solvation) data and their prediction with COSMO-based and equation-of-state approaches.

Deviations results		COSMO-RS		COSMO-SAC-dsp		PSRK		UMR-PRU	
$\Delta_{solv}\bar{s}_i^{\infty}$	N_{points}^{tot}	AAD [cal/mol.K^{-1}]	SD [cal/mol.K^{-1}]	AAD [cal/mol.K^{-1}]	SD [cal/mol.K^{-1}]	AAD [cal/mol.K^{-1}]	SD [cal/mol.K^{-1}]	AAD [cal/mol.K^{-1}]	SD [cal/mol.K^{-1}]
All binaries	23586	2.1127	11.3976	2.1095	11.3644	2.3699	11.4305	3.4955	12.8297
Aqueous systems	448	6.7120	5.1403	6.0145	5.4465	8.1271	5.5060	6.9827	8.1061
Non-aqueous systems	23138	2.0236	11.4670	2.0338	11.4357	2.2585	11.4867	3.4279	12.8948
BAC = 1	5445	1.1929	1.5104	1.0545	1.2849	1.2211	1.3826	1.8188	5.8779
BAC = 2	3997	1.6078	2.9433	1.5583	2.9391	1.7387	3.0603	3.0413	7.3874
BAC = 3	921	1.9922	4.2126	1.9982	4.2634	2.1394	4.2782	2.4298	5.0790
BAC = 4	2031	1.4220	2.1520	1.3675	2.1003	1.6499	2.2945	2.7346	2.7952
BAC = 5	4240	3.1030	25.9568	3.5305	25.8987	4.0006	25.9434	5.6694	27.1489
BAC = 6	398	2.3441	2.7709	2.3254	2.6162	2.3953	2.2935	4.8350	4.8055
BAC = 7	676	2.3251	2.3577	2.2597	2.1099	2.0210	2.3240	3.5300	3.0890
BAC = 8	3670	2.7838	4.2289	2.6367	4.0958	2.8666	4.4770	3.9430	6.0326
BAC = 9	2208	2.8567	4.6770	2.7475	4.6004	3.2496	4.7266	4.4260	7.3806
No association (BAC = 1-4)	12394	1.4236	2.4356	1.3384	2.3808	1.5265	2.4837	2.4085	6.0229
self-association only (BAC = 5)	4240	3.1030	25.9568	3.5305	25.8987	4.0006	25.9434	5.6694	27.1489
cross-association only (BAC = 6)	398	2.3441	2.7709	2.3254	2.6162	2.3953	2.2935	4.8350	4.8055
cross and self-association (BAC = 7-9)	6554	2.7610	4.2403	2.6351	4.1233	2.9084	4.4078	4.0631	6.3084

Table 4 – Average absolute deviations (AAD) and standard deviations (SD) between experimental solvation enthalpy (or, more precisely, partial molar enthalpy of solvation) data and their prediction with COSMO-based and equation-of-state approaches.

Deviations results		COSMO-RS		COSMO-SAC-dsp		PSRK		UMR-PRU	
$\Delta_{solv}\bar{h}_i^{\infty}$	N_{points}^{tot}	AAD [kcal/mol]	SD [kcal/mol]	AAD [kcal/mol]	SD [kcal/mol]	AAD [kcal/mol]	SD [kcal/mol]	AAD [kcal/mol]	SD [kcal/mol]
All binaries	23586	0.80	4.35	0.95	4.53	0.81	4.34	1.28	4.82
Aqueous systems	448	2.47	1.67	1.90	1.58	2.00	1.54	2.66	2.61
Non-aqueous systems	23138	0.75	4.39	0.92	4.58	0.77	4.39	1.24	4.87
BAC = 1	5445	0.49	0.68	0.75	2.13	0.39	0.48	0.67	1.72
BAC = 2	3997	0.58	0.94	0.85	1.76	0.58	0.96	1.12	3.04
BAC = 3	921	0.72	1.72	0.74	1.69	0.77	1.73	0.95	2.13
BAC = 4	2031	0.53	0.74	0.54	0.77	0.70	0.87	1.00	1.07
BAC = 5	4240	1.16	9.72	1.48	9.82	1.31	9.71	2.02	10.09
BAC = 6	398	0.98	1.10	1.02	1.20	0.89	1.07	1.70	1.77
BAC = 7	676	0.88	0.93	0.95	0.96	0.94	1.01	1.49	1.38
BAC = 8	3670	1.00	1.35	0.92	1.27	0.91	1.42	1.34	1.87
BAC = 9	2208	1.18	1.76	1.07	1.68	1.15	1.67	1.78	2.80
No association (BAC = 1-4)	12394	0.54	0.89	0.75	1.82	0.53	0.86	0.89	2.22
self-association only (BAC = 5)	4240	1.16	9.72	1.48	9.82	1.31	9.71	2.02	10.09
cross-association only (BAC = 6)	398	0.98	1.10	1.02	1.20	0.89	1.07	1.70	1.77
cross and self-association (BAC = 7-9)	6554	1.05	1.46	0.97	1.39	0.99	1.47	1.49	2.18