COSMO-RS for aqueous solvation and interfaces

By Andreas Klamt[1,2]

[1] COSMO*logic* GmbH&CoKG, Imbacher Weg 46, 51379 Leverkusen, Germany

[2] Inst. of Physical and Theoretical Chemistry, University of Regensburg, Germany

* corresponding author, Tel.: +49-2171-731681, e-mail: klamt@cosmologic.de

keywords: COSMO-RS; hydration; solubility; water interfaces; aqueous micelles

Abstract

The quantum chemically based conductor-like screening model for realistic solvation COSMO-RS is a rather general method for predicting the chemical potentials – and thus the activity coefficients – of almost arbitrary molecules in almost any dense pure or mixed liquid. Originally it was developed for the prediction of partition coefficients and Henry's law constants which are important in environmental and life science, almost all of them involving the aqueous phase. Partitioning between organic phases, as essential in many chemical engineering applications, was considered only in later stages. Therefore the solvent water had and still has a great importance in the parameterization and validation of COSMO-RS with the result, that aqueous solvation in generally is well described. In this article we review the performance of COSMO-RS for aqueous solvation, discuss some problems arising for this special fluid water, and review some recent extension of COSMO-RS to pK_a prediction, electrolyte systems, and aqueous interfaces.

1. Introduction

The quantum chemically based Conductor-like Screening Model (COSMO) [1], which is a modification of the general class of dielectric continuum solvation models [2], and its extension to Realistic Solvation (COSMO-RS) [3-6] were originally developed at the chemical company Bayer in projects focusing on the prediction of environmental distribution and toxicity. Therefore the prediction of aqueous Henry's law constants, i.e. ΔG_{hydr}, octanol-water and alkane-water partition coefficients, adsorption from water to soil, and similar equilibrium constants had been the primary concern of COSMO-RS. Almost all of these environmentally relevant properties involve the aqueous phase. The same is true for life science and medicinal chemistry applications which also are of great interest for industrial computational chemistry. Only much later the method was taken up by chemical engineers, who most often consider activity coefficients of organic molecules in organic solvents. Water as a

solvent is definitely also of importance in chemical engineering, but it does not play such an exceptional role as in life science and environmental science. Therefore the proper description of aqueous solvation was of primary importance during the development of COSMO-RS with the result, that aqueous solvation was highly represented in the parameterization and validation data sets of the original versions of COSMO-RS and still is in the COSMO*therm* line of COSMO-RS versions.

The success of COSMO-RS has stimulated several re-implementations, e.g. COSMO-SAC, COSMO-RS(Banerjee), COSMO-RS(Ol), and the ADF-COSMO-RS [7-10]. Most of these are parameterized based on more chemical engineering focused data sets and therefore may have different and most likely worse performance with respect to aqueous solvation than the COSMO-RS and COSMO*therm* versions in the original COSMO-RS tradition. Due to the lack of representative data for these COSMO-RS re-implementations, this article will focus on the original COSMO-RS development line.

The article will be structured as follows: In chapter 2 a short introduction to the basic principles of COSMO-RS will be given. Chapter 3 considers the description and the special features of the water molecule in the framework of COSMO-RS. Chapter 4 gives an overview of the accuracy of COSMO-RS for hydration free energies and activity coefficients of neutral compounds in water. Chapter 5 concerns ions in aqueous solution, with specific consideration of pK_a and recent approaches to for electrolyte systems. Chapter 6 covers the extension of COSMO-RS to interfaces, interfacial tensions and micellar systems. In chapter 7 we summarize the results and provide an outlook for improvements.

2. The COSMO-RS method

A detailed description of COSMO-RS is beyond the scope of this review and given elsewhere [3-8]. Thus only the basic concept of the method shall be presented here.

COSMO-RS is composed of two fundamental steps. First quantum chemical calculations have to be performed for all compounds of interest. In these calculations a virtual conductor embedding the molecule is taken into account by the continuum solvation model COSMO [1,2]. The molecule induces a polarization charge density σ on the interface to the conductor, and these charges interact with the molecule, generating a more polarized electron density than in vacuum. Simultaneously solving the quantum chemical equations and the conductor boundary condition, the solute molecule is optimized with respect to electron density and geometry to its energetically optimal state in a conductor, and the energies, geometries and polarization charge densities on the

surface segments get stored in a COSMO file. The standard quantum chemical method used in the COSMO*therm* versions of COSMO-RS since 2001 is density functional theory (DFT) and the DFT functional B88-P86 [11,12] with a triple zeta valence polarized basis set (TZVP) and the RI approximation [13,14]. This has been used throughout most applications mentioned in this article, unless the results are declared as BP-TZVPD-FINE level. This more recent level involves a refined construction of the COSMO cavity and adds diffuse functions to the TZVP basis set. All DFT/COSMO calculations have been performed by the quantum chemical program TURBOMOLE [15,16], although other programs may be used for the generation of such COSMO files as well.

In the second step of COSMO-RS, the statistical thermodynamics of the molecular interactions, this polarization charge density is used for the quantification of the interaction energy of pairs of surface segments. As most important molecular interaction modes, electrostatics and hydrogen bonding is taken into account in this way. The electrostatic energy deviations to the conductor reference state are taken into account as misfit energies

$$E_{misfit}(\sigma,\sigma') = \tfrac{1}{2} a_{eff} \alpha_{misfit} (\sigma + \sigma')^2 \qquad (1)$$

and the additional hydrogen bond interaction energy is quantified approximately as

$$E_{hb}(\sigma,\sigma') = a_{aeff} c_{hb}(T) \min(0, \sigma\sigma' - \sigma_{hb}^2) \qquad (2)$$

where σ and σ‘ are the polarization charge densities of the two interacting surface segments and a_{eff} is the effective size of a thermodynamically independent contact. The value of a_{eff} usually is fitted to during the COSMO-RS parameterization and turns out to in the order of 7Å², which corresponds to a coordination number of ~6 for a molecule as water. Although not derived from geometrical arguments, this nicely agrees with values for the coordination number as typically found in molecular dynamics calculations. α_{misfit} is the misfit energy coefficient, which can be reasonably well derived from theoretical considerations, but finally is fine-tuned vs. experimental data. It takes into account the average electronic polarizability of the neighboring molecules. The empirically adjusted hydrogen bond coefficient $c_{hb}(T)$ has a temperature dependence which expresses the entropy loss going along with hydrogen bond formation. σ_{hb} is a kind of minimum polarity for hydrogen bond formation. The less specific dispersive interactions are described to first order based on element specific surface energies.

The quantum chemical information about the polarization charge densities σ plays the key role for the evaluation of the molecular interactions in the liquid phase. As a preparation for an efficient statistical thermodynamics treatment, the 3D distribution of the polarization charge densities σ on the surface of each molecule X is converted into a surface composition function $p^X(\sigma)$, usually called σ-profile. It describes the amount of molecular surface with polarity σ on the surface of molecule X. σ-profiles provide detailed information about the molecular polarity distribution [4-6]. The σ-profiles of water and five representative organic compounds are shown in figure 1. For the purpose of the calculation of vapor-liquid equilibria, separate vacuum quantum chemical calculations are performed for each molecule and its relevant conformations.

Based on the surface pair interaction model introduced above, the statistical thermodynamics itself is done using a coupled set of non-linear equations for the activity coefficients of the surface segments, the so-called COSMOSPACE equations [18], which are equivalent to an exact solution of the quasi-chemical approximation of Guggenheim [19]. This results in a solvent specific free energy response function $\mu_S(\sigma)$, called σ-potential, which gives the chemical potential of a surface segment of polarity σ in the solvent S. Finally, the chemical potentials of the compounds in a pure or mixed solvent are calculated by summation of the σ-potentials of the surface segments of a compound and slightly corrected by an empirical combinatorial term which takes into account surface areas and volumes of solutes and solvents resulting from the COSMO cavities.

Flexible compounds are represented as ensembles of minimum energy geometries, called conformations or conformers, and the free energy of the compound in a solvent or mixture is consistently calculated from the conformer partition function, based on the sum of DFT/COSMO energies and COSMO-RS chemical potentials of the individual conformers. This ensures that the proper conformations being most favorable in the current solvent of interest is dominating the thermodynamic property calculations. In rare, but significant cases this multiple conformation treatment, which is only available in the COSMO*therm* line of COSMO-RS, can be crucial, since the most important conformation may change from one solvent to the other. An extreme example is 4-hydroxy-2-butanone, which can form a very nice intra-molecular 6-ring hydrogen bond. In non-polar environments this dominates completely, e.g. in hexane we find it 94% populated, and in the gas phase even 97%, while in water this reduces to 36%, and in methylamine it is only 1.6% populated.

While in the original version of COSMO-RS the polarization charge density σ was the only local surface descriptor taken into account, later versions make use of additional descriptors. Examples are the $\sigma^{\perp}$ descriptor [4], expressing the local variance of σ, or the local polarizability, or just the atomic number of the underlying element. Nevertheless, the polarization charge density σ keeps being the most important descriptor for the local interactions of the surface segments.

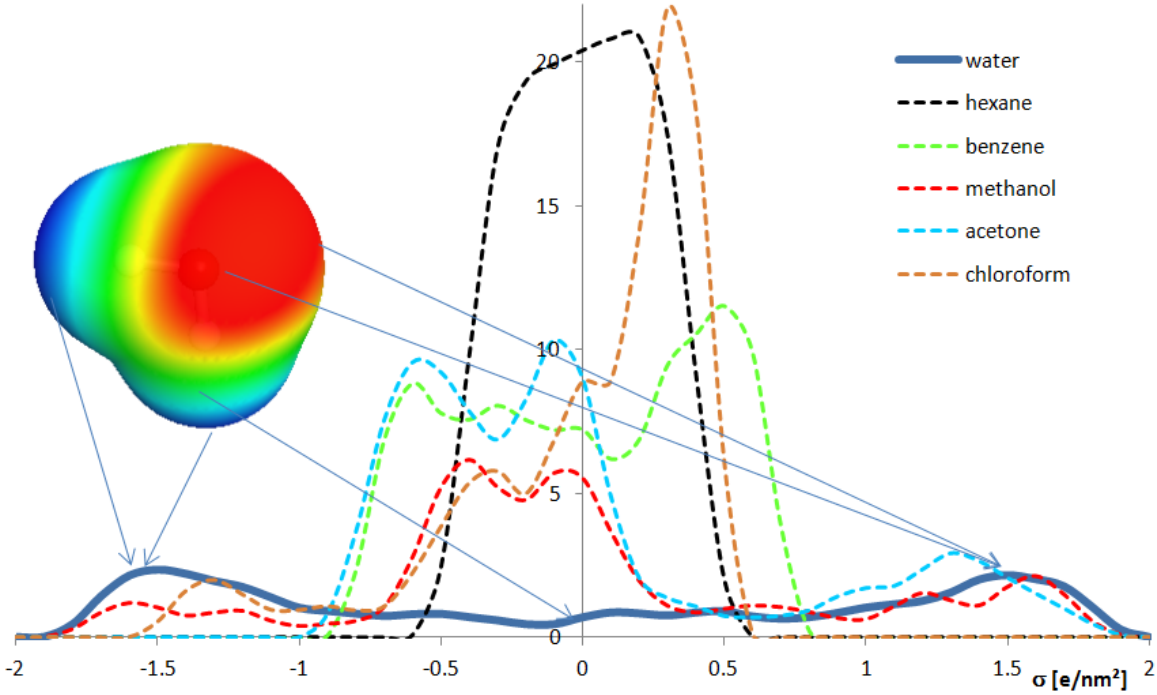

Figure 1: σ-surface of water and σ-profiles of water and 5 other solvents

3. The description of water in the COSMO-RS framework

Figure 1 shows the σ-surface and the σ-profile of water, together with those of five representative solvents. The σ-surface of water is dominated by the blue regions of strongly negative σ on the hydrogen bond donors and the red regions of the oxygen lone pairs, which are hydrogen bond acceptors. These hydrogen bonding areas cause two almost symmetric peaks in the σ-profile of water, which are centered at ±1.6 e/Å², respectively, and dominate the σ-profile. The amount of non-polar, green surface area is relatively small. Indeed, it is the smallest within all solvents considered so far. The exceptional breadth and symmetry of the σ-profile and the relatively low portion of non-polar surface area are responsible for the unique properties of water, because on the one hand it almost perfectly provides the optimal partners of opposite polarity for all surface segments, and on the other hand it has a very small concentration of suitable partner surfaces for non-polar, i.e. green surface segments of solutes.

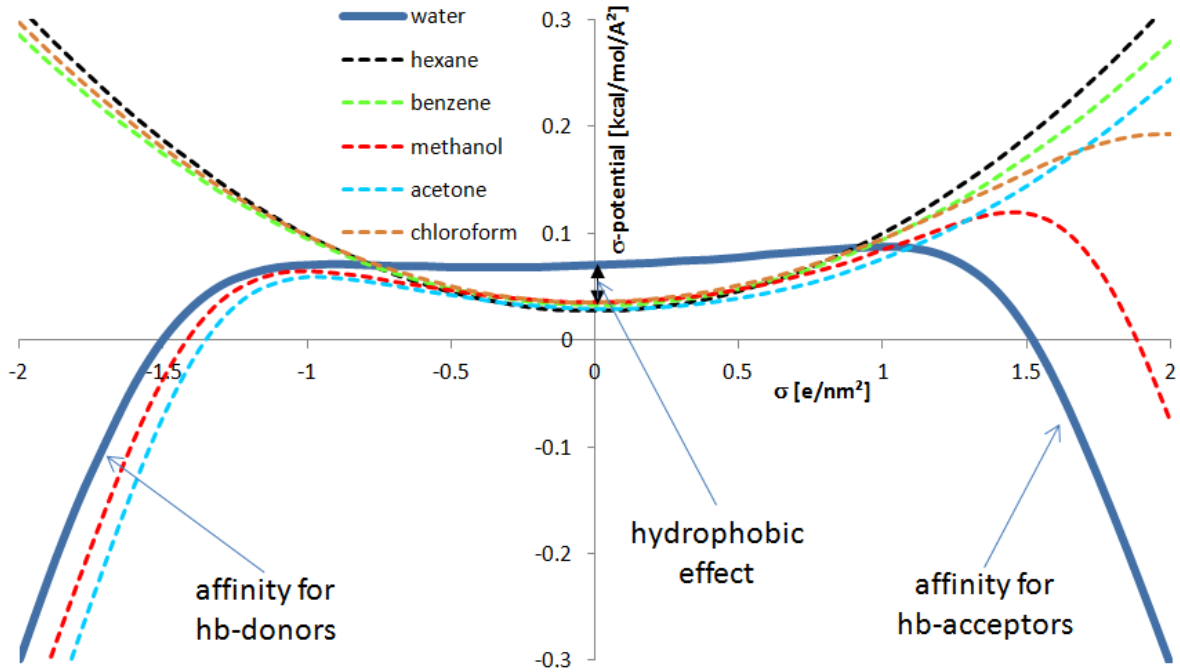

Figure 2: σ-potentials of water and five representative organic solvents

The solvent behavior of water can be better studied in the resulting σ-potential as illustrated in figure 2. In the less and medium polar range, i.e. for $|\sigma| < 1.2 e/nm^2$, we see an almost flat σ-potential with a very small curvature, which essentially corresponds to the electrostatic response of a strong dielectric medium, but in the outer ranges hydrogen bonding causes strongly negative chemical potentials, i.e. strong attractions, for hydrogen bond donor surface segments (left side) and acceptor segments (right side). It should be noticed, that methanol, although even having slightly less polar acceptors as water, as can be seen in σ-profiles, is more attractive for donors. This is a result of the identical number of donors and acceptors in water, which has the consequence, that in water almost no free acceptors are available for additional donors, while in methanol, which has only one donor, but two acceptors, plenty of free acceptors are available. On the other hand, water is one of the most attractive solvents for hydrogen bond donors, because almost all other solvents have an excess of acceptors. A very special feature of the σ-potential, i.e. of the solvent behavior of water is the relatively high chemical potential of non-polar surface in water (see the range $|\sigma| < 0.5 e/nm^2$ in the σ-potential). Therefore non-polar surface area prefers any other solvent compared to water, a behavior, which is well known as hydrophobicity. Because at room-temperature the strong hydrogen bonds between the polar parts of water can hardly be broken, only the small percentage of non-polar (green) surface area of water is available for non-polar molecular surface

dissolving in water, resulting in a strong loss of entropy associated with each non-polar surface contact in water and thus explaining the entropic nature of hydrophobicity [20].

While the basic features of water can be nicely expressed in terms of the σ-profile and σ-potential, it turned out already in the first quantitative parameterization of COSMO-RS [4] that water as a solute gets too strongly negative chemical potential in all polar solvents and especially in water. The reason for this is still not completely understood. As an empirical way out it is convenient to scale the surface area of water during the COSMO-RS calculations by an empirical factor in the range of 0.7. Since the solvent behavior does only depend on the normalized σ-profiles, such scaling keeps the solvent behavior or water unchanged and selectively modifies its solute behavior. Similar or slightly different water specific corrections are also applied in the COSMO-RS re-implementations, e.g. in COSMO-SAC a scaling of the σ-profile along the σ-axis was applied in order to reduce the polarity of water. Only the most recent BP-TZVPD-FINE15 parameterization of COSMO-RS, which includes an improved COSMO cavity construction, a larger basis set, a more detailed description of hydrogen bonding [21,22] and a final dielectric correction term for correlated misfit charge densities, is able to describe the solute water without such special fitting.

4. Water as a solvent for neutral solutes

The prediction of the free energy, activity coefficients, and solubility of neutral compounds in water is of great practical importance. One quantity, which gives quite unbiased information about the ability of a model to predict aqueous solvation is the free energy of hydration, which is identical with the Henry's law constant for vaporization from water. Depending on the community, the one or the other expression is preferred. A reference data set of ΔG_{hydr} values for 274 solutes was provided by Marenich et al. in the context of the SM_x series of solvation models [23]. Using a COSMOtherm parameterization from 2007, on this data set we achieved an accuracy of 0.58 kcal/mol (MUE), without having used these data in the parameterization [24]. Recently we achieved an accuracy of 0.52 kcal/mol (MUE) on these data [25], using the latest BP-TZVPD-FINE15 parameterization of COSMOtherm, corresponding to an average error of a bit more than a factor of 2 in the related Henry's law constants. For comparison, the accuracy on the full dataset of 2346 solvation free energies, mainly in non-aqueous solvents, was 0.49 and 0.42 kcal/mol (MUE), respectively. Hence the prediction error for aqueous solvation appears to be slightly larger for aqueous than for non-aqueous solvation, which is not surprising, since the aqueous solvation data also have a larger variance, i.e. the range of hydration free energies span a range from -14.1 to 4.3 kcal/mol and have a variance of 3.3 kcal/mol, while the non-aqueous solvation data range from -14.1 to 1.7 kcal/mol with a variance of only 2 kcal/mol. It needs to be said the experimental

data quite definitely contribute an experimental error the deviations, which may be conservatively estimated as 0.3 kcal/mol.

An early comparison of the COSMO-RS prediction accuracy on aqueous and non-aqueous infinite dilution partition coefficients (IDACs) [26] was already published in 2003. In that paper the performance COSMO-RS for 106 aqueous IDACs was 44 %AAD (i.e. % absolute deviation), while the accuracy on 297 non-aqueous data was 46%AAD, corresponding to an error of 0.67 ln-units or 0.4 kcal/mol MUE. It was shown that for these relatively small compounds, the various UNIFAC variants [27] were more accurate for the non-aqueous solvents, but performed much worse for aqueous IDACs, while COSMO-RS had rather robust error bars for both subsets. Again it must be noted, that a major part of the error most likely results from wrong experimental data, e.g. from 2-pentanone, the experimental of which was completely out of the trend of the other ketones.

The latest released version of COSMOtherm [28] achieves an accuracy of 0.5 ln-units MUE for a validation set of 560 aqueous logarithmic IDACs not used for the parameterization. The performance on 8318 non-aqueous IDACs is slightly better, i.e. 0.28 ln-units MUE. Summarizing, aqueous activity coefficients and aqueous solvation free energies are predicted with approximately the same accuracy as those for non-aqueous solvents. Mixtures can be easily treated by COSMO-RS, introducing no additional error. Therefore the error bars derived for the solvation of pure water also apply to the aqueous mixtures. An example is for a larger drug molecule is shown in figure 3.

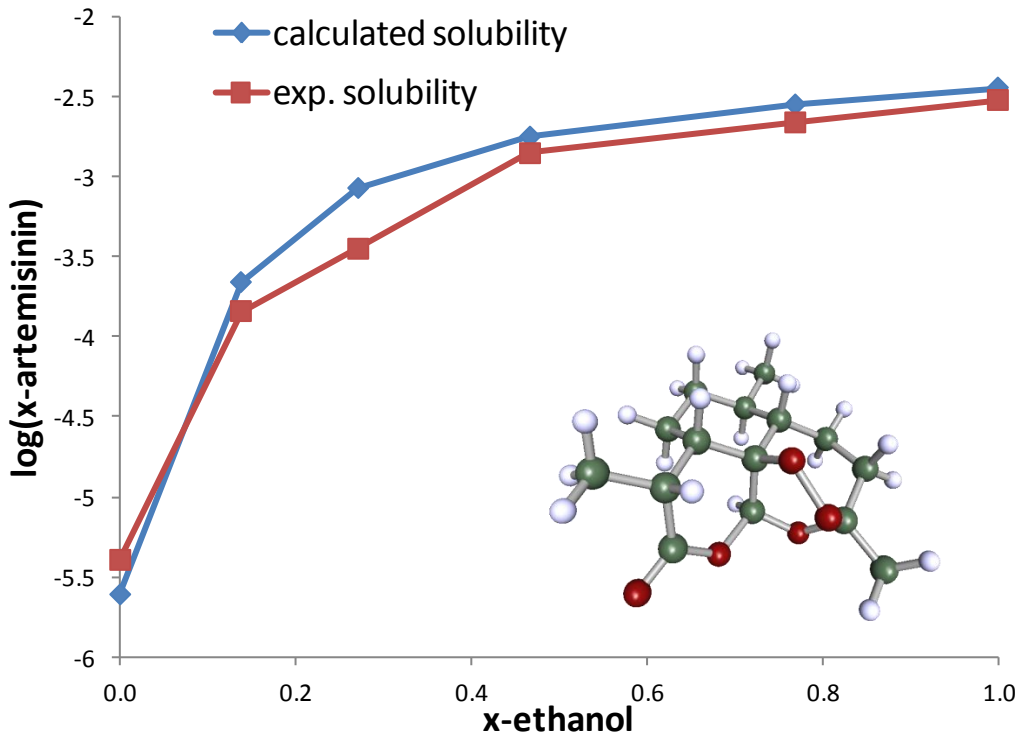

Figure 3: Comparison of COSMO-RS predicted and experimental solubilities of artemisinin in ethanol–water mixtures at 303 K as reported by Lapkin et al. [29]

5. Aqueous solvation of ions

The COSMO-RS method was originally developed solely for neutral solutes and solvents. For ions COSMO-RS theory bears some severe problems. One problem is the very high polarization charge density, which can occur on small or highly charged ions, or if the ionic charge is very localized in an ion. Such high polarization charge density has the effect that the most polar available solvent surface segments of the opposite polarity get associated with the ionic surface during the COSMO-RS thermodynamics. If the polarization charge density of the solvent is not high enough to compensate the polarization charge density on the ionic surface, then all the polar surface segments get a misfit charge of the same sign. The energetic costs for such correlation of misfit charges are not accounted for in the COSMO-RS thermodynamics, and therefore COSMO-RS will fail to deal with ions which have a polarization charge density outside the range of the solvent σ-profile. The most severe and most important cases for such high charge densities are the alkali and earth-alkali metal cations (see figure 4) The situation is less critical for the anions, since for these the polarization charge densities are in the range of the maximum polarization charge densities of the water donors, just with the flouride anion being somewhat

outside the water range. The hydroxyl anion is very far out of the range of water, as well as its counterpart, the hydronium cation.

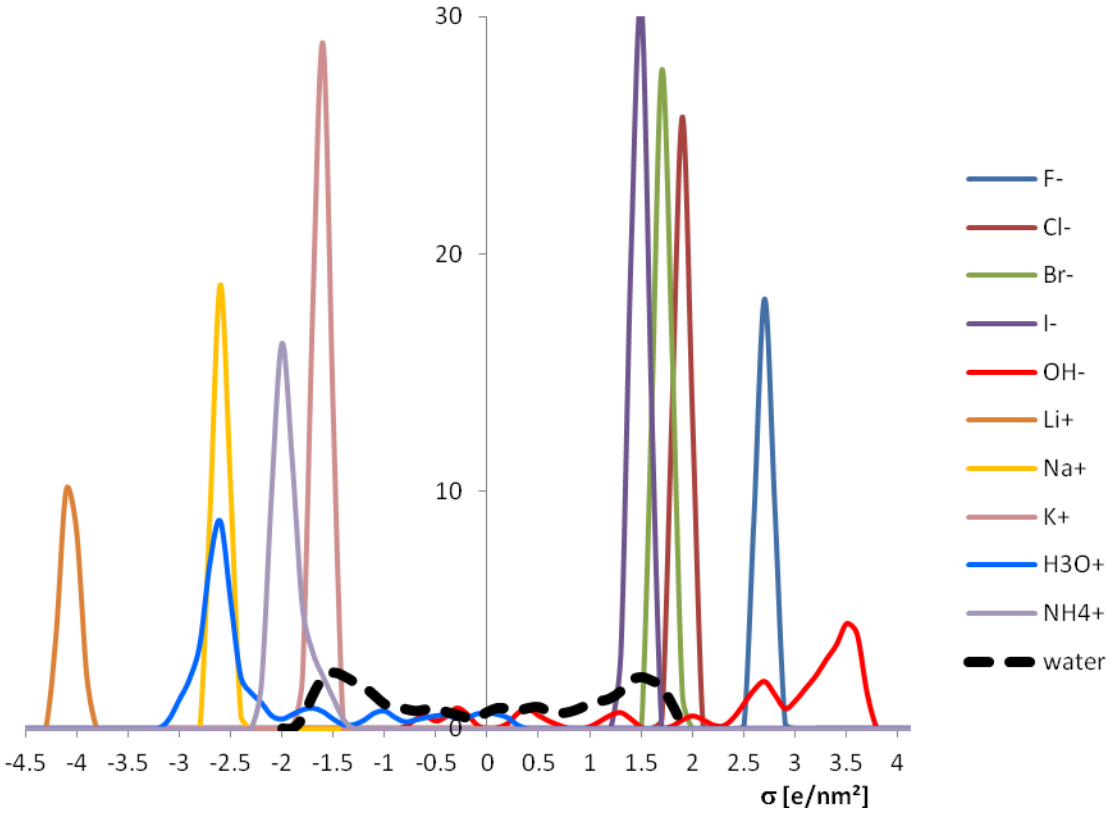

Figure 4: σ-profiles of small ions. The profiles for spherical ions are slightly broadened for better visualization.

Another fundamental problem for COSMO-RS arises in dilute electrolyte systems, in which the long range interactions of ions lead to the Debye-Hückel attraction. Since COSMO-RS only takes into account short range interactions, the Debye-Hückel effect cannot be described by COSMO-RS itself. Nevertheless, it is straight forward to supplement COSMO-RS with an external expression for the Debye-Hückel energy, in a similar way as done for the external combinatorial term, as it has been done in other G^{ex}-models before [30]. Such approach should work very well in the limit of low ionic strengths, but at very high ionic strength the Debye-Hückel expressions partially include nearest neighbor interactions of ions. That raises the danger of double-counting, because direct interactions of ions are already covered by COSMO-RS.

Two different pathways have been followed in order to overcome the problems arising from the high charge densities on ions. The first approach is straight forward from a physical point of view: At very high charge densities the interactions of ions and their nearest solvent molecules become very strong and can be considered

as ion-solvent cluster formation, or in other words the formation of rather tightly bound solvation shells, as they are well known for small ions as Li^+. Therefore they should be described within the quantum chemical part of COSMO-RS as explicit solute-solvent clusters. On the outer surface of such clusters the ionic charge is much better distributed and the σ-profile most often is back in the range of typical σ-profiles. Andersson and Stipp [31] applied the explicit first hydration shell approach systematically to 40 ions with charges ranging from -3 to +4. As shown in figure 5, they achieved very good correlations and slopes close to unity vs. experimental ΔG_{hydr}. Especially for the singly and doubly charged ions they demonstrated a clear improvement resulting from COSMO-RS over a treatment in which the clusters are just handled by COSMO continuum solvation. Since they considered only ions in infinite dilution, they did not need to add a Debye-Hückel expression.

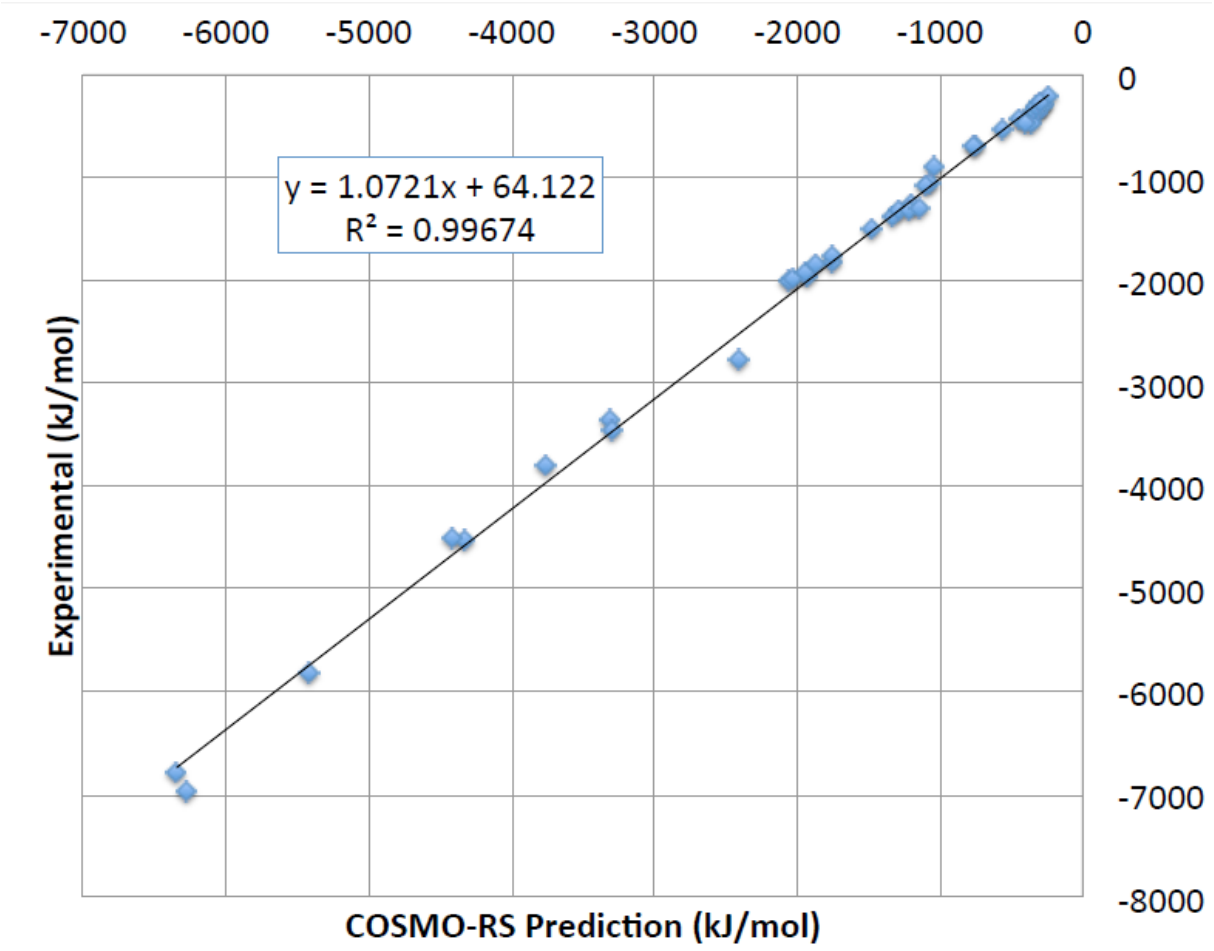

Figure 4: Experimental and calculated free energies of hydration for ions of valency -3 to +4 according to Andersson et al. [31].

Toure et al. followed a similar pathway for the simulation of activity coefficients in electrolyte systems [32-34]. Instead of introducing complete explicit first solvation shells they determined the minimum number of explicit water molecules required for the transferable description of the electrolyte behavior of each ion in combination

with several different counter ions. For the long range ion-ion-interactions they used the Pitzer-Debye-Hückel (PDH) ansatz [35]. With this COSMO-RS-PDHS they achieved remarkably good descriptions of many important electrolyte systems, as shown in figure 5. The advantage of this approach is that it does not require special parameter fitting, once the number of explicit water molecules for an ion is determined. A disadvantage is that the mole scale has to be carefully recalculated, because some water molecules are now introduced in form of the hydrated ions. Furthermore this approach comes to its limit at high ionic strength, when the number of water molecules available for the explicit solvation is no longer abundant.

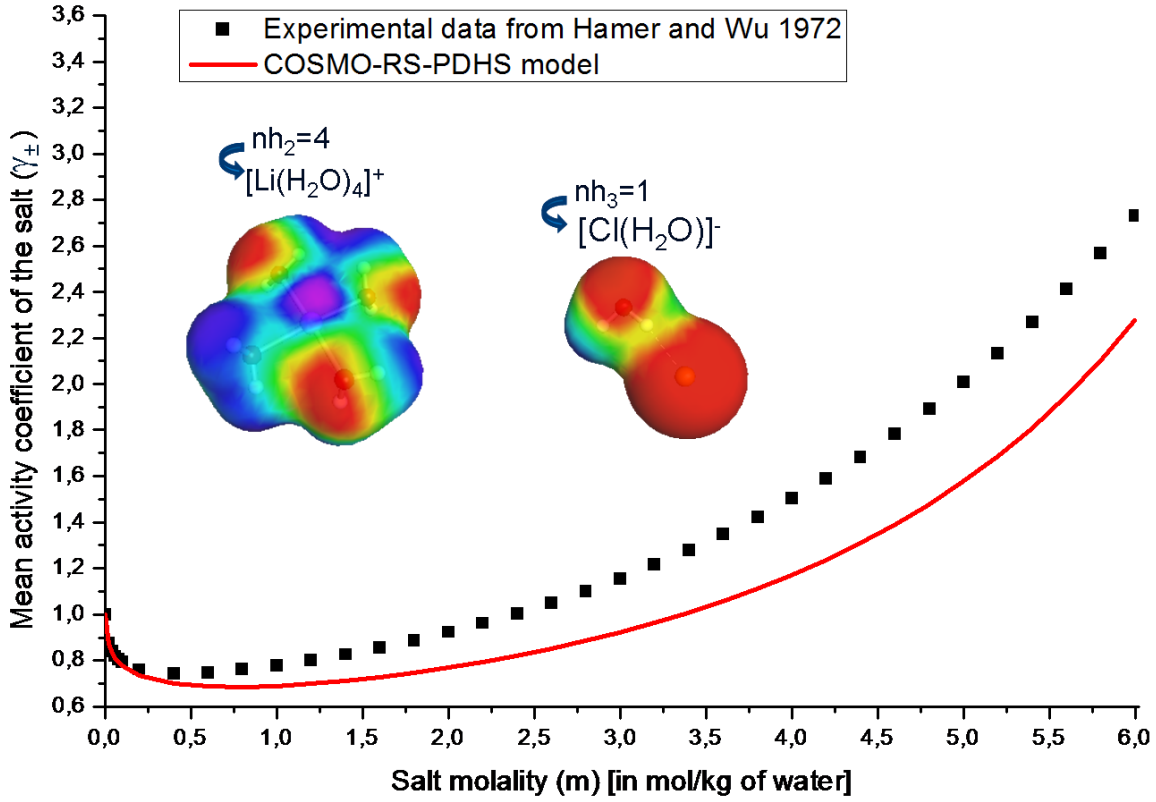

Figure 5: Activity coefficient of LiCl in a binary LiCl-water mixture, with Li^+ being represented as $[Li(H_2O)_4]^+$ and Cl^- as $[Cl(H_2O)]^-$ according to Toure [34]

A very different route for the handling of small ions in electrolytes was taken by the Smirnova and coworkers [36]. For the correction of the missing parts of the ion-solvent interactions they introduced additional interactions of small cations with strongly negative parts of the solvent surface, using the functional form as it was used for hydrogen bonding, i.e. they allowed small cations to be hydrogen bond donors. Fitting element specific hydrogen bond constants and fine-tuning the radii of the small ions, they also achieved good descriptions for electrolyte activity coefficients, using a PDH expression for the long range ion-ion interactions. A practical advantage of

this approach is that no rescaling of the mole fractions is required, and that this approach is applicable up to higher concentrations of ions and to electrolytes in solvent mixtures and ionic micellar systems.

A very important application of COSMO-RS involving ionic species is pK_a prediction. Just calculating the total free energy difference of the acid-base pair, i.e. of the corresponding protonated and deprotonated species, within the COSMO-RS scheme yields a surprisingly good correlation of experimental pK_a values [37] of very different acids, all falling on one regression line. Unfortunately, the slope of this regression line is only about 60% of the theoretically expectable value. This annoying behavior, which also appears in several other solvation models, has been analyzed in detail, initially without an explanation [37]. Later it was shown that the addition of explicit water molecules, algorithmically placed at the most polar surface patches of the neutral and ionized species, raised the slope to 90% with one explicit water molecule, and to 102% of the expected slope employing two explicit water molecules on both sides [38]. It appears that the explicit water molecules are required for the ions in order to correct for the strong ion problem of COSMO-RS. Nevertheless, the explicit water molecules add some noise to the free energies, and therefore the direct prediction of pK_a using the empirical regression line with the lower slope yields more accurate predictions with an accuracy of ~0.6 pK_a-units (RMSD).

An important property to predict in aqueous systems is the influence of salt on the activity coefficients and solubilities of neutral compounds, often called salting-out effect. This effect is usually quantified by the Setschenow coefficient. In two independent studies [39,40] it was demonstrated that COSMO-RS, without any ion corrections, predicts the trends of Setchenow coefficients qualitatively correct, but overestimates it, so that the slope of experimental Setchenow coefficients versus predicted ones is roughly 0.6. Whether this slope inconsistency can be removed by explicit water molecules, as it was in the case of pK_a, has not yet been investigated systematically.

Another interesting study was done on the partitioning of a non-trivial series of anion, ranging from the simple, spherical Cl^- to the triply charged, flexible citric acid anion, between water and a the organic solvent 4-(3-phenylpropyl)pyridine (PPP)[41]. The free energies of transfer were measured by ion-transfer voltammetry. A straight prediction of the free energies of transfer with COSMO-RS yielded a good correlation with the experimental data, but with a slope of 0.5, i.e. COSMO-RS over-predicted the free energies of transfer by a factor 2. In a second step the amount of water dissolved in the organic phase was calculated with COSMO-RS. A water mass fraction of 8% was predicted. Measurement in the lab yielded 10%. Taking into account the 10% of water in PPP in the COSMO-RS predictions, the predicted free energies of transfer reduced strongly and the

slope of the regression line became almost exactly unity. This proves that for the solvation of ions in organic solvents it is very important to take into account even small contents of water.

Summarizing, COSMO-RS theory can be applied to several questions involving ions in water, but in some cases additional corrections as an external Debye-Hückel term and explicit water molecules or corrected ion-water interactions may be required.

6. Aqueous interfaces and micelles

Originally COSMO-RS theory considers molecules embedded in a homogeneous solvent. In this case the chemical potential of the solute is calculated by integration of the σ-potential of the solvent S over the entire surface of the solute X. A FLATSURF extension has been developed, which enables the application of COSMO-RS to molecules at simplified, flat interfaces of two solvents, and even to liquid-vacuum interfaces. It samples the center positions and orientations of a solute at an interface, calculates the free energy of the solute in each realization by applying the σ-potential of solvent S to the surface located in S, and the σ-potential of solvent S' to the surface parts in S', and finally derives the free energy of the solute at the interface from the partition sum of all interface exposed states of the solute. This method can be used to quantify the free energy of transfer of a solute from a bulk solvent to an interface. Goss used the FLATSURF method for the quantification of the adsorption of organic molecules to air-water interfaces [42]. He proved that the experimentally determined adsorption constants of organic molecules to aqueous aerosols or very small water droplets, as they exist in fog, can be well described using the COSMO-RS/FLATSURF predicted free energies of transfer of the molecules from the water-air interface to the gas phase, while the use of air-water Henry's law constants, i.e. free energies of bulk water hydration, yields much worse correlation of the experimental data. In this way he proved that for such small droplets surface absorption is the dominating absorption mode in fog.

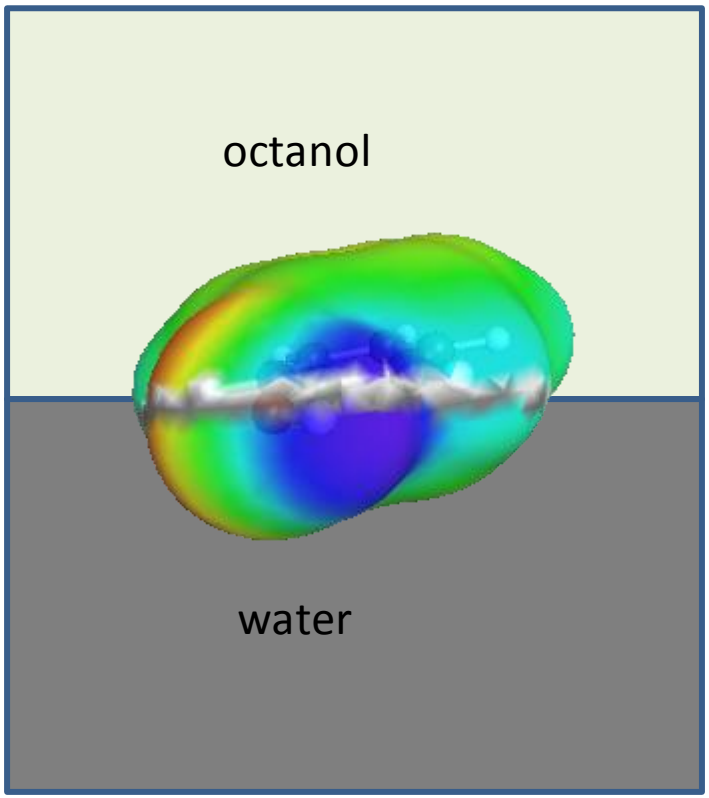

Figure 6: preferential arrangement of phenol at the interface of water and octanol as predicted by COSMO-RS/FLATSURF

Andersson et al. recently extended the FLATSURF method to the prediction of interfacial tensions [43], mostly for pure or mixed aqueous-organic interfaces. They introduced an additional interfacial layer, which allows for an accumulation of amphiphilic species at the interface. Without any empirically adjusted parameter they achieved quantitative agreement of the predicted interfacial tensions with experimental data.

Another situation of inhomogeneous aqueous solvation is the partitioning of solutes between micelles and water. Considering a micelle as a layered liquid with typically up to 40 layers and following the same concept as described for the FLATSURF extension, the COSMOmic extension of COSMO-RS is able to predict water micelle partition coefficients [44]. Most experimental data for micellar partition coefficients are measured for the bio-membrane-water systems, because these are of utmost importance in life science. In the first COSMOmic paper a very good performance of the COSMOmic approach for the prediction of DMPC-water partition coefficients was reported for neutral solutes, without the need for any additional parameter fitting. Recently this has been extended to ionic solutes, adding a simple empirical expression for the zeta-potential of the bio-membrane [45]. Equally good results for anions, neutral solutes and cations have been achieved with an accuracy of ~0.7 log-units (RMSE), which can be considered as being close to the experimental accuracy of these data. Beyond the bare micelle-water partition coefficients, COSMOmic also generates detailed information about the free energy profiles for solutes penetrating through micelles and bio-membranes. By two independent groups COSMOmic has recently been compared with molecular dynamics (MD) calculations for the same systems [46, 47]. Both studies report that COSMOmic produces at least comparably accurate information at a computational expense which is less than 0.01% of the computation time required for MD. Storm et. al. also reported applications to other ionic and non-ionic micelles [48], i.e. SDS and CTAB.

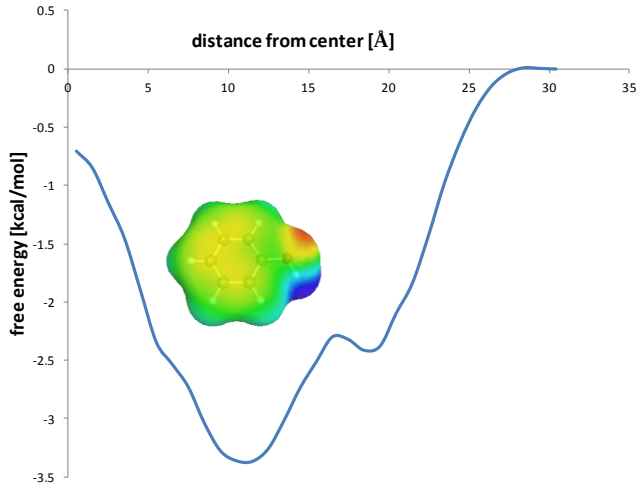

Figure 8: Free energy profile of phenol in an aqueous DMPC membrane calculated with COSMOmic.

7. Summary and outlook

By reducing the interactions of molecules in liquid systems to surface interactions quantified based on surface polarization charges resulting from quantum chemistry, the COSMO-RS method provides an efficient alternative approach to the simulation of pure and mixed liquid systems. Aqueous systems have been a special focus during the development of COSMO-RS and therefore the very special and complicated solvent water was described well from the beginning. Quantitative predictions of solvation free energies and free energies of transfer for neutral solutes involving aqueous phases can be expected to be within the same overall intrinsic accuracy of ~0.3-0.4 kcal/mol (RMSE) as it is achieved for non-aqueous liquids. Within the COSMO-RS approach the specialty of water shows up when water is considered as a solute, and empirical water corrections are required in this case.

Promising extensions of COSMO-RS to ionic species have been developed by various groups, enabling the prediction of dissociation constant in aqueous and non-aqueous solvents, free energies of transfer of ions, and recently also the simulation of electrolyte solutions. Nevertheless, caution is still recommended when using COSMO-RS for electrolyte systems and further work is still required for the development of a robust and easy to use treatment of ions within the COSMO-RS framework.

Much progress has been achieved in the extension of COSMO-RS towards inhomogeneous systems mostly involving one aqueous phase, such as molecules in micelles or at phase boundaries, or the prediction of interfacial tensions.

Nevertheless, some fundamental limitations of COSMO-RS still remain, especially its restriction to dense, incompressible liquids, which prohibits its application to near and super-critical systems in general, and to super-critial water in special. While combinations of COSMO-RS with equations of state may help in this direction, they always require the fitting of special parameters, because water turns from a bulk liquid of closely interacting molecules to an ensemble of more separated molecules at high temperatures and pressures, causing the problems of the description of the single molecule water coming up. A project for the development of a COSMO-RS theory overcoming this limitation is preparation, but it will take several years before a consistent COSMO-RS description of compressible fluids may become available.

Declaration of potential conflict of interest

The author is the original inventor and developer of COSMO and COSMO-RS and distributes the method commercially in form of the software COSMOtherm out of his company COSMOlogic GmbH&CoKG.

References

[1] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans.2* (1993) 799

[2] J. Tomasi, B. Mennucci, R. Cammi, *Chemical Reviews* 105 (2005) 2999

[3] A. Klamt, *J. Phys. Chem.* 99 (1995) 2224

[4] A. Klamt, V. Jonas, T. Buerger, J.C.W. Lohrenz, *J. Phys. Chem. A* 102 (1998) 5074

[5] A. Klamt, F. Eckert, W. Arlt, Annual Review Chem. Biomol. Eng. 1 (2010) 101-122

[6] A. Klamt, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier, Amsterdam, (2005)

[7] S.T. Lin, S.I. Sandler, *Ind. Eng. Chem. Res*. 41 (2002) 899

[8] D. Kundu, T. Banerjee, . *Eng. Chem. Res*. 50 (2011) 14090

[9] H. Grensemann, J. Gmehling, *Ind. Eng. Chem. Res*. 44 (2005) 1610

[10] C.C. Pye, T. Ziegler, E. van Lenthe, J.N. Louwen, *Can. J. Chem.* 87 (2009) 790

[11] A. D. Becke, *Phys. Rev. A* 38 (1988) 3098

[12] J. P. Perdew, *Phys. Rev. B* 33 (1986) 8822

[13] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Letters* 242 (1995) 652

[14] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.* 97 (1997) 119

[15] R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Letters* 162 (1989) 165

[16] Turbomole, Version 6.20, TURBOMOLE GmbH, Germany, 2010.

[17] A. Schäfer, A. Klamt, D. Sattel, J.C.W. Lohrenz, F. Eckert, *Phys. Chem. Chem. Phys.* 2 (2000) 2187

[18] A. Klamt, G. J. P. Krooshof, R. Taylor, AICHE Journal 48 (2002) 2332

[19] E.A. Guggenheim, Mixtures, Oxford University Press, Oxford, 1952

[20] A. Klamt, *Fluid Phase Equilibria* 206 (2003) 223

[21] A. Klamt, J. Reinisch, F. Eckert, A. Hellweg, M. Diedenhofen, *Phys. Chem. Chem. Phys.*14 (2012) 955

[22] A. Klamt, J. Reinisch, F. Eckert, J. Graton, J.Y. Le Questel, *Phys. Chem. Chem. Phys.*15 (2013) 7147

[23] A.V. Marenich, R.M. Olson, C.P. Kelly, C.J. Cramer, D.G. Truhlar, *J. Chem. Theory Comput*. 3 (2007) 2011

[24] A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco F. J. Luque, *Acc. Chem. Res*. 42 (2009) 489-492

[25] A Klamt, M Diedenhofen, *J. Phys. Chem. A* (2015) DOI: 10.1021/jp511158y

[26] R. Putnam, R. Taylor, A. Klamt, F. Eckert, M. Schiller, *Ind. Eng. Chem. Res.* 42 (2003) 3635

[27] A. Fredenslund, R.L. Jones, J.M. Prausnitz, AIChE J. 21 (1975) 1086

[28] F. Eckert, A. Klamt, L. Pohler, COSMO*therm* program, Release C3.0, Revision 1822, COSMO*logic* GmbH&CoKG, Leverkusen, Germany (2014)

[29] A.A. Lapkin, M. Peters, L. Greiner, S. Chemat, K. Leonhard, M.A. Liauw, W. Leitner, *Green Chem.*12(2010) 241

[30] I. Kikic, M. Fermeglia and P. Rasmussen, *Chem. Eng. Sci*. 46 (1991) 2775

[31] M.P. Andersson, S.L.S. Stipp. *J. Comput. Chem*. 35(2014) 2070

[32] O. Toure, F. Audonnet, A. Lebert, C.G. Dussap, *Chem. Eng. Res. Des*. 92 (2014) 2873

[33] O. Toure, F. Audonnet, A. Lebert, C.G. Dussap, *Can. J. Chem. Eng*., 93 (2015) 443

[34] O. Toure, PhD thesis, Université Blaise Pascal, Clermont-Ferrand 2014

[35] K.S. Pitzer, *J. Am. Chem. Soc.*, 102 (1980) 2902

[36] T. Ingram, T. Gerlach, T. Mehling, I. Smirnova, *Fluid Phase Equilibria* 314 (2012) 29

[37] A. Klamt, F. Eckert, M. Diedenhofen, M.E. Beck, *J. Phys. Chem. A.* 107 (2003) 9380

[38] F. Eckert , M. Diedenhofen , A. Klamt., *Molecular Physics* 108 (2010) 229

[39] S. Oleszek-Kudlak, M. Grabda, E.Shibata, F. Eckert, T. Nakamura, 24 (2005) 1368

[40] S. Endo, A. Pfennigsdorff, and K.-U. Goss, *Environ. Sci. Technol.* 46 (2012) 1496

[41] S.M. MacDonald, M. Opallo, A. Klamt, F. Eckert, F. Marken, *Phys. Chem. Chem. Phys.* 10(2008) 3925

[42] K.-U. Goss, *J. Phys. Chem. A* 113 (2009) 12256

[43] M.P. Andersson , M.V. Bennetzen, A. Klamt, S.L.S Stipp, *J. Chem. Theory Comput.* 10 (2014) 3401

[44] A. Klamt, U. Huniar, S. Spycher, und J. Keldenich, *J. Phys. Chem. B* 112 (2008) 12148

[45] K. Bittermann, S. Spycher, S. Endo, L. Pohler, U. Huniar, K.-U. Goss, A. Klamt, *J. Phys. Chem. B* 118 (2014) 14833

[46] S. Jakobtorweihen, A. Chaides Zuniga, T. Ingram, T. Gerlach, F. J. Keil, I. Smirnova, *J. Chem. Phys.* 141 (2014) 045102

[47] M. Paloncýová, R.DeVane, B. Murch, K. Berka, M. Otyepka, *J. Phys. Chem. B* 118 (2014) 1030

[48] S. Storm, S. Jakobtorweihen, I. Smirnova, A.Z. Panagiotopoulos, *Langmuir* 29 (2013) 11582