COSMO-RS: 
The Beauty of Fluid Phase Thermodynamics from the Perspective of Quantum Chemistry

On-line Number: 0057

Andreas Klamt

COSMOlogic GmbH&CoKG, Burscheider Str. 515, 51381 Leverkusen, Germany, klamt@cosmologic.de

ABSTRACT

COSMO-RS is a relatively novel approach for the prediction of fluid phase thermodynamic properties. It is based on quantum chemical calculations for the individual molecules, followed by a statistical thermodynamics of interacting surfaces, which uses the quantum chemical information to quantify the molecular surface interactions. During the past few years COSMO-RS has been widely accepted as the most innovative and powerful tool in this area of thermodynamics property predictions. This has been underlined by winning the VLE part of the “First Industrial Fluids Simulation Challenge” in November 2002.

Due to the almost general applicability of modern quantum chemical methods, COSMO-RS can be applied in many situations, where group contributions fail. New or complicated multifunctional chemical compounds, reactive intermediates, and even ions and zwitterions can be handled with no problems. Electronic and hydrogen bonding intramolecular interactions are well described. As a result, isomer differences in thermodynamic data can be resolved. Due to the accurate statistical thermodynamics, COSMO-RS is equally reliable at finite concentration and infinite dilution, even for strongly interacting and associating compounds. Although some problems still exist for electrolytes, COSMO-RS appears to be a rather general predictive method for the thermodynamics of simple and complex fluid systems. In addition to some standard applications, successful applications to ionic liquids, electrolyte systems, and reactive systems are reported. Applications in the pharmaceutical industry are described as well.

Beyond the numerical results, COSMO-RS also provides a novel perspective to the interactions of molecules in the liquid phase and to fluid phase thermodynamics. This will be illustrated based on the COSMO surface charge densities in several cases, providing a colourful and vivid insight into fluid phase thermodynamics.

KEYWORDS

fluid phase thermodynamics, a priori prediction, molecular interactions
INTRODUCTION

Reliable thermophysical data are the fundament of chemical engineering process design and development, because feasibility, efficiency, safety, and costs of a chemical process can only be estimated based on a quantitative process model. Modern process simulation software nowadays allows for the simulation of large and complex processes and even of entire plants, but these software tools remain empty shells unless the large number of physicochemical equilibrium properties can be quantified. Although in most cases experimental measurement is the most reliable source of thermophysical data, there are many situations where the large number of data required for a process simulation or for a pre-screening of process alternatives cannot be provided by measurement, since experiments are expensive or time-consuming, or because of safety or stability concerns, or just since the respective compounds are not available in the required purity.

In this situation predictive calculation methods for thermophysical data become more and more important. Group contribution methods (GCMs) like UNIFAC [1,2] presently are most widely used in for the prediction of such data. The basic idea of these methods is to consider chemical compounds as assemblies of predefined functional groups, and to assume that the interactions of molecules can be represented as interactions of these groups. If the large matrix of interaction parameters of the groups is appropriately adjusted to a very large set of experimental data, then the GCM approach allows for the prediction of the thermophysical properties of compounds and mixtures which are composed of the parameterised groups. Despite of the undoubted partial success of this approach, GCMs suffer from several fundamental limitations:

- Many interaction parameters are not available due to missing experimental data.
- Intramolecular interactions cannot be taken into account.
- Differences between most isomers and tautomers of compounds cannot be resolved.
- GCMs are parameterised for simple organic liquids and cannot be extended to more complicated situations like charged species, transition states, molecular complexes and associates, polymers, inorganic compounds, etc.
- GCMs make use of the quasi-chemical approximation of Guggenheim for the statistical thermodynamics which causes considerable errors for strongly interacting molecules in high dilution.

Due to these limitations of GCMs there is a strong need for alternative approaches to the prediction of thermophysical properties.

In order to overcome the limitations of GCMs, alternative methods should avoid the artificial concept of chemical groups and try to consider the full molecules. The most straightforward and best known of such molecular methods are molecular dynamics and Monte Carlo simulations (MD/MC), which try to simulate a representative ensemble of molecules in order to derive from these thermophysical properties as thermodynamic averages. Unfortunately very large ensembles are required for such simulations, which cause the need for an approximate representation of the quantum chemical interactions by classical force-fields. Although good force-fields have been developed for some important regions of organic chemistry, the force-field approach and the corresponding approximations limit the MD/MC methods to almost the same area of standard chemistry in which GCMs give reliable results. In addition, MD/MC simulations are still very time-consuming, because they require the demanding dynamic or statistical simulation for each different point in a phase diagram. Thus, although MD/MC simulations probably are the only way to get access to dynamic properties, it is questionable whether they are the most appropriate supplement or alternative to GCMs in the area of thermodynamic equilibrium data.
By the rapid developments of algorithms and computers, reliable quantum chemical calculations on molecules of up to 50 - 100 atoms can nowadays be performed within the time-scale of a day on cheap personal computer hardware. Thus quantum chemistry (QC), and here especially the density functional theory (DFT), today is - and will be even more in future - an efficient source of information for molecular properties, i.e. total energies, relative energies, electrostatic properties, polarisability, spectra and many more. The big advantage of quantum chemical calculations is that they can be applied to almost any chemical compound, even to rare and exotic chemicals and to meta-stable and transition states - without concerns about compound availability and purification, stability, toxicity, volatility etc. These aspects often make experimental measurements very expensive and time-consuming, if not impossible. Nevertheless, the traditional quantum chemistry is restricted to the calculation of single molecules in vacuum or in small clusters. Hence QC cannot be directly used for the calculation of properties of molecules in liquids or even for fluid phase equilibrium properties. Therefore the use of QC methods for the treatment of the real-life problems of most physical chemists and chemical engineers has been quite small, so far.

In this paper we will illustrate that COSMO-RS (Conductor-like Screening Model for Realistic Solvation) is well suited to bridge the gap between QC and fluid phase thermodynamics. COSMO-RS combines the almost general applicability of QC with an extremely efficient and accurate statistical thermodynamics of interacting surfaces. Thus COSMO-RS has the potential to become a versatile and efficient supplement or alternative to GCMs. Since the details of the COSMO-RS approach and its theoretical justification have been published several times [3-7] this article focuses on the suitability and versatility of COSMO-RS for the prediction of thermophysical data in physical chemistry and chemical engineering. Furthermore, the beauty and vividness of this novel description of molecular interactions will be demonstrated.

**THEORY**

*The Basic Concept of COSMO-RS*

COSMO-RS is a two step procedure. The first step is a special continuum quantum chemical calculation (QC/COSMO) for each chemical species under consideration, and the second step is a fast statistical thermodynamics calculation which takes the information from the QC/COSMO calculations for the quantification of intermolecular interactions and yields the desired thermodynamic properties.

**Step 1: Quantum chemical COSMO calculations**

Quantum chemical calculations are usually performed for molecules in vacuum. But for a condensed state a vacuum environment is not very realistic. The near neighbours of a molecule in a condensed phase will provide some electrostatic polarisability which lowers the energetic costs of molecular polarity. Hence molecules in a condensed state are generally more polar than they would be in vacuum. The overall effect of the surrounding polarisability can be taken into account by the use of continuum solvation models (CSMs) [6,8]. These models integrate the effect of a dielectric continuum virtually surrounding a molecule into the quantum chemical calculation and thus lead to a state which is more realistic for the situation of the molecule in a condensed state. The conductor-like screening model (COSMO) [3,6] is one of these CSMs. By a small physical approximation COSMO achieves a considerable algorithmic simplification and an increased numerical stability compared to normal dielectric CSMs. Therefore COSMO meanwhile is widely used for dielectric continuum QC calculations.
Although the concept of dielectric CSMs has been quite successful for the description of molecules in water and in alkanes, they do not provide a physically correct and detailed description of the solvent, which is desired for a quantitative model. Therefore in the context of COSMO-RS continuum calculations are only used for the generation of a reference state which is suitable for molecules in solution. COSMO-RS considers the state of a molecule embedded in a virtual conductor, which is identical to a dielectric of \( \varepsilon = \infty \), as the reference state for molecules in condensed phases. Thus the first step in the COSMO-RS procedure is to perform a QC/COSMO calculation for each molecular species involved in the considered problem, i.e. for solvent and solute compounds. As a result of such calculation we yield the total energy of each molecule in a conductor and we get the polarisation charge distribution \( \sigma \) which the conductor puts on the surface of the molecule in order to compensate the electric field of the solute. This polarisation charge density \( \sigma \) turns out to be an excellent local measure of molecular polarity, and it is the key entity for the quantification of molecular interactions in solution. It can easily be visualized on the molecular surface, providing a detailed, vivid and colourful description of the molecular surface polarity as shown in figure 1. Details like the stronger polarity of the carbonyl oxygen in acetic acid compared with formyl fluoride, the lone-pair directions on the carbonyl atoms, or the extremely strong lone-pair of the amine nitrogen can be easily recognized. Beyond such 3D-visualization of the surface polarization charge density \( \sigma \), histograms of \( \sigma \) with respect to the amount of surface area provide a valuable quantitative description of the polarity distribution. Such \( \sigma \)-profiles of solutes \( p^X(\sigma) \) and of solvents and solvent mixtures \( p^S(\sigma) \) play a key role in the COSMO-RS theory (see examples in figures 2 - 4).

![Figure 1: COSMO-surfaces of acetic acid, formyl fluoride, and methylamine coloured by the polarization charge density \( \sigma \). The details of the molecular surface polarity can clearly be recognized.](image-url)

Based on theoretical estimates and on a detailed optimisation, a set of element specific radii (about 1.17 times Bondi radii) has been derived [5] which is used for the construction of the molecular cavities in the conductor. DFT has turned out to be a good level of QC in order to achieve a reliable energies and polarisation charges of ground state molecules at reasonable computation times. DFT calculations for molecules of up to about 30 atoms can nowadays be performed in less than a day on personal computers. If efficiently implemented, COSMO causes less than 20% overhead compared to vacuum DFT calculations, so that roughly the same time-requirements hold for the DFT/COSMO calculations. Good and approved DFT/COSMO implementations are presently available in the programs DMol3, Turbomole, Gaussian03, and PQS [9-13]. Others are in preparation.

It is important to note that the QC/COSMO calculations need to be done only once for each molecular species. The results are stored in a COSMO-files which can be stored in a database for later use in COSMO-RS calculations. A database of about 2500 common solvents and important chemicals is
available [14]. Hence a new thermophysical problem only those compounds need to be calculated by QC/COSMO which have not been calculated before. This database also enables efficient solvent or entrainer screening.

**Step 2: COSMO-RS thermodynamics**

While the QC/COSMO calculations generate the necessary input information about the molecules, the second step provides the thermodynamic calculations. As done in GCMs the interactions of the molecules are now considered as pair-wise contact interactions of molecular surfaces. In GCMs the only information about each piece of molecular surface is the group index $k$ and the interactions of the surfaces are quantified by the large, empirically fitted group interaction matrix $a_{kl}$. This is replaced by a generic interaction functional $e(d, d')$, where $d$ and $d'$ are sets of local properties of the two contacting surface pieces which are known from the QC/COSMO calculations. The most important of these properties is the polarisation charge density $\sigma$. Based on physical arguments, electrostatic and hydrogen bond interactions can be quite well expressed as functions of the polarisation charge densities $\sigma$ and $\sigma'$ [4-7], i.e. as $e(\sigma, \sigma')$.

In analogy to GCMs, the statistical thermodynamics of COSMO-RS is split into a combinatorial part and a residual part. The combinatorial contribution, which takes into account size and shape effects of the molecules, is small in most cases and it can be treated in COSMO-RS almost as in GCMs.

Based on the interaction functional $e(\sigma,\sigma')$, COSMO-RS solves the thermodynamics for the residual part by an iterative, but exact algorithm for pair-wise interacting surface pieces [4-7,15] which first calculates the activities $\gamma(\sigma)$ of the different types of surface segments in an ensemble $S$ as

$$\gamma(\sigma)^{-1} = \sum_{\sigma'} p_S(\sigma') \gamma(\sigma') \tau(\sigma,\sigma')$$

(1)

using the solvent $\sigma$-profile $p_S(\sigma)$ and the interaction parameter

$$\tau(\sigma,\sigma') = \exp \left\{ - \frac{e(\sigma,\sigma')}{kT} \right\}$$

(2)

These COSMOSPACE equations [15] have been derived from the partition sum of an ensemble of independently pair-wise interacting surface segment. Although being an off-lattice theory, COSMOSPACE extreme well describes the results of lattice Monte-Carlo simulations. Meanwhile it has been shown that they are identical with an exact solution of the Guggenheim’s quasi-chemical theory [16], although the latter is a lattice fluid theory. Thus COSMO-RS uses a very sound thermodynamic model and any approximations to the quasi-chemical model are avoided which are responsible for the inability of most GCMs to generate reliable results for finite and infinite dilution activity coefficients with the same parameters.

From these solutions for the combinatorial and residual parts of the statistical thermodynamics activity coefficients, excess properties, and many other thermodynamic properties are easily available. In contrast to GCMs, even vapour pressures of the compounds can be calculated without knowledge of pure compound properties.

Despite of the rather physical model, a small number of constants in the interaction functional as well in the thermodynamic equations had to be fine-tuned based on experimental data. Depending on the details of the COSMO-RS version this number is about 25, most of them being close to the initial physical guess. This number may be compared with the about 5000 freely adjusted parameters in a GCM like mod-
UNIFAC. It is important to note that after these numbers have been fixed once, COSMO-RS can be applied as a purely predictive tool to almost any chemical compound consisting of the presently about 10 most important elements taken into account in the parameterisation. The COSMO-RS thermodynamics is available as software COSMOtherm [17], which runs on all common PC and workstation systems.

APPLICATIONS

Vivid Application Examples

The ability to visualize the surface polarity $\sigma$ on the molecular surface and to convert it into a $\sigma$-profile, i.e. a histogram of the surface polarity of a molecule enables a much more graphic and vivid understanding of fluid phase thermodynamics than other methods can provide. As one example we may consider the binary mixture of propanone (acetone) and chloroform, which is well known for its large negative heat of mixing. Looking to the surface polarities and $\sigma$-profiles shown in figure 2, it can clearly be seen that propanone and chloroform have almost perfectly complementary polarities which allow for a very favourable pairing of surface segments with opposite polarities in the mixture, while each of the solvents itself does not allow for a good pairing due to the asymmetry of the individual $\sigma$-profiles. These leads to a strong reduction of the electrostatic misfit energy in the mixture and hence to a large negative heat of mixing.

![Figure 2: Surface polarities, $\sigma$-profiles, and excess properties of the binary mixture of propanone and chloroform as calculated by COSMOtherm](image)

Figure 2: Surface polarities, $\sigma$-profiles, and excess properties of the binary mixture of propanone and chloroform as calculated by COSMOtherm
In contrast, the mixture of propanone and dimethyl ether shown in Figure 3 behaves much more like an ideal mixture, because the \( \sigma \)-profiles of the two solvents are very similar. Here the old recipe of "similis similibus solvantur" applies.

As a third example the mixture of water and hexane shall be mentioned. As shown in Figure 4, the \( \sigma \)-profile of water mainly consists of two pronounced peaks at \( \pm 1.5 \) e/\( \text{nm}^2 \), arising from the two polar hydrogens and from the lone-pairs of the oxygen, respectively, while there is very few non- or weakly polar surface area on the water molecule. This well balanced strong polarity makes water to feel very well in itself and hence to have a low vapour pressure. Hexane in contrast only has weakly polar surface in the region of \( \pm 0.5 \) e/\( \text{nm}^2 \). Although there is no strong interaction, these almost neutral surface segments can also very well pair up with each other. But if we put a hexane molecule into water at moderate temperature, there is not sufficient thermal energy available to break the very strong polar surface pairs of water. Thus hexane has to select the very few moderately polar surface segments of water as partners. Resulting a loss of entropy, but almost no excess enthalpy of mixing. In this way COSMO-RS also correctly describes the unusual feature that the low miscibility of hydrocarbons in water is mainly an entropic effect, and not enthalpic, as one would intuitively assume. This has been shown in more detail in a recent paper [18].
Figure 4: Surface polarities, $\sigma$-profiles, and excess properties of the binary mixture of water and hexane as calculated by COSMOtherm

Figure 5: Surface polarities of typical ionic liquid cations and anions

Figure 6: $\sigma$-profiles of ionic liquid cations and anions

As a final example of the robustness and graphicness of COSMO-RS we like to mention the promising solvent class of the ionic liquids. To our surprise, the COSMO-RS is able predict the activity coefficients of solutes in ionic liquids, although it was never specially parameterized for the treatment of charged species. Several applications examples are shown in [24]. The good performance of COSMO-RS for ionic can be understood, if we consider their surface polarities and $\sigma$-profiles as shown in figures 5 and 6. Although carrying a net charge, ionic liquids show smaller surface polarities than other polar solvents, e.g. water. The charge is well distributed over a larger part of the molecule, either by symmetry or by conjugation. Most likely this is also the reason for the low melting points of these room-temperature liquid salts, because polarity hotspots would results in association and subsequent crystallization

Range of Applicability of COSMO-RS

Due to the almost general feasibility of the underlying DFT/COSMO calculations and the generic and physically based description of the interactions, COSMO-RS has a very broad range of applicability. Obviously the core area for COSMO-RS applications is the calculation of VLE and LLE equilibria of binary and multinary liquid mixtures. Several examples of such applications have been published [8,18-21]. COSMO-RS can be applied to any rare kind of chemical functionality without significant loss of accuracy, because DFT as well as the COSMO-RS interaction functional are general in this regard. Intramolecular interactions as electronic push-pull effects or hydrogen bonds are well covered by COSMO-RS. Thus wide areas of industrially important chemistry can be treated with COSMO-RS which are just inaccessible by group contribution methods. The reliability of COSMO-RS in chemical engineering VLE prediction has been further underlined by winning the VLE section of the first industrial fluid property simulation challenge [23]. The capability of COSMO-RS to resolve small differences of compounds and to treat most diverse solvents opens unique opportunities for an efficient solvent screening in the search for a good solvent, absorbent, or entrainer for isomer separation, especially if combined with the COSMObase database of about 1000 pre-calculated solvents. Such solvent screening has become a standard procedure in several chemical companies, and some important successes have been reported.

In its COSMOtherm implementation, COSMO-RS also is able to treat more complex chemical phenomena as association, dissociation, and complex reacting systems including meta-stable reaction
intermediates and transition states [25]. Thus, due to its sound physical basis, it allows for a much more complete description of complex systems than most other chemical engineering model.

Although COSMO-RS was primarily developed for liquid systems, it turned out that in some cases it can be applied to amorphous solids as well. One of these cases is solubility in polymers. With a few additional assumptions and approximations, the basic ideas of COSMO-RS can as well be used for the calculation of adsorption, absorption, and partition coefficients involving phases or matrices of chemically less well defined or unknown compositions. Examples are the prediction of adsorption constants to activated carbon [26], soil-sorption coefficients [27], and physiological partition coefficients as blood-brain-partitioning or intestinal absorption [28].

Conclusions

By the unique combination of quantum theory with a continuum solvation model, and by a subsequent statistical thermodynamics of interacting surfaces based on the quantum chemical surface properties, COSMO-RS provides a novel, vivid, and predictive access to fluid phase thermodynamics. Although presently not being more accurate than modern group contribution methods in their core region of validity, COSMO-RS enables a priori predictions of thermophysical data in wide application areas of chemical engineering where presently no other predictive tools are available.

References