This is a postscript of Fluid Phase Equilibria, 172(1), 43-72. The original article can be found under
http://www.sciencedirect.com/science/article/pii/S0378381200003575

# COSMO-RS: A Novel and Efficient Method for the *a priori* Prediction of Thermophysical Data of Liquids

By Andreas Klamt* and Frank Eckert
COSMO*logic* GmbH&Co.KG, Burscheider Str. 515, 51381 Leverkusen, Germany
* corresponding author, Tel.: +49-2171-731681, e-mail: andreas.klamt@cosmologic.de

keywords: molecular simulation, COSMO-RS, vapor-liquid equilibria, liquid-liquid equilibria, activity coefficient, chemical potential

## Abstract

A novel and very efficient method for the a priory prediction of thermophysical data of liquids is presented. It is based on unimolecular quantum chemical calculations which provide the necessary information for the evaluation of molecular interactions in liquids. Combined with a very fast and accurate statistical thermodynamics the new method is an alternative to structure-interpolating group contribution methods. The most important advantages are the essentially general applicability, the sound physical basis, and the graphicness of the procedure, which easily allows for a chemical interpretation and understanding of thermophysical behavior. A methodological comparison with group contribution methods is given. Example applications are presented.

## 1. Introduction

Structure-interpolating group contribution methods (GCMs) presently are the most reliable and most widely accepted way of predicting activity coefficients and other thermophysical data of compounds in liquid multi-component mixtures without explicit use of experimental mixture data. Such activity predictions often are of high value for chemical engineers in the synthesis and design of chemical processes and plants, since they often face the situation that no experimental data for certain compounds in their mixtures are available, neither from databases nor from measurement at affordable costs. UNIFAC [1] is the most widely used and probably the most accurate of such group contribution methods. The theory of UNIFAC has been summarized in an excellent way by Sandler [2]. An overview over the present status of GCMs has been given recently by Gmehling [3].

In this paper a novel and efficient method for the *a priori* prediction of thermophysical data of liquids is presented as an alternative to GCMs, which does not make such an extensive use of experimental data as GCMs but which yields the necessary information about the molecular interactions in solution from quantum chemical calculations on the chemical compounds combined with a rather rigorous and physically meaningful description of the interactions of molecules in solution. This method, which is called 'Conductorlike Screening Model for Real Solvents' (COSMO-RS), was originally published in 1995 [4]. A generalisation and careful parametrisation of the method was given in 1998 [5]. In addition, there is a general review of the basic model COSMO and its extension COSMO-RS [6]. According to the scientific roots of the method, all of these papers were written from the point of view of computational chem-istry, using the terminology and assumig the background of that special scientific community.

Only recently it was shown, e.g. by the thesis of Clausen [7] that COSMO-RS is a valuable tool for the handling of chemical engineering problems regarding activity coefficients and other thermophysical data of compounds in the fluid phase. Therefore the scope of this paper is to present the COSMO-RS method in a way more appropriate and more familiar for chemical engineers, making use of analogies to chemical engineering methods like UNIFAC and chosing application examples out of chemical engineering practice. In order to achive conceptual clarity while keeping this article comprehensive, a slightly simplified description of COSMO-RS is presented here, and a few technical details are neglected. The reader interested in these details is refered to references [4 - 6]. In forthcoming papers some of these details will be considered in the context of conceptional improvements and extensions of COSMO-RS.

The rest of the article will be organized as follows:

- Section 2 considers interactions of molecules in condensed media. In section 2.1 the underlying physical situation is described. Section 2.2 considers the approximations to molecular interactions made in GCMs, while section 2.3 presents the COSMO-RS view of molecular interactions.

- In section 3 the statistical thermodynamics is considered. Subsection 3.1 describes the exact solution of the statistical thermodynamics problem of interacting surface segments, as it is used in COSMO-RS. Subsection 3.2 considers the approximations which are used in standard chemical engineering methods UNIFAC and UNIQUAC and show, that UNIFAC could be improved by using the exact statistical thermodynamics of COSMO-RS.

- Section 4 summarizes the results of the parametrisation of COSMO-RS based on density functional calculations. In 4.1 the optimized parameters are given and discussed, while in 4.2 the overall accuracy of the method is illustrated based on 6 thermodynamic properties considered in the parametrisation.

- In section 5 six representative solvents are discussed in more detail in order to illustrate the basic meaning of $\sigma$-profiles and $\sigma$-potentials and to show the wealth of information contained in this COSMO-RS view of molecules in solution.

- Section 6 presents several applications of COSMO-RS to binary mixtures of common solvent systems.

- Section 7 compares the special strengths and weaknesses of COSMO-RS and GCMs and gives a list of typical problems for which COSMO-RS right-now can provide solutions while other methods are not applicable.

- Section 8 gives a summary of COSMO-RS and an outlook towards future improvements.

## *2. Molecular Interactions in the Condensed Phase*

### 2.1 The basic physical situation

In contrast to the gas-phase which may be considered as an ensemble of entirely separated and almost non-interacting molecules which undergo rare events of collision, condensed phases are characterized by rather closely packed molecules which are strongly interacting among each other. The interactions between the molecules may be classified in three classes, which shall be considered and characterized seperately, below.

### 2.1.1 Dispersive or van der Waals interactions

The most generally occuring class of intermolecular interactions in the condensed phase are the dispersive (or van der Waals) interactions. Dispersive interactions are short-range forces of molecular surfaces. They are the sum of two parts: An attractive part arising from the intermolecular correlation of the electrons in the vicinity of the adjacent molecular surfaces. This part is generally assumed to scale with $r^{-6}$. The second part is a repulsive term arising from the quantum chemical Pauli repulsion, which tends to restrict the interpenetration of the occupied molecular orbitals of different molecules. This is often considered to scale with $r^{-12}$ (Lennard-Jones potential), but other strongly decaying, empirical potentials may be used as well.

The net behaviour of dispersive interactions is a weak attraction of two adjacent molecules which causes them to approach each other until Pauli-repulsion rather abruptly stops further approximation. Due to the short range character of dispersive interactions the resulting equilibrium distance is characteristic for the two atoms in the adjacent molecules, which are closest to the molecular contact point. These two atoms will keep a distance, called van der Waals distance, which can be considered to be the sum of the two van der Waals (vdW) distances of the adjacent atoms. Only very strong additional interactions or extreme external pressure can cause the molecules to penetrate eachother beyond this vdW distance. Due to the fact that the concept of vdW radii is very useful, but somewhat approximate, many sophisticated vdW parametrisations have been proposed, but the element specific set of vdW radii derived by Bondi [8] still is most generally accepted and will be used as reference below.

Because of the short-ranged nature of the vdW interactions it is useful to consider vdW interactions as interactions of molecular surfaces. In good approximation the amount of vdW energy gained by a unit area of molecular contact surface of two molecues can be considered as

$$E_{vdW}(e,e') = a_{vdW}(e)\, a_{vdW}(e') = \tau(e) + \tau(e') + \delta_{vdW}(e)\delta_{vdW}(e') \qquad (1)$$

with

$$\delta_{vdW}(e) = a_{vdW}(e) - \hat{a}_{vdW} \qquad (2)$$

and

$$\tau(e) = \bar{a}_{vdW}\left(\tfrac{1}{2}\hat{a}_{vdW} + \delta_{vdW}(e)\right) \qquad (3).$$

In these equations the usual product approach for vdW energies has been applied. $\hat{a}_{vdW}$ is some average vdW parameter for all elements while $\delta_{vdW}(e)$ is the specific deviation from the average for element e. Assuming that $\delta_{vdW}(e)$ generally is small, the product on the right side of eq. 1 can be neglected. In this approximation the originally intermolecular vdW interactions reduce to unimolecular interactions of molecules with some average vdW continuum. We will apply this approximation throughout the rest of the article.

In non-polar liquids and crystals dispersive interactions are the only or at least the dominating intermolecular interactions. For medium sized molecules like hexane dispersive interactions are of the order of 10 - 15 kJ/mol. This is a large amount of energy which is responsible for the condensation of larger non-polar molecules. But due to the fact, that the specific part of the vdW energy, i.e. the term $\delta_{vdW}(e)\ \delta_{vdW}(e')$ in eq. 1, typically is about one order of magnitude less vdW interactions are not very specific, i.e. they do not give rise to specific pairing of certain parts of molecular surfaces. Therefore non-polar compounds tend ot have quite low freezing points.

### 2.1.2 Coulomb or electrostatic interactions

Coulomb interactions are the electrostatic interactions between the charged particles (nuclei and electrons) of the entire condensed system. Although sometimes considered separately, we shall subsume induced electrostatic interactions, i.e. electronic polarizability, in this class. In principle, Coulomb interactions are general to all condensed systems. Nevertheless, for some classes of compounds, usually called non-polar compounds, the net intermolecular electrostatic interaction of each molecule with its environment is rather small due to a good intramolecular compensation of nuclei and electrons.

Coulomb interaction energies of neighboring polar molecules typically are in the order of 20 - 30 kJ/mol. Thus, in such systems they are even stronger than vdW interactions. They are rather specific, because for each piece of molecular surface the gain in electrostatic energy strongly depends on the choice of the right partner of opposite polarity.

Due to the relatively slowly decaying $r^{--1}$-behaviour of the Coulomb potential these interactions are generally considered as long-ranging. This makes the explicit treatment of Coulomb interactions in condensed systems computationally very expensive and often this is the bottleneck for a good computational treatment of extended condensed systems.

Nevertheless, at least in systems of neutral molecules the Coulomb interactions can be represented by local dipoles and higher multipoles, and the resulting potentials decay with $r^{-3}$ or stronger. In that representation the vast amount of Coulomb interaction energy arises from nearest neighbor interactions.

### 2.1.3 Hydrogen bonding

Hydrogen bonding (hb) is important in systems containing strongly positively polar hydrogen atoms as donors and negatively polar heteroatomic lonepairs as acceptors. In such systems donors and acceptors may form a very short-ranged and directed pseudo-bond, the hydrogen bond. Hydrogen bonding apparently goes along with a large Coulomb energy gain of the two oppositely polar groups. Sometimes this Coulomb energy is considered to be part of the hb-energy. Within this article only the quantum chemical energy gain arising from the penetration of the groups below vdW distance and from the resulting additional electron delocalisation is considered as hb-energy.

In this sense the energy of a single hydrogen bond typically is in the order of 10 kJ/mol. Hydrogen bonding strongly depends on the right choice of the nearest neighbor contact partner and even on its orientation. Thus it is even more specific than Coulomb interaction.

## 2.2 Intermolecular interactions in Group Contribution Methods

In GCMs as UNIFAC, which will be considered as a representative GCM below, all interactions between molecules are considered as nearest neighbor interactions of pairwise contacting molecular surfaces. Hence there is no way to catch long-range effects beyond the nearest neighbor shell as they may appear in Coulomb interactions. Therefore charged species cannot be correctly described by GCMs.

The only information which is known of the molecular surfaces is the type of the functional groups i and j, to which the two pieces of surface belong. Each functional group has a specific amount of interaction surface $q_i$, which is measured in relative units, where an ethylene group is used as reference. The number z of independent molecular contacts per unit area is called coordination number. In UNIFAC z is set to 10. Hence the area of a single molecular contact is assumed to be 0.1 units.

The basic assumption in UNIFAC is that each contact between two groups i and j goes along with a specific interaction energy $a_{ij}$. This is assumed to subsume all kinds of interactions, i.e., vdW, Coulomb, and hb-interactions. The values of $a_{ij}$ for all pairs of groups have to be fitted to experimental data.

Two important approximations are implicitly introduced by this assumption:

- The identification of chemical groups is independent of the intramolecular environment. This basically disables any possibility to take into account intramolecular electronic effects like push-pull-effects, which may considerably modify the polarity and thus the interactions of a group. Even more serious is the fact that for hb groups like OH there is no way to take into account whether they are freely available or almost inavailable due to an intramolecular hydrogen bond.
- The assumption that any kind of contact between the two groups is associated with the same energy, independent of the the individual locations and directions. Obviously, in reality two groups, e.g. like $H_2O$ and C=O, can form vastly different contacts, some being attractive and others even repulsive. But in UNIFAC all of these contacts are summarized by a single interaction parameter. I.e., the UNIFAC interaction parameters themselves cover a considerable amount of thermal averaging, although formally they are considered as temperature independent. In addition, the spatially non-specific vdWs interactions and the strongly directed hb interactions are treated alike. Therefore the difference in the interaction entropies of such kinds of interactions cannot be correctly described by UNIFAC.

The fact, that there is no information on the physical origin of the interaction of two groups makes any systematic interpretation of UNIFAC calculations impossible. Especially no systematic temperature dependence of the interaction parameters can be derived due to the unknown mixture of interactions. Hence, if temperature dependent parameters are introduced, as it is done in mod-UNIFAC [9,10], the temperature dependency has to be parametrized seperately for each pair of groups.

## 2.3 The COSMO-RS view of molecular interactions

Now the novel view of molecular interactions as it has been introduced by COSMO-RS will be discussed. Although it was developed independently of GCMs, starting from a very different class of solvation models used in computational chemistry, COSMO-RS shows some parallels to UNIFAC. For the purpose of this paper it appears to be useful to take these analogies as starting points.

As in UNIFAC molecular interactions in the condensed phase are essentially considered as pairwise interactions of molecular surface pieces. But the interaction energy of two pieces of molecular surface is not described by pairwise group parameters, but expressed as by a generic functional of local surface descriptors, which can be calculated on the molecular contact surface of molecules using quantum chemical methods. The most important descriptor used in COSMO-RS is the local screening charge density $\sigma$, which would be induced on the molecular surface if the molecule would be embedded in a virtual conductor. This descriptor, which can be calculated at reasonable costs by quantum chemical programs using the continuum solvation model COSMO [11], turns out to be an extremely valuable descriptor for the local polarity of molecular surface. Before we go on with a detailed descripion of the interaction energy functional in section 2.3.2. a short introduction into COSMO will be given in 2.3.1.

## 2.3.1 The Conductorlike Screening Model (COSMO)

Quantum chemical methods originally have been developed for isolated molecules, i.e. for molecules in vacuum or at best in the gas phase. But due to the extreme importance of solvents in bio- and technical chemistry, and due to the practical impossibility of a rigorous representation of a molecular environment by quantum chemical tools, attempts have been made since almost 25 years to combine the quantum chemical description of a molecule with an approximative continuum description of the surrounding solvent. In order to describe the electrostatic behaviour of solvents, dielectric models have been chosen for this purpose, following the early ideas of Born [12]. Reviews on the such continuum solvation models (CSM) have been given by Tomasi and Persico [13] and by Cramer and Truhlar [14]. The Conductorlike Screening Model COSMO introduced by Klamt and Schüürmann in principal is nothing else than a more efficient variant of these CSMs. It avoids the complicated solution of the dielectric boundary conditions by approximating the screening charges of a dielectric medium of permeability $\varepsilon$ with the scaled screening charges of a conductor, where the scaling factor is $f(\varepsilon) = (\varepsilon-1)/(\varepsilon+0.5)$. Thus only the much simpler boundary condition of a conductor appears in the equations. The COSMO approximation is exact in the limit of $\varepsilon = \infty$ and it is within 0.5% accuracy for strong dielectrics like water ($\varepsilon = 80$). Even in the lower dielectric limit of solvents, which typically is $\varepsilon = 2$, COSMO can be shown to coincide with the exact dielectric model within 10% in all relevant cases. Due to its inherent technical advantages the COSMO approach is going to become the standard CSM in quantum chemical codes.

The concept of a COSMO calculation is simple:

- For a given molecular geometry construct a molecular cavity using element specific radii which typically are in the range of 1.2 times Bondi radii.
- Discretize the cavity in $m$ small segments, so that on each segment i a constant screening charge density $\sigma_i$ can be assumed, resulting in a screening charge $q_i = a_i \ \sigma_i$, with $a_i$ being the area of the segment and calulate the Coulomb interaction matrix $\underline{\underline{A}}$ of the segments.
- For each molecular charge distribution, i.e. nuclear charges and electronic density of the solute, calculate the set of ideal screening charges $\underline{q}^* = (q_1^*, \ldots, q_m^*)$ from the conductor condition of vanishing potential

$$\underline{\Phi}^{tot} = \underline{\Phi}^X + \underline{\Phi}(\underline{q}^*) = \underline{\Phi}^X + \underline{\underline{A}}\,\underline{q}^* \qquad (4)$$

  where $\underline{\Phi}^X$ is the electrostatic potential generate by the solute X on the $m$ segments.
- Calculate the screening charges $\underline{q}$ from $\underline{q} = f(\varepsilon)\,\underline{q}^*$.
- Resubstitute the screening charges as an external field into the quantum formalism.

Since matrix $\underline{\underline{A}}$ and its Cholesky factorisation, which is needed for an efficient solution of eq. 4, only have to be calculated once for each geometry, the additional computational costs of COSMO in the quantum chemical algorithms are small. In addition COSMO may even slightly speed-up the quantum chemical self-consistency part. Thus COSMO calculations are usually no more expensive than gas phase quantum chemistry. At the end of the quantum chemical self-consistency and geometry optimisation loops a self-consistent state is reached, i.e. energy, density, and geometry of the solute X are as if embedded in a dielectric of strengths $\varepsilon$. In addition, the screening charge density $\sigma$ supplied by the continuum on each position of the molecular contact surface and the dielectric interaction energy of the solute with the continuum is known now. The dielectric energy is an excellent measure for the polarity of molecules, much better than dipole moment, because it subsumes all ways of electrostatic interactions of a molecule with its surrounding.

If combined with accurate quantum chemistry methods like density functional theory or correlated *ab initio* methods, CSMs have proven to give very reasonable results for Henry law constants for air/water and air/alkane systems [13,14]. Successful application to the prediction of partition coefficients have been reported as well [15,16]. But despite of this remarkable success, CSMs are not capable of a general description of the activity of molecules in solvents, because they are missing any idea of statistical thermodynamics and thus cannot simulate general solvents, different temperatures, and mixtures. Even more, the application of the concept of the linear response theory of dielectrics to the situation of polar solutes in polar solvents is untenable, because the interactions in such systems are much too strong to be treated by with a linear response theory.

## 2.3.2 The COSMO-RS interaction energy concept

The premises of the COSMO-RS interaction energy concept are threefold:

i) CSMs like COSMO give very reasonable results for solvation in water and alkanes.
ii) The dielectric theory used in CSMs is untenable.
iii) Concepts of temperature and mixture are missing in CSMs.

In this situation COSMO-RS uses the self-consistent state of molecules embedded in a virtual conductor ($\varepsilon = \infty$), as it is readily achievable by COSMO calculations, as a new reference state for molecules in solution, without assuming that this state is of any real meaning. There can be no doubt that this state of ideal screening is much closer to the real situation of molecules in solution, at least in polar solvents, than the unually chosen reference state of molecules in vacuum, because it takes account of the electrostatic screening of molecules by their surrounding, and even of the resulting back-polarisation of the solutes. It is important, that we are able to calculate the state of the molecules in vacuum using the same quantum chemistry method, because this will enable us to calculate the energy difference between the gas phase and the condensed phase, which is needed for vaporisation data.

A liquid now is considered to be an ensemble of almost closely packed ideally screened molecules, as schematically shown in Figure 1. Each piece of surface now has a direct contact partner. But in reality there is no conducting medium between them as has been assumed in the reference state. Therefore the energy difference between the real situation of such contact and the ideally screened situation has to be defined as a local electrostatic interaction energy, which results from the contact of the molecules. Considering a contact on a region of molecular surface of area $a_{eff}$ (effective contact area), and considering the two contacting pieces of molecular surface had average ideal screening charge densities $\sigma$ and $\sigma'$ in the conductor, it is possible to calculate this interaction energy as the energy which is necessary to remove the residual screening charge density $\sigma + \sigma'$ from the contact. Apparently, in the special situation of $\sigma = -\sigma'$ there is nothing to remove and hence the interaction energy is zero. Such a contact here is called "ideal electrostatic contact", in which each molecule just screens its partner just like a conductor would do. In the general case there is some misfit of the partners, i.e. where $\sigma + \sigma'$ does not vanish, the energy of removal is given by elementary electrostatic theory as

$$E_{misfit}(\sigma,\sigma') = a_{eff} e_{misfit}(\sigma,\sigma') = a_{eff} \frac{\alpha'}{2} (\sigma + \sigma')^2 \qquad (5)$$

where $e_{misfit}(\sigma,\sigma')$ means the misfit energy density on that contact surface and $\alpha'$ is a general constant which can be calculated approximately, but in the end will be fitted to experimental data as fine-tuning. As shown in refs. 3-5, this misfit term to a great deal subsumes quite well the polarisation response of the molecules to the misfit situation. If there is no correlation between the different misfit charges, as we may hope to be the case for neutral compounds, the total electrostatic energy difference between the ideally screened ensemble and the real ensemble is given by the integral of misfit energy functional $e_{misfit}(\sigma,\sigma')$ over the entire intramolecular contact area. In the case of ions misfit interactions have to be taken into account. Presently this is not possible in COSMO-RS but it will be covered by an extension of COSMO-RS which is in preparation. Thus, presently application of COSMO-RS is not recommended.

As discussed above, hydrogen bonding is to a large degree covered by the electrostatics, but we have to parametrize the extra hb-energy resulting from interpenetration of the atomic electron densities in some reasonable way. This energy should only come into play, if two sufficiently polar pieces of surface of opposite polarity are in contact, and it should be the more important, the more polar both surface pieces are. Taking the screening charge density $\sigma$ as a local measure of polarity, the following function realizes such behaviour:

$$E_{hb}(\sigma,\sigma') = a_{eff} e_{hb}(\sigma,\sigma') = a_{eff}\ c_{hb}\ \min\{0, \min(0, \sigma_{don} + \sigma_{hb}) \max(0, \sigma_{acc} - \sigma_{hb})\} \qquad (6)$$

with $\sigma_{don} = \min(\sigma,\sigma')$ and $\sigma_{acc} = \max(\sigma,\sigma')$. $\sigma_{hb}$ is some threshold for hydrogen bonding and $c_{hb}$ is the strength coefficient. Both parameters are general and have to be adjusted to experimental data. Eq. 6 may look rather complicated, but it does the following: It is zero, unless the more negative of the two screening charge densities is less than the threshold $-\sigma_{hb}$, and unless the more positive exceeds $\sigma_{hb}$. Because positiv molecular regions get negative screening charge, the negative $\sigma$ now is the donor part of the H-bond, and the positive is the acceptor. In this case the hb-energy is proportional to the product of the excess screening charge densities, i.e. to $(\sigma_{don}+\sigma_{hb})$ $(\sigma_{acc}-\sigma_{hb})$, which is a negative, i.e. attractive, energy contribution due to the opposite sign of $\sigma_{don}$ and $\sigma_{acc}$.

As derived in section 2.1.1, it is assumed that the vdW energy contributions can be expressed by element specific parameters $\tau(e)$, which have to be fitted to experimental data. Then the vdW energy gain of a molecule X during the transfer from the gas phase to any solvent is given by the integral of $\gamma(e)$ over the molecular surface, respectively:

$$E^X_{vdW} = \sum_{\alpha \in X} a^X_\alpha \tau(e(\alpha)) \qquad (7)$$

with $a_\alpha^X$ being the molecular surface on atom α, and e(α) indicating the element of atom α. Because this vdW contribution is independent of any neighborhood relations, it is not really an interaction energy, but it may be considered as an energy contribution to the reference state in solution. Thus, the state of a molecule embedded in a vdWs interacting conductor is considered as reference state. Within this approximation vdW interactions do not contribute to the free energies of transfer between different liquid states, i.e. they do not contribute to activities, but only to the liquid-gas transfer, i.e. to vaporisation data.

Summarizing, the interactions of molecular surfaces in COSMO-RS are given by an interaction energy functional $e(\sigma,\sigma') = e_{misfit}(\sigma,\sigma') + e_{hb}(\sigma,\sigma')$ , which does only depend on the polarities, i.e. on the screening charge densities, of the interacting surfaces, plus a vdW contribution which is subsumed in the reference state energy. The generic interaction functional e(σ,σ') has three adjustable parameters α', $\sigma_{hb}$, and $c_{hb}$, while the vdW term has one adjustable parameter per element. In addition to these explicit parameters it should be noted, that the screening charge densities and hence the entire parametrisation depends on the detailed values of the element specific radii which are used in the cavity construction. Although these are roughly fixed by the rule 1.2 * $R_{Bondi}$, they should be considered as parameters, because they have been optimized in the parametrisation. Thus in total so far there are 3 general parameters plus 2 parameters per element. For the 9 elements H, C, N, O, F, S, Cl, Br, and I considered so far we have altogether 25 parameters. This is almost negligiable compared to the several thousands of group interaction parameters needed for the present suite of groups in UNIFAC, which still cover only a subset of the compounds treatable by COSMO-RS.

## *3. Statistical Thermodynamics*

### 3.1 The self-consistent solution in COSMO-RS

Having now available a concept of the interaction energies, the next step is the statistical thermodynamics which is necessary for any consistent model of molecules in solution. The exact solution of the thermodynamical problem would require an extensive sampling over all different arrangements of all molecules of the systems, weighting the contribution of each arrangement by its Boltzmann factor. This direct approach, which is used in the molecular dynamics (MD) and Monte Carlo (MC) type methods, is extremely time consuming and it always requires compromises regarding sampling and regarding the accuracy of the energy evaluations. COSMO-RS follows a different concept. The basic approximation is, that the ensemble of interacting molecules may be replaced by the corresponding ensemble of independent, pairwise interacting surface pieces. This approximation, which means the neglect of any neighborship information of surface pieces on the moleculat surface and hence the loss of sterical information, is made in exactly the same way in the GCMs like UNIFAC. The great advantage of this approximation is the extreme reduction of the complexity of the problem which allows for a fast and almost exact solution.

The screening charge density σ is the only descriptor determinig the interaction energies. Thus, the ensemble of surface pieces characterizing a liquid system S is sufficiently described by the distribution function $p_S(\sigma)$, which describes the amount of surface in the ensemble having a screening charge density between σ and σ+dσ. In the following the notation "σ-profile" is used for this distribution function.

If our liquid system under consideration is a pure liquid X, then the σ-profile $p_S(\sigma)$ of the system is identical with the σ-profile $p^X(\sigma)$ of the pure compound X. Note that subscipt S always refers to the systems (or solvent) while a superscript X refers to a unimolecular property of a compound X. The σ-profile of a single compound can easily be derived from the quantum chemical COSMO output for that molecule, applying some local averaging algorithm in order to take into account, that only screening charge densities averaged over an effective contact area are of physical meaning in COSMO-RS. For details of the averaging see ref. 4. s of representative solvents are shown and discussed in section 4.

In general a system may be a mixture consisting of several compounds $X_i$ with molar concentrations $x^i$. Then the σ-profile of the system is given by the weighted sum of the σ-profiles of the components, i.e.

$$p_S(\sigma) = \sum_{i \in S} x^i p^{X_i}(\sigma) \qquad (8)$$

For the statistical thermodynamics it is convenient to consider a normalized ensemble. Since the integral of $p^X(\sigma)$ over the entire σ-range is the total surface area $A^X$ of a compound X, the normalized σ-profile of the system is defined as

$$p'_S(\sigma) = p_S(\sigma) / A_S = p_S(\sigma) \Big/ \sum_{i \in S} x^i A^{X_i} \qquad (9)$$

Now the assumption is made that there is no free surface in the bulk of the liquid, i.e. that each piece of the molecular surface has a direct contact partner. This assumption corresponds to a state of a liquid far from the critical region. Under this boundary condition the statistical thermodynamics of the ensemble can be solved using the exact equation

$$\mu_S(\sigma) = -RT \ln\left\{\int d\sigma' \, p_S(\sigma') \exp\left(\frac{\mu_S(\sigma') - a_{eff} e(\sigma,\sigma')}{RT}\right)\right\} \qquad (10)$$

In this equation $\mu_S(\sigma)$ is the chemical potential of an average molecular contact area of size $a_{eff}$ in the ensemble S at temperature T. The function $\mu_S(\sigma)$ appears on both sides of the equation. Therefore the equation must be solved by iteration, starting with $\mu_S(\sigma') = 0$ on the rhs. Fortunately, the solution converges rapidly and $\mu_S(\sigma)$ can be iterated up to numerical precision within milliseconds, even on a personal computer. Then he have achieved an exact solution of the exact equation 10. In this equation the chemical potentials on the rhs are fully consistent with those on the lhs. Therefore, according to conventions in physics, this solution is called self-consistent.

Defining an activity factor $\gamma_S(\sigma)$ of a  piece of surface by

$$\ln \gamma_S(\sigma) = \mu_S(\sigma) / RT \qquad (11)$$

eq. 10 may be rewritten as

$$\ln\{\gamma_S(\sigma)\} = -\ln\left\{\int d\sigma' \, p'_S(\sigma')\gamma_S(\sigma') \exp\left(-\frac{a_{eff} e(\sigma,\sigma')}{RT}\right)\right\} \qquad (12)$$

which may look more familiar to chemical engineers. In this equation it becomes apparent that the exactness of the statistical thermodynamics solution of COSMO-RS results from the consistent use of the activity factor of the partner surface segments σ' on the right hand side. A formal derivation of eq. 10 is given in ref. 3.

Now it is straightforward to define the chemical potential of a solute X in the ensemble S by

$$\mu_S^X = \mu_{res\,S}^X + \mu_{comb\,S}^X = \int p^X(\sigma)\mu_S(\sigma) d\sigma + \mu_{comb\,S}^X \qquad (13)$$

where the residual part, i.e. the part resulting from the interactions of the surfaces in the liquid, is given by the surface integral of function $\mu_S(\sigma)$ over the solute surface, which is expressed using the σ-profile of the solute in eq. 13., and a combinatorial contribution, which arises from the different shapes and sizes of the solute and solvent molecules. For the combinatorial part originally the simple expression

$$\mu_{comb\,S}^X = -\lambda RT \ln A_S \qquad (14)$$

is used, where $A_S$ is the average solvent surface area as defined in eq. 9 and λ is a parameter to be adjusted. In the present parametrisation a Staverman-Guggenheim [17]  (SG) term is added to the combinatorial part as it is usually used in GCMs. Therefore eq. 14 was replaced by

$$\mu_{comb\,S}^X = -RT\left\{\lambda \ln A_S + L_1^{SG} + \frac{z}{2}\frac{V^X}{V_0} L_2^{SG}\right\} \qquad (15)$$

with

$$L_1^{SG} = 1 - \frac{A^X}{A_S} + \ln\frac{A^X}{A_S} \qquad (16)$$

and

$$L_2^{SG} = 1 - \frac{V^X}{V_S}\frac{A_S}{A^X} + \ln\frac{V^X}{V_S}\frac{A_S}{A^X} \qquad (17)$$

where $V^X$ is the molecular volume of compound X, $V_S$ is the average molecular volume of he solvent, and $V_0$ is the partial volume of an ethylene unit used for normalisation purposes. For methodological consistency the COSMO areas and volumes were used here instead of the tabulated group increment areas and volumes of the GCMs. The coordination number z then was adjusted in the general parametrisation described in section 4.

Now the activity coefficient of a solute X in a liquid S is given by

$$\gamma_S^X = \exp\left\{\frac{\mu_S^X - \mu_X^X}{RT}\right\} \qquad (18)$$

Thus, now a complete framework is available to calculate activity coefficients of almost arbitrary solutes in almost arbitrary liquids.

Although in chemical engineering partial pressures, i.e. gas phase compositions, routinely are calculated using activity coefficients supplemented by experimental vapor pressures of the pure components, there may be situations in which even a good direct estimate of the partial pressures and vapor pressures is required. In contrast to GCMs COSMO-RS is able to provide a reasonable estimate using the expression

$$\mu_{gas}^{X} = E_{gas}^{X} - E_{COSMO}^{X} - E_{vdW}^{X} + \omega n_{ring}^{X} + \eta_{gas} \qquad (19)$$

for the chemical potential of the solute in the ideal gas. Here $E_{gas}^X$ and $E_{COSMO}^X$ are the quantum chemical total energies of the molecule X in the gas phase and in the conductor, respectively, and $E_{vdW}^X$ is the vdW energy of the solute, which we addressed to the conductor reference state. The adjustable, but universal parameter $\eta_{gas}$ connects the reference states in solution and in the gas phase, which are 1 mol(X)/mol(S) and 1 bar, respectively, in COSMO-RS. The second last term in eq. 15 is a heuristic but highly significant correction for ring compounds with $n_{ring}^X$ being the number of ring atoms in compound X and $\omega$ being an adjustable parameter, which turns out to be about 0.85 kJ/mol.

Because COSMO-RS trivially yields the Gibbs free energy of a thermodynamic system as the sum of the chemical potentials of the components, and since the temperature derivative can be computed by finite differences, COSMO-RS is readily able to calculate enthalpies and entropies in the standard way.

## 3.2 Comparison with UNIFAC and UNIQUAC

In principle the statistical thermodynamical problem in UNIFAC is analogous to that in COSMO-RS. In both cases an ensemble of pairwise interacting surface segments is considered. Only the description of the interaction energies is different. While in COSMO-RS the screening charge density $\sigma$ is the relevant property for interaction energies, in UNIFAC the underlying group number *i* determines the interaction of a surface segment. Consequently, the analogon of the $\sigma$-profile $p'_S(\sigma)$ is the normalized group constitution function

$$\Theta_k = \frac{\sum_i x^{(i)} \nu_k^{(i)} Q_k}{\sum_i \sum_j x^{(i)} \nu_j^{(i)} Q_j} \qquad (20)$$

in which $\nu_k^{(i)}$ is the number of groups of type k in compound *i* and $Q_k$ denotes the relative surface area of group *k*. The integration over $\sigma$ has to be replaced by a summation over group types, and the group interaction parameters $a_{k,l}$ take the place of the interaction energy $a_{eff}\, e(\sigma,\sigma')$. Making these substitutions the analogon to the exact and self-consistent solution of eq. 12 would read

$$\ln\{\gamma_k\} = -\ln\left\{\sum_l \Theta_l \gamma_l \exp\left(-\frac{a_{k,l}}{RT}\right)\right\} \qquad (21)$$

Instead UNIFAC makes use of the approximation

$$\ln\{\gamma_k\} = 1 - \ln\left\{\sum_l \Theta_l \exp\left(-\frac{a_{l,k}}{RT}\right)\right\} - \sum_l \frac{\Theta_l \exp\left(-\frac{a_{k,l}}{RT}\right)}{\sum_m \Theta_m \exp\left(-\frac{a_{m,l}}{RT}\right)} \qquad (22)$$

where the surface specific logarithmic activities $\ln\{\gamma_k\}$ are related to the group activity factors $\Gamma_k$ in UNIFAC by $\ln \Gamma_k = Q_k \ln\{\gamma_k\}$. The molecular activity coefficients are derived by a summation of the logarithmic group activity factors, in perfect analogy to COSMO-RS.

The approximation of eq. 22 implicitly assumes $\gamma_l = 1$ on the rhs of eq. 21, which means that the activity coefficients are no longer self-consistent. This corresponds to a mean field assumption. Thus, the UNIFAC approximation in statistical thermodynamics implies severe errors if the activity of certain group species in the mixture is much higher than 1. This regularly is the case if strongly interacting species, e.g. hydrogen bonding groups, are considered in high or infinite dilution. Hence original UNIFAC is not very well suited for the calculation of infinite dilution activity coefficients and a special parametrisation was worked out for this limit [18]. This is also the reason why a standard parametrisation of UNIFAC does not perform very well for the prediction of octanol/water partititon coefficients, which involve the activity coefficients of solutes in the hydrogen bonding solvents octanol and water. Only by a special parametrisation [19] satisfying results could be achieved.

The mean-field approximation in UNIFAC may have been justified 20 years ago for the reason of being explicit, i.e. for avoiding any recursive calculations. It must be emphasized here, that at least nowadays - having powerful computers available - there is no reason left to use this severe approximation, because the self-consistent solution of the exact statistical thermodynamics expressed in eq. 21 can be done in milliseconds on any personal computer. Thus a self-consistent UNIFAC (SC-UNIFAC) could be reparametrized using eq. 21. However, this is not the scope of the present paper.

## *4. Parametrisation of COSMO-RS*

The first parametrisation of COSMO-RS was done based on the DMol400 density functional theory (DFT) code [20] using the COSMO implementation [21] in this quantum chemical program. Meanwhile additional parametrisations have been worked out for DMol3 [22], which is the latest version of DMol, having slight changes in the basis sets and functionals, as well as for different basis sets in Turbomole [23] -DFT, which recently got a COSMO implementation [24]. Below we use the present standard parametrisation DMol3-SG, which uses the Staverman-Guggenheim term in the combinatorial part.

The basic parametrisation was done using about 650 room temperature data points [25,26] for the six properties air/water-Henry law constant, vapor pressure, and the 4 partition coefficients octanol/water, hexane/water, benzene/water, and diethylether/water for 217 compounds of broad chemical functionality based on the elements H, C, N, O, and Cl. Meanwhile the elements F, S, Br, and I have been added, using additional data sets of about 20 compounds and 60 data per element.

### 4.1 Results for parameters

The radii used for the cavity construction in COSMO are the most basic parameters in COSMO-RS. In the original implementation the values of the radii have been optimized in a tedious procedure which involved about 20.000 quantum chemical COSMO calculations because each change in the radii requires new COSMO calculations. The results are given in Table 1 together with those of the additional elements. These radii have been kept throughout all other parametrisation since there was no indications that these radii are very sensitive to the quantummechanical method.

As can be seen in Table 1 all radii are 17 (±4) % larger than the corresponding Bondi vdW radius, the only exception being H, for which the optimized COSMO radius is 8% larger than the Bondi radius. The relatively low radii for H and O result from the compromise between normal contacts and hydrogen bonding. Taking into account the fact, that vdW radii are only roughly defined and differ by much more than 4% for different sources, this result can be considered rather systematic. It agrees well with the empirical finding of a 20% increase of cavity radii compared to vdW radii in dielectric CSMs. But in COSMO-RS this result is consistent in another sense. An increase of radii of about 17% yields cavity volumes which in average quite well agree with the molecular volumes derived from experimental density. This means that such cavities are in average space-filling, while vdW cavities are only contacting along nearest neighbor distances, leaving open space between the molecules. However, as illustrated in Figure 1, the COSMO-RS view of contacting cavities requires space filling cavities. Thus the independent optimisation of the radii with respect to thermodynamical data leads to consistent values regarding molecular volumes.

The second set of element specific parameters are the dispersion constants $\tau(\alpha)$. These are given for the DMol3 parametrisation in column 4 of Table 1. The detailed values, which were optimized by linear regression, are slightly sensitive to the details of the parametrisation. Nevertheless, qualitatively they agree quite well with the polarizabilities of the elements, which roughly determine the disperive interactions.

The effective contact area for a single independent molecular contact resulted in $a_{eff}$ = 6.15 Å². Considering the area of 45Å² of a water molecule this corresponds to about 7 independent neighbors for a water molecule. This result for this 'coordination number' at least appears to be a reasonable results.

The electrostatic misfit energy coefficient is $\alpha'$ = 6635 kJ/mol/Å². This agrees reasonably well with the estimate of 8300 kJ/mol/Å² which can be roughly derived from electrostatic considerations.

For the hydrogen bonding parameters we have no a prori guess. The values found in the parametrisation are $c_{hb}$ = 33800 kJ/mol/Å² and $\sigma_{hb}$ = 0.084 e/Å².

For the ring correction coefficient $\omega$ we find 0.89 kJ/mol. Using the reference states 1 mol/mol and 1 bar for the fluid phase and for the gas phase, respectively, we find $\eta_{gas}$ = 20.2 kJ/mol.

Finally, the coefficient $\lambda$ in the combinatorial part of the chemical potentials comes out as 0.3. The coordination number z in the Staverman-Guggenheim term was found to be z = 7.2, whereas in GCMs usually z = 10 is assumed.

Because all properties so far are considered at room temperature, the values of the parameters are best values at room temperature, only. Although the dominant part of the temperature dependence of the thermodynamic properties arises from the statistical thermodynamics and hence is automatically taken into account in COSMO-RS, some parameters require an explicit temperature dependence in order to achieve a good agreement with temperature dependent data. Because temperature dependent vapor pressures are most easily available we chose the temperature derivatives of the logarithmic vapor pressure at room temperature for a parametrisation of the temperature dependencies.

Generally the liquid-gas transfer constant $\eta_{gas}$ has to be split into a temperature independent part $\eta^0_{gas}$ plus a temperature dependent term RT $\eta^T_{gas}$. For $\eta^T_{gas}$ the fit yielded almost exactly the value -6.0. Consequently $\eta^0_{gas}$ was 34.8 kJ/mol, yielding the original value of $\eta_{gas}$ at room temperature.

The vdW energies require an explicit temperature dependence as well. In order to keep the number of parameters as small as possible, a general temperature dependence for all elements was assumed, which is expressed by a factor

$$f_{vdw}(T) = 1 - c^T_{vdW} + c^T_{vdW}\,\frac{T}{298.15\,K}. \qquad (23)$$

Optimisation yielded $c^T_{vdw}$ = -0.61. This means that the dispersive energy contributions generally decrease with temperature, which is reasonable since by thermal expansion vdWs energies are considerably weakened. A second source of a negative $c^T_{vdw}$ may be surface proportional entropic contributions, which arise from the partial restriction of vibrational degrees of freedom in the liquid, and which probably are subsumed in the dispersive interactions by the fitting procedure.

Finally the H-bond coefficient $c_{hb}$ requires a temperature dependence, because the formation of directional H-bonds goes along with a considerable decrease of entropy. Hence the H-bond energy should be interpreted as a free energy and we expect that the extra H-bond energy is zero at infinite temperature. Presently we make the simple assumption

$$c_{hb}(T) = c_{hb}\,\max\left(0,\quad 1 - c^T_{hb} + c^T_{hb}\,\frac{298.15\,K}{T}\right). \qquad (24)$$

The coefficient $c^T_{hb}$ was found to be 1.5.

## 4.2 Results for properties

For the 6 properties considered in the parametrisation a rather homogeneous rms deviation of 1.7 kJ/mol for the chemical potential differences was found, which corresponds to 0.3 log units for the partition properties. A more detailed discussion of the results is given in ref. 4. Although this accuracy surely is less than the accuracy equired for many applications it must be taken into account that it was achieved for a broad variety of properties and a broad range of chemical functionality, which covers alkanes, alkenes, alkines, alcohols, ethers, esthers, carbonyls, acids, mono- and multifunctional aromatic compounds, amines, amides, nitrogen aryls, nitriles, nitro-compounds, chlorinated compounds, and diverse others.

The only relevant data which had to be taken out of the data set due to very large errors have been the Henry contant of water in water (being equivalent with the vapor pressure of water) and the Henry coefficients and partition coefficients of amines. In the first case of water the problems arise from the fact that the number of hydrogen bonds of a water molecule in the solvent water is overestimated by approximately one due to the fact that for the loss of steric information COSMO-RS is not capable to take into account the extremly high degree of order which is necessary to build 4 H-bonds per molecule for water in water. In the case of amines COSMO-RS presently is not able to resolve the strong H-bond acceptor capacity of substituted neutral amines since in such cases the lonepair is almost hidden by the substituent. Such rather long-lived H-bonds must be treated as explicit molecular complexes.

For the additional elements F, S, Br, and I, which have not been considered in the original parametrisation, comparable accuracies were achieved, by fitting the two element specific parameters, only. For F and I the rms is worse about 25%, indicating that for these elements the approximation of an average vdW continuum made in the dispersive energy functional is questionable due to the extreme values of the polarizabilities of these elements.

## *5. COSMO-RS representation of common solvents*

In this section the 6 solvents hexane, water, methanol, acetone, benzene, and chloroform will be considered in more detail, in order to make the reader familiar with σ-profiles and σ-potentials and to illustrate the wealth of information contained in the COSMO-RS representation of molecular interactions. The σ-profiles and room-temperature σ-potentials for the six solvents are given in Figures 2 and 3, respectively. While reading these diagrams please be aware, that the molecular polarity is inverted on the σ-scale, because positive polarities like polar hydrogen atoms cause negative screening charge density, while negative polarities cause positive screening charge. All σ-values are in e/Å².

### 5.1 Hexane

Hexane is used as a representative alkane compound. Because alkanes do not have significant electrostatic moments, the σ-profiles of alkanes are rather narrow. Nevertheless, two peaks can be detected, resulting from the hydrogens on the negative side and from the carbons on the positive side. But the two peaks are close enough to be of no relevance for the σ-potential of hexane. The σ-potential, which describes the affinity of the solvent for a molecular surface of polarity σ almost exactly agrees with a parabola centered at $\sigma = 0$. Since there is no hydrogen bonding in hexane, the σ-potential is exclusively determined by electrostatic misfit, which increases quadratically with the polarity of the surface, i.e. with σ. Since such quadratic electrostatic behaviour is equivalent to dielectric behaviour, COSMO-RS just reflects the dielectric electrostatic screening of alkanes.

### 5.2 Water

The σ-profile of water is very broad and, apart from small fluctuations, it is almost symmetric. On the negative side water shows a broad peak at about -0.015 resulting from the two polar hydrogen atoms, while on the positive side the same broad peak at +0.015 is present, resulting from the lonepairs of the oxygen atom while inbetween there is only very little surface resulting from the intermediate regions on the molecule. The symmetry of the σ-profile is of high importance for the properties of water, because symmetry implies at least the principal availability of the ideal partner surface for each piece of surface in the system. Hence water shows a lot of very favorable and almost ideal interaction, regarding electrostatics as well as H-bonding. This is the reason, why water "likes itself" very much and why it has a high boiling point and surface tension.

Regarding the σ-potential of pure water an almost flat region at a level of about 0.2 kJ/mol/Å² in the range $-0.01 < \sigma < 0.01$ is visible. The slight slope is of no relevance for neutral compounds. Within this region water electrostatically almost behaves like a conductor, apart from a constant surface tension of 0.2 kJ/mol/Å². It is important that among almost all solvents water has the highest σ-potential at $\sigma = 0$, i.e. it likes non-polar surface less than all other solvents. In this way COSMO-RS reflects the fact that water is less lipophilic than all other solvents or that lipids and other non-polar compounds are hydrophobic.

In the outer ranges of the σ-potential H-bondig is dominating. The stronger polar a piece of molecular surface is, the stronger H-bonds it can form with water. Both, H-bond acceptors and donors can undergo efficient H-bonding in water if they overcome a threshold of about ±0.01.

### 5.3 Methanol

Comparing the σ-profile of methanol with that of water an acceptor peak of the oxygen is visible which is slightly less broad than in water but occurs at almost the same polarity σ. The donor peak from the polar hydroxy-hydrogen is just half the peak of the two hydrogens in water. In addition there is a double peak resultig from the CH3 group, having about the same width and shape as the peak of hexane, but being shifted to the negative side due to polarisation by the oxygen.

The σ-potential of methanol looks very different from water. In the middle range it might be approximated by a parabola with low curvature, corresponding to a dielectric behaviour with much higher ε than alkanes . But this curvature can be roughly converted to a value of $\varepsilon = 3.5$ which is much less than the macroscopic value of $\varepsilon = 32.6$ at room temperature. This underlines the arguments given in refs. 3 - 5, that the macrosopic dielectric constant is of no meaning for the electrostatic solvent behaviour of polar solvents. Even the H-bond regions look very dissimilar to water. The sigma potential for donors is much is lower than that of water and has a less negative threshold. This reflects the fact that in contrast to water methanol is not saturated with donors because each oxygen may build two H-bonds, but only one donor is available per oxygen. Therefore additional donors are highly desired and they can undergo a H-bond interaction without replacing an existing H-bond donor. The opposite is true for acceptors. Due to the lack of donors methanol does not want more acceptor functionality since this would compete with the existing acceptors for he rare donors. Therefore only very strong acceptors are able to undergo efficient H-bonding in methanol and the σ-potential of methanol is much higher than that of water in this region.

## 5.4 Acetone

The σ-profile of acetone is highly asymmetric. The carbonyl oxygen forms a peak at about 0.011, while on the other side the 6 hydrogens of the two methyl groups are carrying the countercharge on a much larger area, resulting in a σ-peak at about -0.006. The carbons appear to be responsible for a neutral peak at $\sigma = 0$. Due to this asymmetry acetone cannot interact favorably with itself because the oxygen is missing appropriately polar partners and because there is too much of the slightly polar surface on the hydrogen, which causes unfavorable interactions with alike polarities. This results in a low boiling point of acetone.

In the σ-potential the asymmetry is even more obvious. On the negative side, i.e. for donors, acetone is as attractive as methanol resulting from the two unsaturated but less polar lone pairs in carbonyl compared to the single unsaturated acceptor in methanol. But on the positive side acetone is almost as unattractive as the non-polar hexane.

## 5.5 Benzene

Benzene has a quite symmetric σ-profile, showing one peak from the π-face of the ring at 0.0055, and another one resulting from the polarized hydrogens at about -0.0065. Obviously, the σ-profile of benzene is much less broad than that of water and shows no intensity in both H-bonding regions, i.e. beyond ±0.01.

As a consequence, the resulting σ-potential is rather symmetric as well, being very well approximated by a parabola with a curvature corresponding to $\varepsilon = 4$ compared to a macroscopic ε of 2.3. Here we have the opposite effect compared to methanol: The effective microscopic "dielectric constant" is larger than the macroscopic one because the quadrupole of benzene contributes to the electrostatic screening of polar solutes while it cannot interact with homogeneous macroscopic fields.

## 5.6 Chloroform

The σ-profile of this solvent is dominated by a huge peak of slightly positive screening charge density at $\sigma = 0.0025$ which is caused by the chlorine atoms. Two smaller side peaks at $\sigma = -0.0025$ and $\sigma = -0.006$ originate from the transition areas between the chlorines on the surface. While the carbon does not have a significant amount of exposed surface area, the strongly polarized hydrogen causes a peak at $\sigma = -0.0135$. This peak, which has a much smaller area than typical hydroxy group hydrogens due to the large chlorine atoms, which hide surface parts of the hydrogen, is just in the range of typical hydrogen bond donors, however in COSMO-RS it is assumed that hydrogens bound to carbon are not able to build real hydrogen bonds. Therefore no extra H-bond term is applied to such surface pieces. Apparently, due to the asymmetric σ-profile chloroform "does not like itself" very much, resulting in a rather high vapor pressure.

As a result of the asymmetric σ-profile the σ-potential of chloroform is asymmetric as well. For negative screening charge densities, i.e. for positive molecular polarities, it almost acts like a low dielectric of $\varepsilon = 2.5$, while for positive screening changes it acts like a much stronger dielectric of $\varepsilon = 5.3$.

Summarizing this discussion of σ-profiles and σ-potentials it can be concluded that these two COSMO-RS representations of solutes and solvents are able to supply a vivid and rich description of the solutional behaviour of chemical compounds.

## *6. Representative applications to binary mixtures*

In this chapter some applications of COSMO-RS to binary mixtures are presented, which shall illustrate the way COSMO-RS works for such problems. More applications can be found in the thesis of Clausen [7].

### 6.1 Acetone - Chloroform

The first example is the binary mixture of acetone (1) and chloroform (2). As discussed above these pure liquids of both compounds have asymmetric σ-profiles, and that they "do not like themselves very much" for this reason. But the σ-profiles of acetone and chloroform are almost complementary. The very polar hydrogen of chloroform fits very well to the lone pairs of the acetone oxygen, and the slightly negative chlorines (having slightly positive σ) fit perfectly with the methyl groups of acetone. Hence these two very dissimilar liquids are ideal partners for each other and they should mix perfectly even with a negative excess enthalpy of mixing. Although this is against the common rule that similar liquids mix best, this behavior is just what is well known from experiment. In Figures 4a and 4b experimental data is compared with the COSMO-RS results. The activity coefficients are in excellent agreement. The negative excess enthalpy of this mixture is well reproduced by COSMO-RS. It is slightly overestimated but the deviation of 0.4 kJ/mol is small. It must be emphasized that no experimental data are used in the COSMO-RS calculation.

### 6.2 Butanol - Water

The performance of COSMO-RS for alcohol - water systems has first been considered by Clausen [7]. These systems show strong electrostatic and H-bond intractions. Clausen found a surprisingly good agreement of COSMO-RS with experimental data for the entire series of methanol to butanol. Here only the system 1-butanol - water is shown. In figures 5a and 5b activity coefficients and gas-phase compositions are compared with experiment. For the latter experimental vapor pressures of the pure compounds are used, as it is recommended, if these data are available. Apparently COSMO-RS matches the experimental data extremely good. Even the miscibility gap of this system is well described.

### 6.3 Benzene - Ethanol

This system was first considered by Clausen, again. An x-y-diagram for the benzene - ethanol system is shown in figure 6. The agreement is very good. Especially, the aceotrope of the system is well met.

### 6.4 Cyclohexane - Aniline

The system cyclohexane - aniline is taken as an example for large non-ideality. Results are plotted in Figures 7. The agreement with experiment is almost perfect.

### 6.5 1-Butanol - n-Heptane

The system 1-butanol - n-heptane shows a characteristic S-shaped excess entropy (see Figure 8). This behaviour is well reproduced by COSMO-RS. The agreement for $G^{ex}$ and $H^{ex}$ is satisfactory as well.

## *7. Strengths and weaknesses of COSMO-RS compared to GCMs*

COSMO-RS and group contribution methods are two quite different approaches to the prediction of activity coefficients of molecules in the liquid phase. In this chapter a comparison of the strengths and weaknesses of both appraches is given.

The greatest advantage of GCMs like UNIFAC is their extreme degree of elaboration which resulted from the 23 years of development by Gmehling's group and a few other acedemic groups around the world and from the mutual exchange of experience with these models between industry and the developing groups. Especially UNIFAC has been parametrized carefully for many different aspects of application using several ten-thousands of experimental data, resulting in a suite of specialized parametrisations, each being optimal for certain purposes. Therefore it will be hard to beat the accuracy of each kind of UNIFAC in its core range of parametrisation.

Another advantage of GCMs appears to be their extreme speed and their low computational requirements. If all group parameters are available the entire calculation takes only milliseconds on a personal computer. COSMO-RS instead requires time-consuming quantum chemical COSMO calculations for each compound under consideration. But as soon as these are available, e.g. from a database, COSMO-RS itself is as fast as UNIFAC.

Obviously the young method COSMO-RS, which has been developed since its early begin in 1994 with an overal man power of about five years, is by far less elaborated and its overall accuracy still is slightly less than that of UNIFAC. There is much less application experience with COSMO-RS. Hence presently it is preferable for industrial users to apply a widely used GCM, if it is applicable to his problem. But COSMO-RS has a large number of methodological advantages, which make it much broader applicable than GCMs could ever be and this opens many new opportunities.

The greatest advantage of COSMO-RS is the fact that it is not structure-interpolating but *a priori*. This causes a rather general applicability. Due to the generic functional for the interaction energies, which is based on the information from quantum chemical calculations and which does only require a few at most element-specific parameters it is applicable to almost the entire organic chemistry. Thus it can be applied to nearly any system for which no group parameters are available in GCMs. These may be rare functional groups, heteroaromatic compounds, instable group combinations like -NCO/water, and even metastable complexes or transition states, for which GCM parameters will never exist.

An important advantage of COSMO-RS is its ability to handle intramolecular interactions of functional groups, while in GCMs each functional group has context-independent interaction parameters. Thus COSMO-RS "sees"' the difference between primary, secondary, and tertiary hydroxy groups in alcohols, because these groups differ in the underlying quantum chemical calculations. COSMO-RS recognizes the much stronger polarisation of a nitro group in a push-pull-compound like p-nitroaniline compared to p-dinitrobenzene. Even more important, it is capable to take into account the extreme reduction of external interactions by internal hydrogen bonds, if these are possible. All these partly very strong effects are not resolved in GCMs, resulting in large errors of these models if one of these effects occurs.

As a result of the resolution of molecular details COSMO-RS is very well able to resolve differences between isomers, which are generally identical in GCMs. Thus it may be used for the efficient search for entrainers [7] in isomer separation problems by doing quantum chemical COSMO calculations for both isomers, which in many cases will run over night,  and by a fast subsequent COSMO-RS screening using a database of COSMO files of common solvents.

The exact thermodynamics is another merit of COSMO-RS compared to present GCMs. The latter tend to fail for infinite dilution activity coefficients of strongly interacting compounds due to the implicit mean field assumption if they are parametrized to reproduce finite concentration activities. COSMO-RS gives rather homogeneous and reliable quality in the entire concentration range, because it does not suffer from any mean field approximation in thermodynamics. In contrast UNIFAC requires separate parametrisations for liquid-liquid and liquid-vapor equilibria [3] and even for infinite dilution and for partition coefficients [18,19].

Temperature dependencies can much better be described by COSMO-RS than by GCMs since for the few energy terms in COSMO-RS physically well founded approaches can be made and only a few parameters have to be adjusted. Regular UNIFAC does not allow for any temperature dependence of the interaction parameters and hence cannot give reliable enthalpies and temperature dependencies at all. Mod-UNIFAC [9,10] is able to account for such temperature dependencies but at the costs of an extreme increase of the adjustable parameters and hence of the required experimental data.

Although in most chemical engineering applications the vapor pressure of the pure compounds is well known, it should be mentioned that in contrast to GCMs COSMO-RS is able to calculate vapor pressures as well. This may be of use for the consistency check of very low vapor pressures, which often are subject to considerable experimental errors, or just for a first estimate for compounds, for which vapor pressures are not yet available.

Finally, an important advantage is the graphicness and physical interpretability of COSMO-RS, as illustrated by the σ-profiles and σ-potentials in section section 5. Thus COSMO-RS does not only give reasonable numbers for activity coefficients, but it also helps the user to understand and interpret his system. Viewing the σ-profiles and σ-potentials of his compounds and mixtures, he may find out which kind of solvent or co-solvent may be required to achieve a certain effect on activities just by analyzing which range of σ-values is missing for beneficial interactions. In the end he is able to trace back the σ-profile and thereby the properties to the molecular surface and thus graphically sees the source of the physicochemical behavior of the compounds displayed on the molecule [31].

## *8. Summary and Outlook*

COSMO-RS is a novel approach to the calculation of activity coefficients of compounds in liquid mixtures and pure compound vapor pressures and thus it is an alternative and/or supplement to group contribution methods, which presently are widely used for such calculations. Like GCMs, COSMO-RS is a surface interaction model, but apart from a few at maximum element specific adjustable parameters, which have been fitted to experimental data it is based exclusively on information derived from individual uni-molecular quantum chemical calculations for each of the compounds in the system. The screening charge density σ supplied by a virtual conductor on the molecular surface and provided by the continuum solvation model COSMO is the key parameter for molecular interactions in COSMO-RS.

Regarding quantum chemical calculations density functional theory provides the best trade-off between computational cost and accuracy. DFT-Codes with efficient COSMO implementation allow for the over-night calculation of many molecules of relevance to chemical engineering even on a simple PC. Systems up to 30 heavy atoms can be treated in about a week. Thus, a large range of chemistry is treatable within a reasonable time frame today and the rapid development of computers and the use of parallel algorithms will increase this range continuously.

Since COSMO-RS only requires pre-calculated quantum chemical output, which can be hold in a database, the thermodynamical calculations themselves are very fast.Thus any calculation of activities of compounds in arbitrary mixtures of pre-calculated compounds at variable temperature can be done in milliseconds. Thus, in combination with a good database of COSMO-files for common solvents, COSMO-RS allows for very efficient and fast thermodynamic calculations and even for large-scale solvent screening.

The statistical thermodynamics of COSMO-RS is to some degree analogous to that of GCMs, but it replaces the implicit mean-field approximation made in GCMs by a self-consistent and exact solution, which nevertheless is very fast. Thus COSMO-RS allows for the efficient and consistent calculation of all thermodynamic quantities in mixtures. The entire framework of a COSMO-RS calculation is summarized as a flow chart in figure 9.

Presently, COSMO-RS is parametrized for the elements H, C, N, O, F, S, Cl, Br, and I. Due to its generic nature it is applicable to almost all neutral compounds with these elements. Molecules with additional elements like Si and P can be handled as well, if these atoms are essentially not exposed to the solvent.

Beyond its present quality and capability, COSMO-RS is open for a large number of qualitative improvements and functional extensions. A correction for misfit charge interactions will improve the accuracy of the electrostatic part and enable calculations for ions. Refined approaches for the dispersive interactions should improve the accuracy mainly for F and I, which presently are subject to some error by a mean-field assumption in the dispersive energies due to their very low and high polarizabilities, respectively. The introduction of a volume decrease by strong interactions should allow for some reasonable pressure-dependency of the activity coefficients, which is neglected so far. First ideas of for an extension of COSMO-RS to the critical region by allowing for vacuum interactions are just being worked out. The progress of the method and more applications will be reported continuously in this journal.

## *9. List of symbols*

| | |
|---|---|
| vdW | van der Waals term |
| e | element |
| $\tau$ | van der Waals coefficient |
| $\varepsilon$ | dielectric constant |
| $\Phi$ | electrostatic potential |
| q | screening charge |

| | |
|---|---|
| $\sigma$ | screening charge density |
| X | chemical compound considered as solute |
| S | solvent, either pure or mixture |
| hb | hydrogen bonding term |
| $\mu$ | chemical potential |
| $p^X(\sigma)$ | sigma profile of a compound |
| $p_S(\sigma)$ | sigma profile of a solvent |
| $\gamma$ | activity coefficient |
| $a_{eff}$ | effective contact area |
| res | residual term |
| comb | combinatorial term |
| SG | Staverman-Guggenheim term |
| A | area |
| V | volume |
| R | gas constant |
| T | temperature |
| $\omega$ | coefficient for ring correction |
| $\eta_{gas}$ | free energy diffeence between reference states in gas and liquid state |
| x | molar fraction |
| $\Theta$ | group constitution function |
| $\nu$ | number of groups in compound |
| Q | relative group area |
| a | UNIFAC group interaction parameter |

## *Acknowledgements*

The authors are grateful to Prof. W. Arlt (TU Berlin), Dr. I. Clausen (TU Berlin / BASF AG), G. Ehlker (RWTH Aachen), Dr. R. Treckmann, Dr. G. Olf, Dr. O. Pfohl, and Dr. J.C.W. Lohrenz (all Bayer AG) for valuable suggestions, discussions, and assistance regarding experimental data, respectively. They are specially indebted to Prof. Arlt for critically reading the manuscript.

## *References*

[1] A. Fredenslund, J. Gmehling, P. Rasmussen, Vapor Liquid Equilibria Using UNIFAC, Elsevier: Amsterdam (1977)
[2] S.I. Sandler, Chemical and Engineering Thermodynamics 3. Edition, Wiley, 1998
[3] J. Gmehling, Fluid Phase Equilibria 144 (1998) 37
[4] A. Klamt, J. Phys. Chem. 99 (1995) 2224
[5] A. Klamt, V. Jonas, T. Buerger, J.C.W. Lohrenz, J. Phys. Chem. 102 (1998) 5074
[6] A. Klamt: 'COSMO and COSMO-RS', in 'Encyclopedia of Computational Chemistry', P. v. R. Schleyer Editor, Wiley, New York (1998)
[7] I. Clausen, PhD-Thesis, Technische Universität Berlin (1999)
[8] A. Bondi, Phys. Chem. 68 (1964) 441-451
[9] G. Nocon, U. Weidlich, J. Gmehling, U. Onken,
Ber. Bunsenges. Phys. Chem. 87 (1983) 17
[10] J. Gmehling, J. Lohmann, A. Jakob, Jiding Li, R. Joh, Ind. Eng. Chem. Research 37 (1998) 4876
[11] A. Klamt, G. Schüürmann, J. Chem. Soc. Perkin Trans.2 (1993) 799
[12] Born,M. Z. Phys. 1 (1920) 45
[13] J. Tomasi, M. Persico, Chem. Rev. 94 (1994) 2027
[14] Cramer,C.J., Truhlar, D.G., in Reviews in Computational Chemistry, Vol. 6 ,
K.B. Lipkowitz, D.B. Boyd, Eds., VCH Publishers, New York (1995)
[15] K. Bienert, A. Klamt, D. Krockenberger, F. Nader, B. Sewekow, R. Wittlinger, Z. Umweltchem.
Ökotox. 5 (1993) 228
[16] S. Maaßen, W. Arlt, A. Klamt, Chemie-Ingenieur-Technik 67 (1995) 476
[17] A.J.Staverman, Recl. Trav. Chim. Pays-Bas, 69 (1950) 163
[18] J.C. Bastos, M.E. Soares, A.G. Medina, Ind. Eng. Chem. Res. 27 (1988) 1269
[19] G. Wienke, J. Gmehling, Toxicological and Environmental Chemistry, 65 (1998) 57;
Erratum: 67 (1998) 275

[20] B. Delley, J. Chem. Phys. 92 (1990) 508-517 and 94 (1992) 7245-7250,
DMol, version 950, Biosym Technologies, San Diego, CA (1995)
[21] J. Andzelm, C. Kölmel, A. Klamt, J. Chem. Phys. 103 (1995) 9312-9320
[22] DMol$^3$ is available from MSI in the Cerius$^2$ program suite.
[23] R. Ahlrichs, M. Bär, M. Häser, H. Horn,
and C. Kölmel, Chem. Phys. Letters 162 (1989) 165
[24] A. Schäfer, A. Klamt, J.C.W. Lohrenz, D. Sattel, to be published
[25] Most of the data were taken from the Bayer inhouse property database is a collection of data
from different sources
[26] THOR database, Daylight CIS, Irvine CA, USA (1990)
26[27] I.B. Rabinovich, P.N. Nikolaev, Zh. Fiz. Khim. 34 (1960) 2298
27[28] A. Apelbal, A.Tamir, M. Wagner, Fluid Phase Equilib. 4 (1980) 229
28[29] I, Nagata, K. Tamura, S. Tokuriki, Thermochim. Acta 47 (1981) 315
29[30] Y. Akamatsu, H. Ogawa, S. Murakami, Thermochim. Acta 113 (1987) 141
[30] E. Schreiber, E. Schuettau, D. Rant, H. Schuberth, Z. Phys. Chem. (Leipzig) 247 (1967) 23[31] C.P. Smith, E.W. Engel, J. Amer. Chem. Soc. 51 (1929) 2660
[32] A. Gusovius, Diploma Thesis, University of Darmstadt (1997)
[31] For examples of molecular surfaces colored by their screening charge density see URL: `http://www.cosmologic.de`

| element | cavity radius [Å] | Bondi radius [Å] | ratio | dispersion coeff. [kJ/mol/Å²] |
|---|---|---|---|---|
| H | 1.30 | 1.20 | 1.08 | 0.0382 |
| C | 2.00 | 1.70 | 1.18 | 0.0348 |
| N | 1.83 | 1.55 | 1.18 | 0.0221 |
| O | 1.72 | 1.52 | 1.13 | 0.0366 |
| F | 1.72 | 1.47 | 1.17 | 0.0265 |
| S | 2.16 | 1.80 | 1.20 | 0.0510 |
| Cl | 2.05 | 2.05 | 1.17 | 0.0514 |
| Br | 2.16 | 1.85 | 1.17 | 0.0550 |
| I | 2.32 | 1.98 | 1.17 | 0.0580 |

Table 1: Element specific parameters

## *Figure captions*

Figure 1: Schematic illustration of contacting molecular cavities and contact interactions.

Figure 2: $\sigma$-profiles of representative solvents.

Figure 3: $\sigma$-potentials of representative solvents at T = 298.15 K.

Figure 4a: Activity coefficients for the binary mixture of acetone and chloroform at T = 298.15 K. Experimental values calculated from thermodynamical data of Rabinovich and Nikolaev [26] resp. Apelblat *et al.* [27].

Figure 4b: Excess Gibbs free energy and excess enthalpy for the binary mixture of acetone and chloroform at T = 298.15 K. Experimental values calculated from thermodynamical data of Nagata *et al.* [28], Akamatsu *et al.* [29] (H-Excess) resp. Rabinovich and Nikolaev [26] and Apelblat *et al.* [27] (G-Excess).

Figure 5a: Activity coefficients for the binary mixture of 1-butanol and water at T = 333.15 K. Experimental values calculated from thermodynamical data of Schreiber *et al.* [30].

Figure 5b: x-y-diagram for the binary mixture of 1-butanol and water at T = 333.15 K. Experimental values taken from Ref. [30].

Figure 6: x-y-diagram for the binary mixture of benzene and ethanol at T = 333.15 K. Experimental values taken from Ref. [6].

Figure 7: x-y-diagram for the binary mixture of cyclohexane and aniline at T = 323.35 K. Experimental values taken from Ref. [6].

Figure 8: Excess properties of the binary mixture of 1-butanol and n-heptane at T = 323.15 K. Experimental values calculated from the thermodynamical data of Smith *et al.* [31] and Gusovius [32].

Figure 9: Flow chart of COSMO-RS

Figure 1: (please convert to black and white for printing)

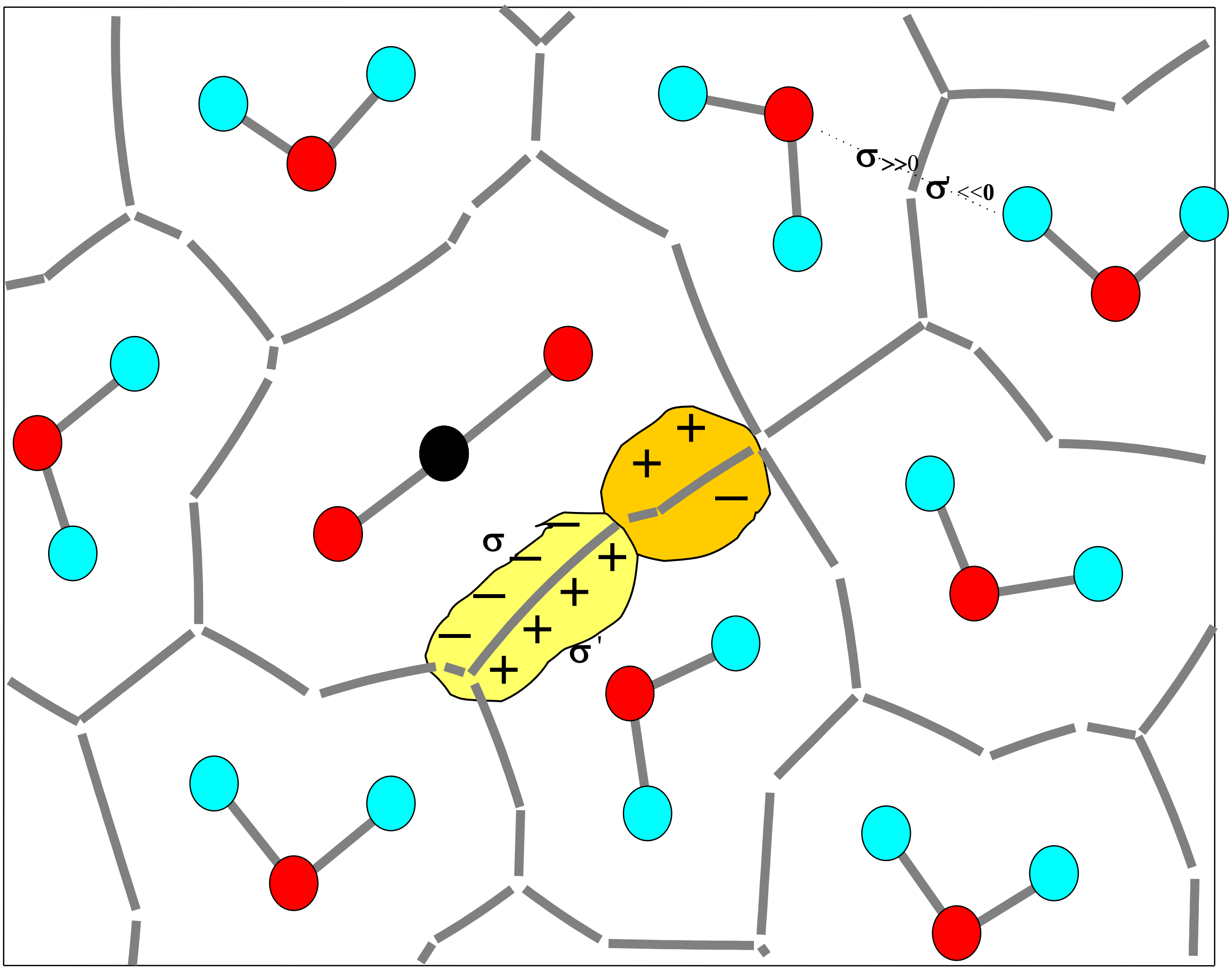


Figure 2:

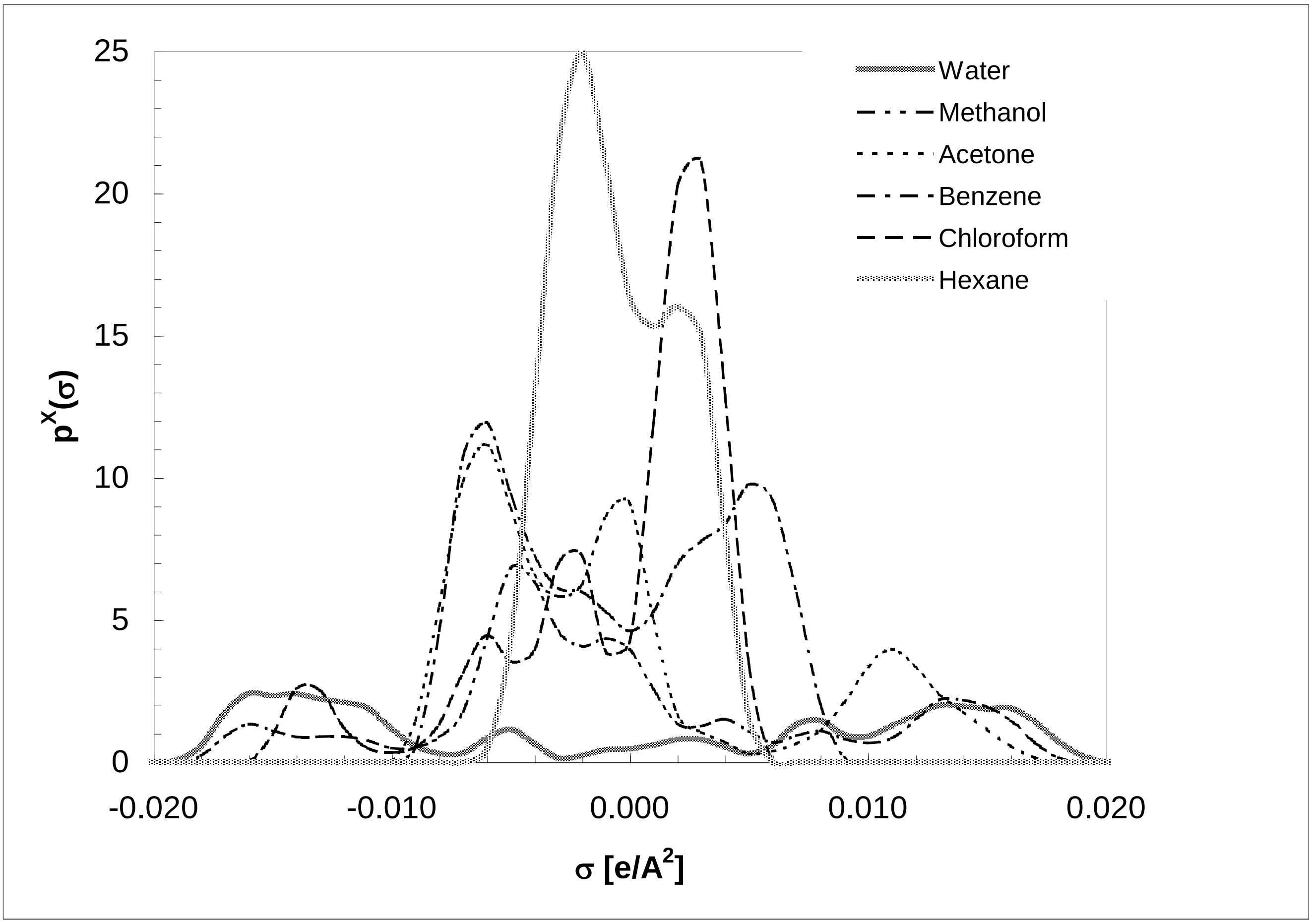

25
20
15
10
5
0
$p^X(\sigma)$
-0.020
-0.010
0.000
0.010
0.020
σ [e/A²]
Water
Methanol
Acetone
Benzene
Chloroform
Hexane

Figure 3:

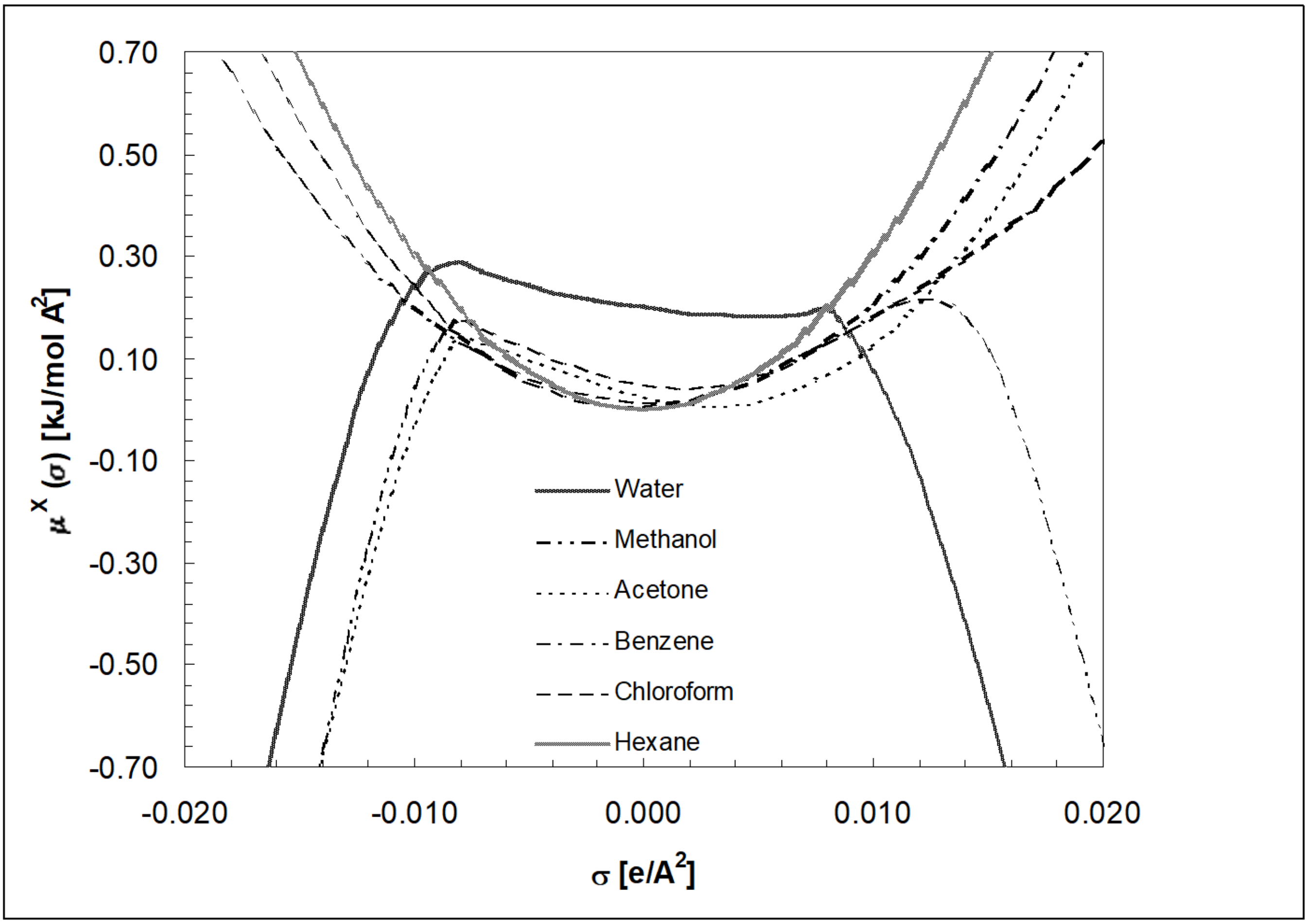

0.70
0.50
0.30
0.10
-0.10
-0.30
-0.50
-0.70
μX (σ) [kJ/mol A2]
-0.020
-0.010
0.000
0.010
0.020
σ [e/A2]
Water
Methanol
Acetone
Benzene
Chloroform
Hexane

Figure 4a:

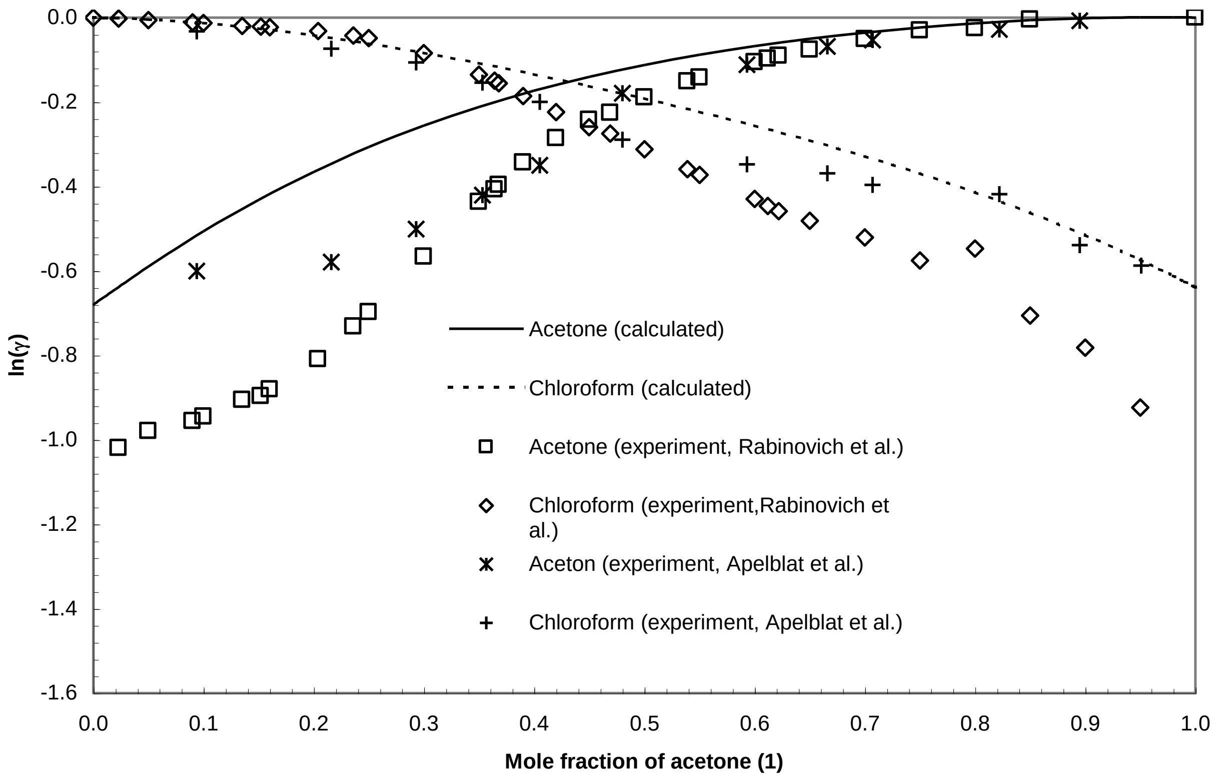


Figure 4b:

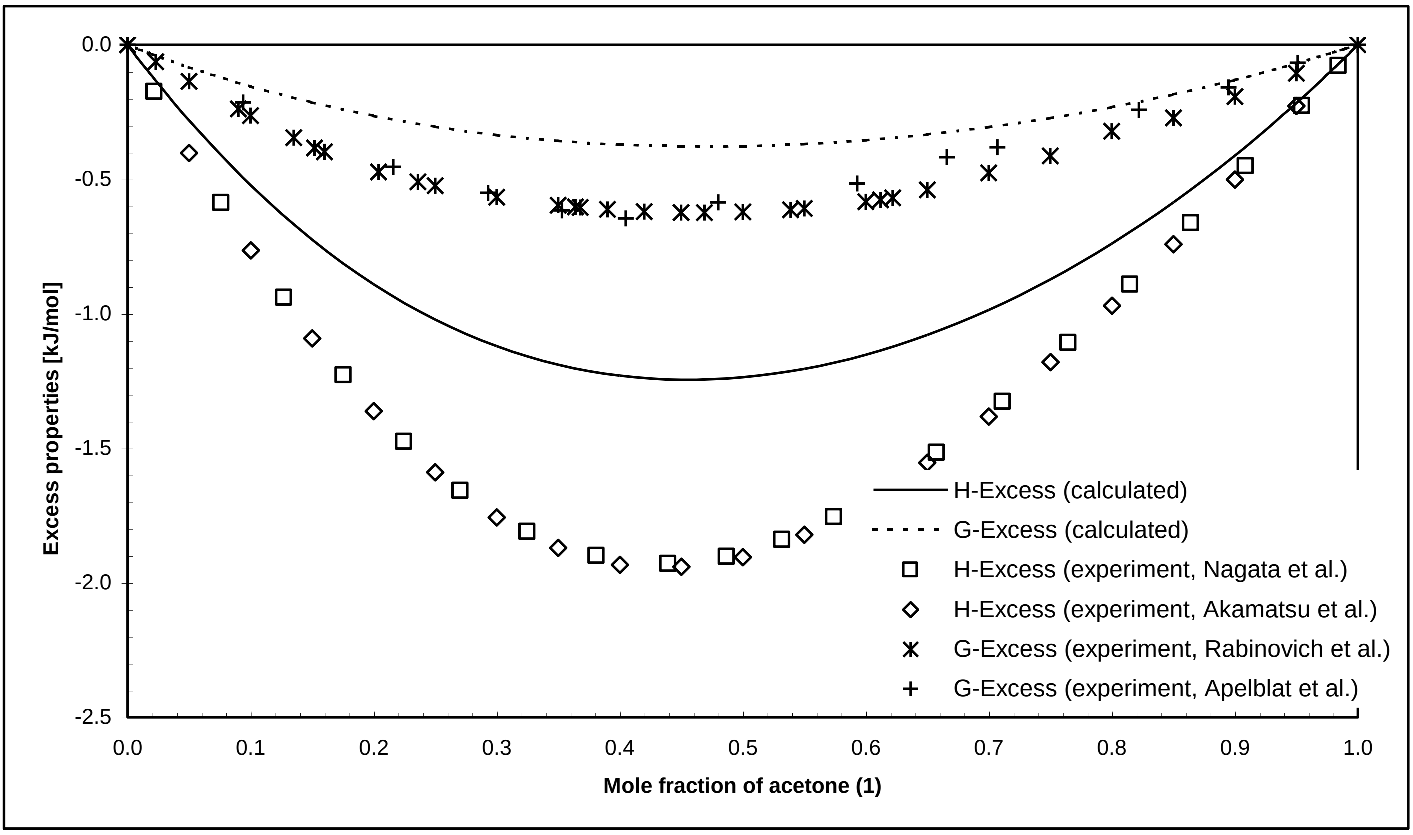


Figure 5a:

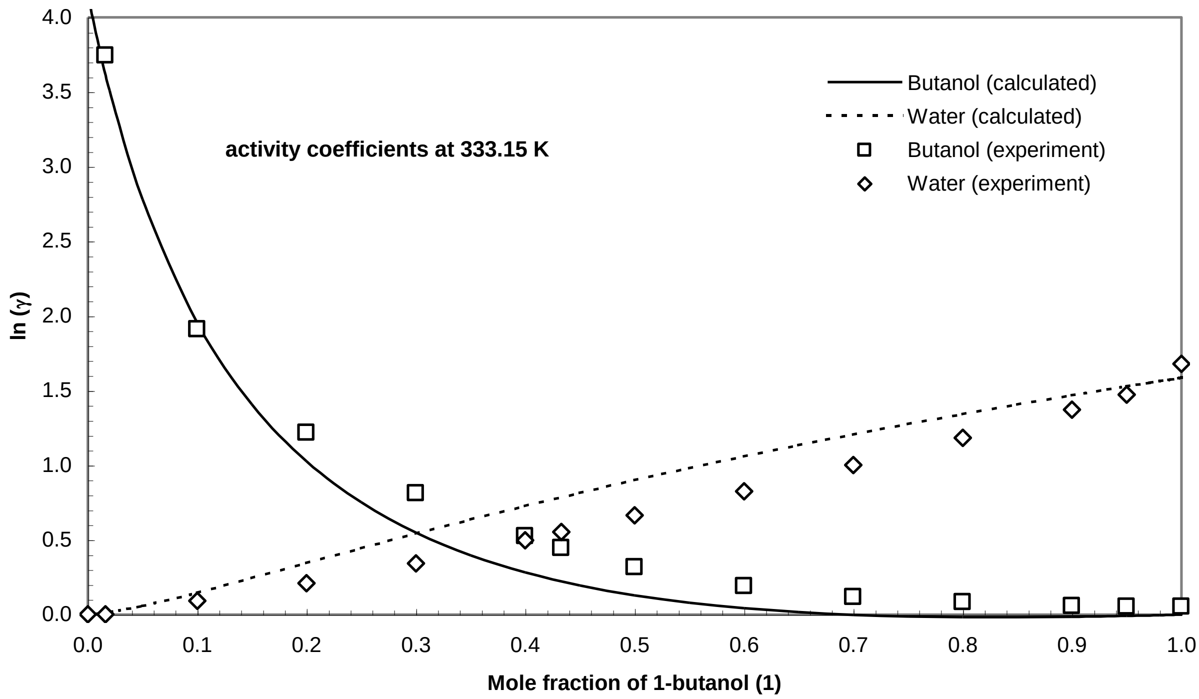


Figure 5b:

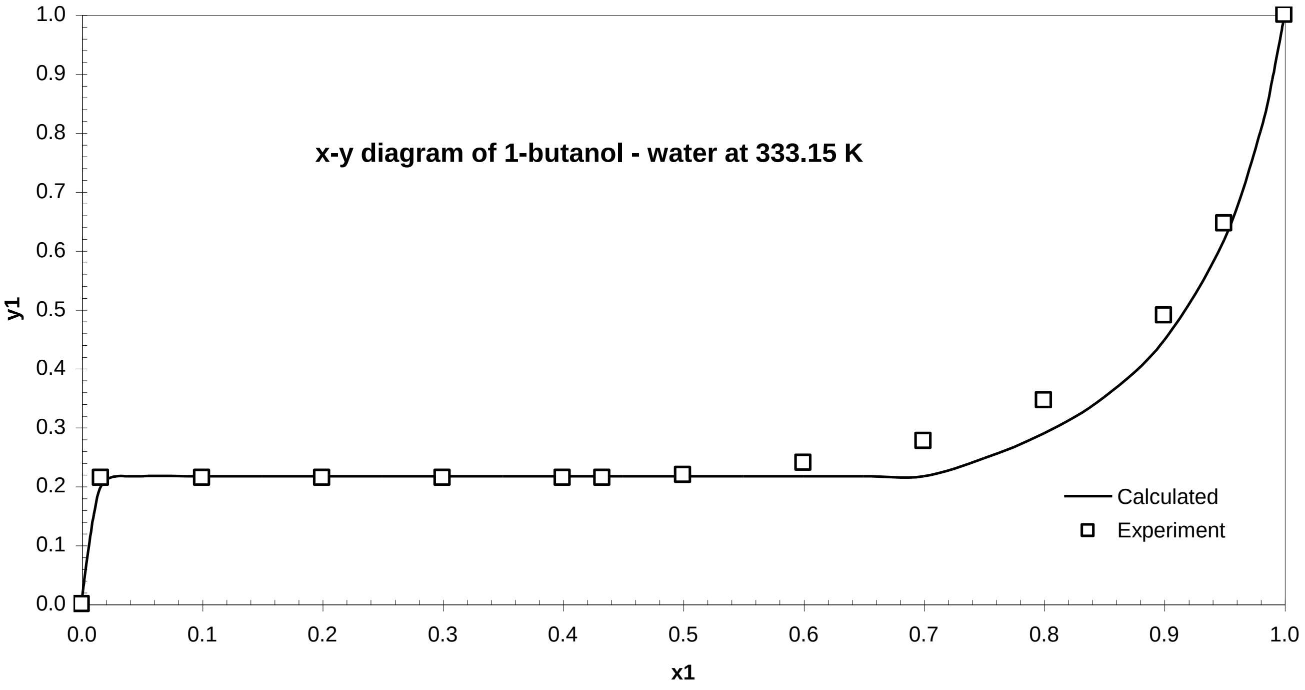


Figure 6:

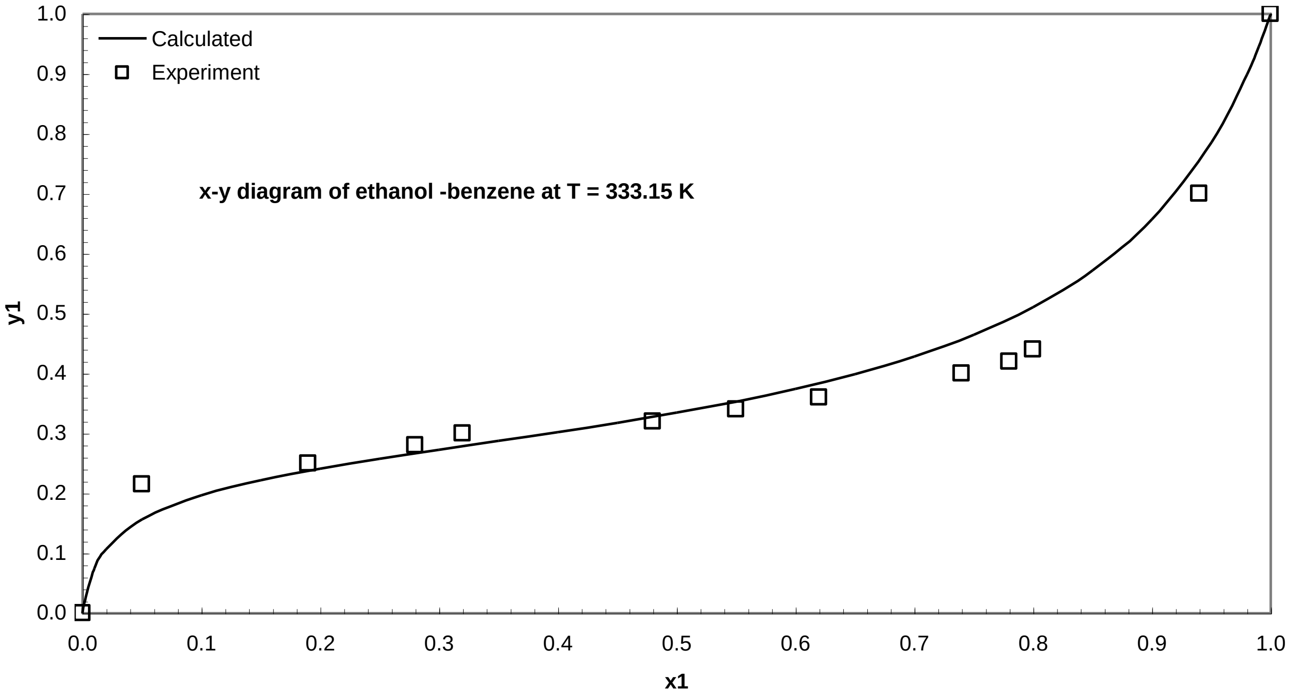


Figure 7:

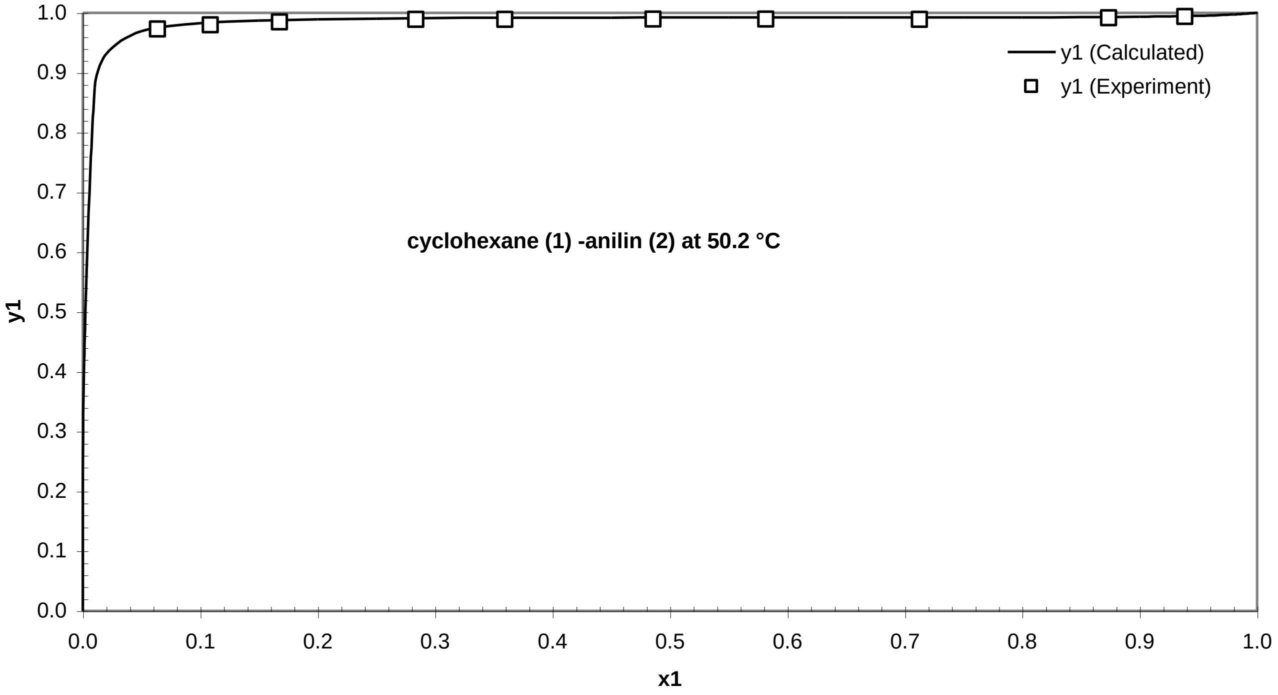
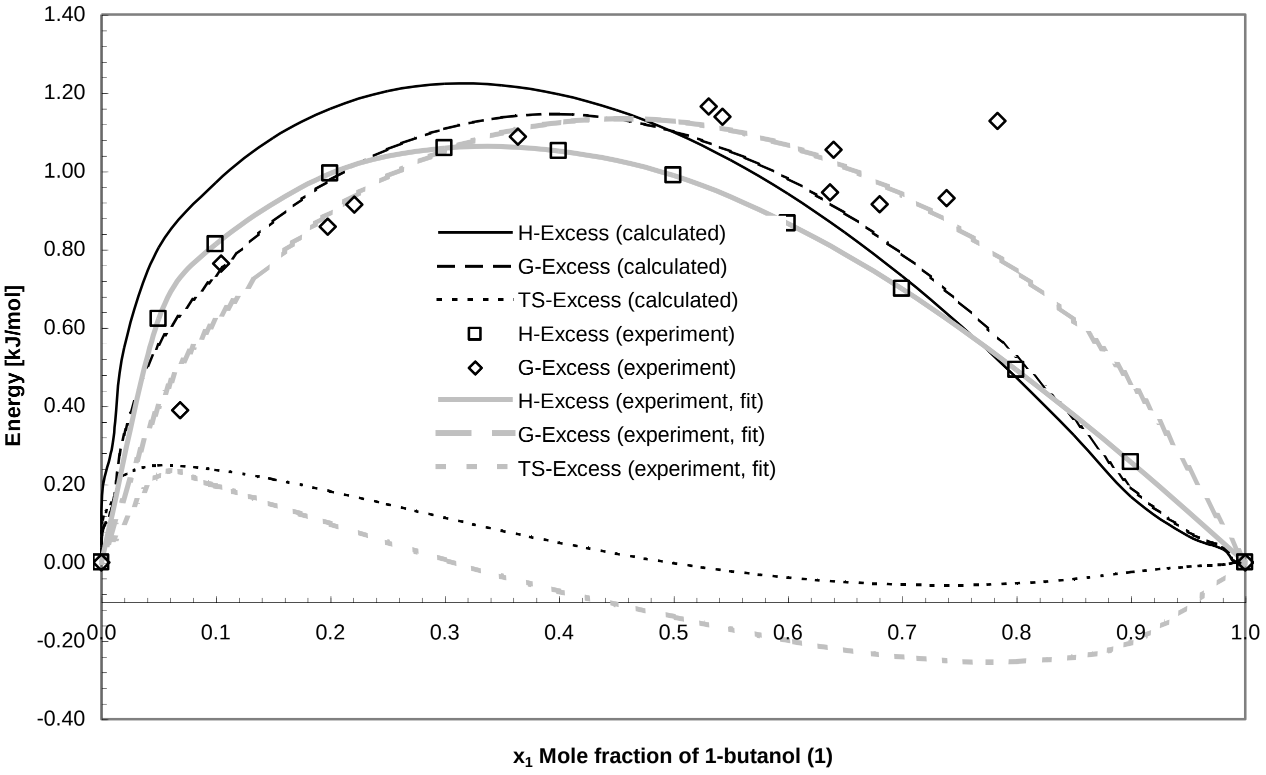
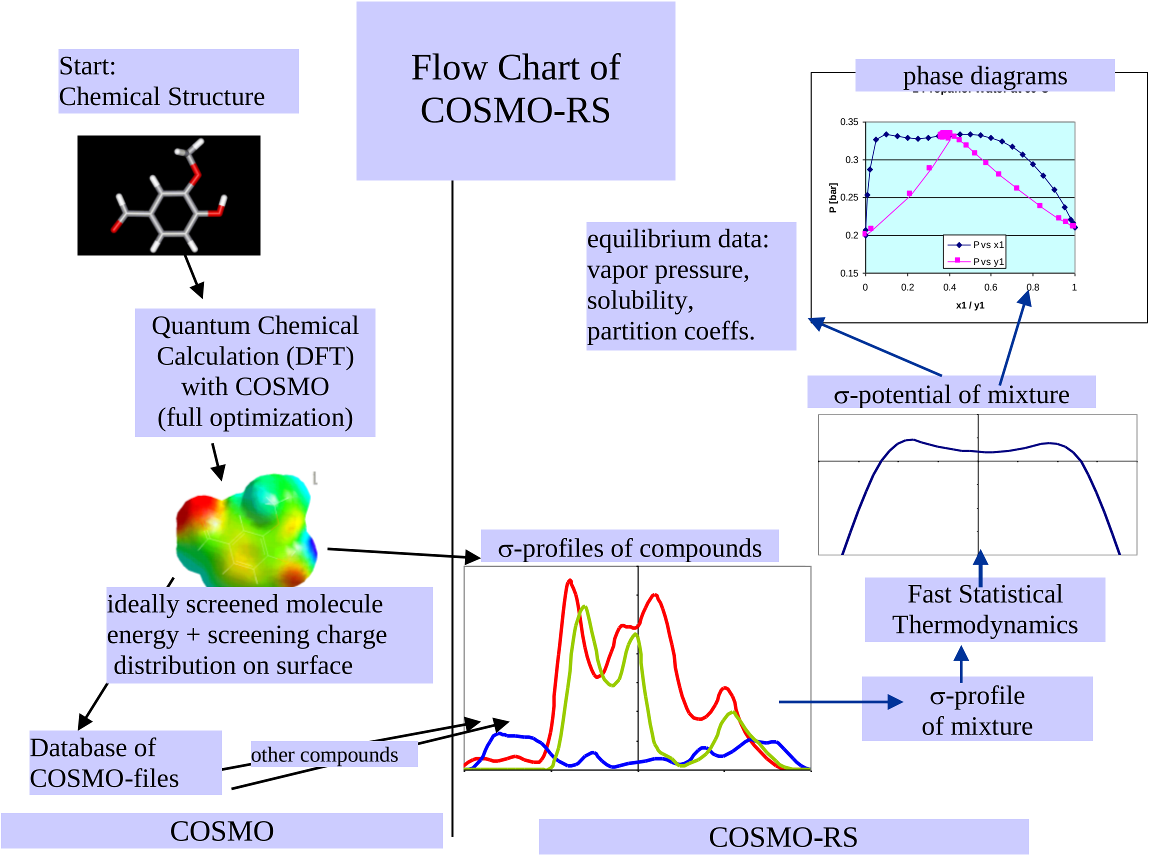


Fig.: 8:



1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
Energy [kJ/mol]
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
x1 Mole fraction of 1-butanol (1)
H-Excess (calculated)
G-Excess (calculated)
TS-Excess (calculated)
H-Excess (experiment)
G-Excess (experiment)
H-Excess (experiment, fit)
G-Excess (experiment, fit)
TS-Excess (experiment, fit)

Fig. 9: (please convert ppt-file to black and white for printing)



Flow Chart of COSMO-RS

Start:
Chemical Structure

Quantum Chemical Calculation (DFT) with COSMO (full optimization)

ideally screened molecule energy + screening charge distribution on surface

Database of COSMO-files

other compounds

COSMO

σ-profiles of compounds

σ-profile of mixture

Fast Statistical Thermodynamics

σ-potential of mixture

phase diagrams

equilibrium data:
vapor pressure,
solubility,
partition coeffs.

COSMO-RS