Fast Solvent Screening via Quantum Chemistry:
The COSMO-RS Approach

By Frank Eckert* and Andreas Klamt

COSMO*logic* GmbH&Co.KG, Burscheider Str. 515, D-51381 Leverkusen, Germany

* Corresponding author, Tel.: +49-2171-731682, e-mail: frank.eckert@cosmologic.de

Abstract

COSMO-RS, a general and fast methodology for the *a priori* prediction of thermophysical data of liquids is presented. It is based on cheap unimolecular quantum chemical calculations, which, combined with a statistical thermodynamics methodology, provide the information necessary for the evaluation of molecular interactions in liquids. COSMO-RS is an alternative to structure interpolating group contribution methods. The method is independent of experimental data and generally applicable. A methodological comparison with group contribution methods is given. The applicability of the COSMO-RS method to the goal of solvent screening is demonstrated at various examples of vapor-liquid-, liquid-liquid-, solid-liquid-equilibria and vapor pressure predictions.

Topic: Thermodynamics

Keywords: solvent screening, vapor-liquid equilibria, liquid-liquid equilibria, activity coefficient, solubility.

1. Introduction

A large part of chemical engineering critically depends upon the knowledge of the thermophysical data of solutions or mixtures of liquids (for example separation processes like distillation, liquid-liquid-extraction, dissolution of solids, crystallization from solution, extractive distillation, membrane processes, absorption and adsorption, but also the description of a chemicals environmental fate). For all of the processes mentioned above it is highly desirable to choose an optimum solvent or solvent mixture. Criteria for such an "optimal" solvent can be cost of the process technology (cost of heating/cooling, simpler or more robust processing technology, product quality) as well as the cost aspects of the environmental, health and safety properties of the process. Numerous industrial solvents and solvent mixtures are available (Flick, 1991; Gessner, 1996; Marcus, 1998). Theoretical-computational models can be valuable tools for the estimation of the thermodynamic properties of solutions and mixtures, thus reducing time, resources and overall cost of the screening of a large number of solvents. The correct description of the dependence on composition, temperature and pressure in multicomponent systems requires reliable thermodynamic models. One can distinguish two approaches to solvent screening: Firstly, methods based upon inter- or extrapolation of a given set of experimental thermodynamic data of the system. Activity coefficient models (for example the NRTL model or UNIQUAC) and equation of state models (for example the Soave-Redlich-Kwong or the Peng-Robinson equations of state) belong to this class. An overview over these methods is given by Sandler (1998), Prausnitz et al. (1999) and Reid et al. (1987). Basically, these methods all involve the fitting of some set of experimental data to a given functional form. The thermodynamic information subsequently is inter- or extrapolated from that functional form. Secondly, methods that are independent of experimental thermodynamic data of the given solutions or mixtures. Estimates are obtained from molecular structure information only. Examples for the second approach are group contribution methods (GCMs) (Fredenslund et al., 1975; Fredenslund et al., 1977; Gmehling, 1998). GCMs are based on interaction parameters that have been obtained previously by analysis of phase equilibrium data of systems containing the same functional groups. Furthermore, the UNIQUAC model has been shown to be independent of experimental data for polar/polar mixture cases when information from quantum chemical calculations is used (Lin and Sandler, 1999). A related model is COSMO-RS the "the conductor like screening model for real solvents" (Klamt, 1995) which is based upon unimolecular quantum chemical calculations of the individual species in the system (i.e. not of the mixture itself). Predictive methods often are indispensable for chemical engineers in the synthesis and design of chemical processes and plants. They are especially well suited for the task of solvent screening if reliable experimental data for a system is missing or not available at affordable cost.

This article mainly concentrates on the second approach to solvent screening. The next section presents the COSMO-RS method and gives a methodical comparison of COSMO-RS with the perhaps most prominent group contribution method, UNIFAC (Fredenslund et al., 1977). The third section

presents applications of COSMO-RS to problems of chemical and engineering thermodynamics such as vapor-liquid equilibria (VLE), liquid-liquid equilibria (LLE) and solid-liquid equilibria (SLE).

2. COSMO-RS and UNIFAC

2.1 Group Contribution Methods / UNIFAC

Group contribution methods 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. The 'UNIquac Functional group Activity Coefficient' (UNIFAC) model (Fredenslund et al., 1977) and its modified versions UNIFAC-Dortmund (Do) (Weidlich and Gmehling, 1987) and UNIFAC-Lyngby (Ly) (Larsen et al., 1987) are probably the most accurate of such group contribution methods. The theory of UNIFAC has been summarized in an excellent way by Sandler (1998). Recent developments in the field of GCMs have been reviewed by Gmehling (Gmehling, 1998).

GCMs are based on the assumption that, with appropriately defined groups, the interaction energy of any system can be well approximated by the sum of functional group interaction energies. I.e. a liquid is considered not a mixture of interacting molecules but a mixture of interacting structural groups. Thus, the number of possibly interacting species is reduced extremely and a moderate number of parameters can be used to predict properties of a relatively wide variety of systems. The introduction of interacting groups also leads to the predictive properties of the GCMs: Experimental VLE data of binary mixtures as well as other experimental thermodynamic data (such as activity coefficients in infinite dilution, LLE and SLE data) are used to fit the interaction parameters of the groups that occur in these systems. The resulting group interaction parameters can be used to predict the properties of other systems if all compounds in the system can be built from the same functional groups. This is the main difference of the GCM UNIFAC to the UNIQUAC model, which is - although algebraically equivalent to UNIFAC – based on molecular interactions.

In practice, the first step in the application of the UNIFAC method is the splitting of the molecules into functional subgroups. The activity coefficient of the system is built by summing up all of the activity coefficients of the fragment groups. The activity coefficients of a species *i* in a mixture is built from two contributions:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{1}$$

The first term, the combinatorial contribution $\ln \gamma_i^C$ accounts for the size and shape differences of the groups.

$$\ln \gamma_i^C = 1 - V_i + \ln V_i - 5q_i\left(1 - \frac{V_i}{F_i} + \ln\left(\frac{V_i}{F_i}\right)\right) \tag{2}$$

Parameters V_i and F_i are given as

$$V_i = \frac{r_i}{\sum_j r_j x_j}, \qquad F_i = \frac{q_i}{\sum_j q_j x_j}, \quad where$$
$$r_i = \sum_k R_k v_k^{(i)} \quad and \quad q_i = \sum_k Q_k v_k^{(i)} \tag{3}$$

$v_m^{(j)}$ is the number of groups of type m in compound j, Q_k denotes the relative surface area and R_k the group volume of group k.

The residual contribution to the activity coefficient $\ln \gamma_i^R$ is computed from

$$\ln \gamma_i^R = \sum_k v_k\left[\ln \Gamma_k - \ln \Gamma_k^{(i)}\right] \tag{4}$$

The residual contribution to the logarithm of the activity coefficient of group k in the mixture and in the pure compound (group activity factors $\ln\Gamma_k$ and $\ln\Gamma_k^{(i)}$) are given as

$$\ln \Gamma_k = Q_k\left[1 - \ln\left(\sum_m \Theta_m \Psi_{mk}\right) - \sum_m \frac{\Theta_m \Psi_{km}}{\sum_n \Theta_m \Psi_{nm}}\right] \tag{5}$$

The surface fraction Θ_m and mole fraction X_m of group m in the mixture are calculated from

$$\Theta_m = \frac{X_m Q_m}{\sum_n X_n Q_n}, \qquad and \qquad X_m = \frac{\sum_j v_m^{(j)} x_j}{\sum_j \sum_n v_n^{(j)} x_j} \tag{6}$$

Finally, Ψ_{mn} is defined as

$$\tag{7}$$

$$\Psi_{mn} = \exp\left(-\frac{a_{mn}}{T}\right)$$

where the a_{mn} are the group interaction parameters between two groups m and n. In *mod*-UNIFAC (Do), eq. (7) is replaced by a more elaborate term holding two additional temperature-dependent group interaction parameters for each group m and n (Weidlich and Gmehling, 1987). The group interaction parameters a_{mn} are obtained from fitting of experimental VLE data of molecules containing the considered groups. For *mod*-UNIFAC (Do), the fitting procedure also included LLE and SLE data (Gmehling et al., 1993; Gmehling et al., 1998). As is clear from eq. (7), the performance of UNIFAC critically depends upon the availability and quality of the group interaction parameters. Although a considerable amount of work of various research groups is devoted to the preparation of group interaction parameters, the matrix of available a_{mn} still shows conspicuous gaps, e.g. for the description of fluorinated compounds or amines (Gmehling et al., 1998). In addition, UNIFAC predictions are problematic if interacting groups are considered in high or infinite dilution (e.g. in the prediction of octanol-water partition coefficients, which permit an estimate of a chemicals environmental fate). This problem of UNIFAC eventually could be met by special parameterizations for the prediction of activity coefficients in infinite dilution and for octanol-water partition coefficients (Wienke and Gmehling, 1998) – however, at the prize of a further reduction of the generality of the approach.

2.2 COSMO-RS

In comparison to GCMs, the COSMO-RS approach to chemical thermodynamics starts from a completely different point of view, namely from the complete molecule or - to be more precise - from the molecular surface as computed by quantum chemical methods (QM). COSMO-RS combines an electrostatic theory of locally interacting molecular surface descriptors (which are available from QM calculations) with a statistical thermodynamics methodology (which as will be shown, holds some similarities to UNIFAC statistical thermodynamics.).

The quantum chemical basis of COSMO-RS is COSMO, the "Conductor-like Screening Model" (Klamt and Schüürmann, 1993), which belongs to the class of QM continuum solvation models (CSMs). In general, basic quantum chemical methodology describes isolated molecules at a temperature of T=0 K, allowing a realistic description only for molecules in vacuum or in the gas phase. CSMs are an extension of the basic QM methods towards the description of liquid phases. CSMs describe a molecule in solution through a quantum chemical calculation of the solute molecule with an approximate representation of the surrounding solvent as a continuum (Tomasi and Persico,

1994; Cramer and Truhlar, 1995; Cramer and Truhlar, 1999). Either by solution of the dielectric boundary condition or by solution of the Poisson-Boltzmann equation, the solute is treated as if embedded in a dielectric medium via a molecular surface or "cavity" that is constructed around the molecule. Hereby, normally the macroscopic dielectric constant of the solvent is used. COSMO is a quite popular model based on a slight approximation, which in comparison to other CSMs achieves superior efficiency and robustness of the computational methodology (Klamt and Schüürmann, 1993; Klamt, 1998). The COSMO model is available in several quantum chemistry program packages: Turbomole (Schäfer et al., 2000), DMOL3 (Andzelm et al., 1995), GAMESS-US (Baldridge et al., 1998) and Gaussian (Frisch et al., 2001). If combined with accurate QM CSMs have been proven to produce reasonable results for properties like Henry law constants or partition coefficients (Cramer and Truhlar, 1999). However, as has been shown elsewhere (Klamt, 1995; Klamt, 1998) the continuum description of CSMs is based on a erroneous physical concept: The macroscopic dielectric continuum theory is a linear response theory whereas the electric fields on molecular surfaces of polar molecules are so strong that the major part of the polarizable continuum (i.e. the "solvent") does not behave linearly. This leads to saturation effects that cannot be captured by a linear response theory. The success of CSMs in some areas of application is due to different reasons (Klamt, 1998). In addition, concepts of temperature and mixture are missing in CSMs.

COSMO-RS, the COSMO theory for "real solvents" goes far beyond simple CSMs in that it integrates concepts from quantum chemistry, dielectric continuum models, electrostatic surface interactions and statistical thermodynamics. Still, COSMO-RS is based upon the information that is evaluated by QM-COSMO calculations. Basically QM-COSMO calculations provide a discrete surface around a molecule embedded in a virtual conductor (Klamt and Schüürmann, 1993). Of this surface each segment i is characterized by its area a_i and the screening charge density (SCD) σ_i on this segment which takes into account the electrostatic screening of the solute molecule by its surrounding (which in a virtual conductor is perfect screening) and the back-polarization of the solute molecule. In addition, the total energy of the ideally screened molecule E_{COSMO} is provided. Within COSMO-RS theory a liquid is now considered an ensemble of closely packed ideally screened molecules as shown in Figure 1. In order to achieve this close packing the system has to be compressed and thus the cavities of the molecules get slightly deformed (although the volume of the individual cavities does not change significantly). As visible in Figure 1, each piece of the molecular surface is in close contact with another one. Assuming that there still is a conducting surface between the molecules, i.e. that each molecule still is enclosed by a virtual conductor, in a contact area the surface segments of both molecules have net SCDs σ and σ' (compare Figure 1). In reality there is no conductor between the surface contact areas. Thus an electrostatic interaction arises from the contact of two different SCDs. The specific interaction energy per unit area resulting from this "misfit" of SCDs is given by

$$E_{misfit}(\sigma, \sigma') = a_{eff} \frac{\alpha'}{2}(\sigma + \sigma')^2 \tag{8}$$

where a_{eff} is the effective contact area between two surface segments and α' is an adjustable parameter. The basic assumption of eq. (8) (which is the same as in other surface pair models like UNIQUAC) is that residual non-steric interactions can be described by pairs of geometrically independent surface segments. Thus, the size of the surface segments a_{eff} has to be chosen in a way that it effectively corresponds to a thermodynamically independent entity. There is no simple way to define a_{eff} from first principles and it must be considered to be an adjustable parameter (Klamt, Krooshof and Taylor, submitted). Obviously, if σ equals -σ' the misfit energy of a surface contact will vanish. Hydrogen bonding (HB) can also be described by the two adjacent SCDs. HB donors have a strongly negative SCD whereas HB acceptors have strongly positive SCDs. Generally, a HB interaction can be expected if two sufficiently polar pieces of surface of opposite polarity are in contact. Such a behavior can be described by a functional of the form

$$E_{HB} = a_{eff} c_{HB} \, min\big[0; min(0; \sigma_{donor} + \sigma_{HB}) max(0; \sigma_{acceptor} - \sigma_{HB})\big] \tag{9}$$

wherein c_{HB} and σ_{HB} are adjustable parameters. In addition to electrostatic misfit and HB interaction COSMO-RS also takes into account van der Waals (vdW) interactions between surface segments via

$$E_{vdW} = a_{eff}\left(\tau_{vdW} + \tau'_{vdW}\right) \tag{10}$$

wherein τ_{vdW} and τ'_{vdW} are element specific adjustable parameters. The vdW energy is dependent only on the element type of the atoms that are involved in surface contact. It is spatially non-specific. E_{vdW} is an additional term to the energy of the reference state in solution. Currently nine of the vdW parameters (for elements H, C, N, O, F, S, Cl, Br and I) have been optimized (see below). For the majority of the remaining elements reasonable guesses are available (Klamt et al., 1998).

The link between the microscopic surface interaction energies and the macroscopic thermodynamic properties of a liquid is provided by statistical thermodynamics. Since in the COSMO-RS view all molecular interactions consist of local pair wise interactions of surface segments, the statistical averaging can be done in the ensemble of interacting surface pieces. Such an ensemble averaging is computationally efficient – especially in comparison to the computationally very demanding molecular dynamics or Monte Carlo approaches which require averaging over an ensemble of all possible different arrangements of all molecules in a liquid. As a result, the computational effort of a COSMO-RS calculation is not significantly higher than that of a UNIFAC calculation. To describe the composition of the surface segment ensemble with respect to the interactions (which depend on σ only), only the probability distribution of σ has to be known for all compounds X_i. Such probability distributions $p^X(\sigma)$ are called "σ-profiles". The σ-profile of the whole system/mixture $p_S(\sigma)$ is just a sum of the σ-profiles of the components X_i weighted with their mole fraction in the mixture x_i.

$$p_S(\sigma) = \sum_{i \in S} x_i p^{X_i}(\sigma) \tag{11}$$

The chemical potential of a surface segment with SCD σ in an ensemble described by normalized distribution function $p_S(\sigma)$ is given by

$$\mu_S(\sigma) = -\frac{RT}{a_{eff}} ln\left[\int p_S(\sigma') exp\left(\frac{a_{eff}}{RT}\left(\mu_S(\sigma') - E_{misfit}(\sigma, \sigma') - E_{HB}(\sigma, \sigma')\right)\right) d\sigma'\right] \qquad (12)$$

$\mu_S(\sigma)$ is a measure for the affinity of the system S to a surface of polarity σ. It is a characteristic function of each system and is called "σ-potential". Please note, that E_{vdW} is not included in eq. (12) (i.e. not part of the statistical averaging) because it is not a function of individual surface contacts. Instead, E_{vdW} is added to the reference energy in solution *a posteriori*. Eq. (12) is an implicit equation. It must be solved iteratively. This is done in milliseconds on any PC. Thus COSMO-RS computations of thermodynamic properties are very fast (see section 3). A detailed description and a rationale of this statistical averaging procedure are given by Klamt (1995). The chemical potential (the partial Gibbs free energy) of compound X_i in system S is readily available from integration of the σ-potential over the surface of X_i

$$\mu_S^{X_i} = \mu_{C,S}^{X_i} + \int p^{X_i}(\sigma)\mu_S(\sigma)d\sigma \qquad (13)$$

$\mu^{Xi}_{C,S}$ is a combinatorial contribution to the chemical potential. It contains one adjustable parameter λ_C:

$$\mu_{C,S}^{X_i} = \frac{\partial G_{C,S}}{\partial N_i} \quad where \quad G_{C,S} = -NkT\lambda_C\left[\sum_i sin^2\left(\frac{x_i\pi}{2}\right)ln\left(\frac{A_i}{A_0}\right) + ln(A_0)\right] \qquad (14)$$

Please note, that the chemical potential of eq. (13) is a "pseudo-chemical potential" which is the standard chemical potential minus $RT\ln(x_i)$ (Ben-Naim, 1987). The chemical potential μ^{Xi}_S of eq. (13) allows for the prediction of almost all thermodynamic properties of compounds or mixtures, such as activity coefficients, excess properties or partition coefficients and solubility. The course of a COSMO-RS calculation is illustrated in Figure 4. The starting point is always a QM-COSMO calculation. However, the time-consuming QM-COSMO calculations have to be done only once for each compound. The results of the QM-COSMO calculations (i.e. the charge distribution on the molecular surface) can be stored in a database. COSMO-RS then can be run from a database of stored QM-COSMO calculations. Thus COSMO-RS is well suited for the task of screening large numbers of solvents or solutes if an appropriate database of QM-COSMO calculations is available.

The COSMO-RS representations of molecular interactions namely the σ-profiles and σ-potentials of compounds and mixtures, respectively, contain valuable information - qualitatively as well as quantitatively. Figures 2 and 3 show the σ-profiles and the room temperature σ-potentials of the four solvents water, acetone, chloroform and hexane, respectively. Of these, hexane is the least polar compound. This is reflected in the narrow distribution of the charge densities around zero in Figure 2. The two peaks can be assigned to the carbon atoms for positive σ and to the hydrogen atoms for

negative σ values (please keep in mind that negative partial charges of atoms cause positive screening charge densities and vice versa). The corresponding σ-potential, which is a measure for the affinity of the solvent to a molecular surface of polarity σ, is a simple parabola centered at σ=0 (see Figure 3). Such a shape arises from misfit contributions only (no hydrogen bonding) and is equivalent to purely dielectric behavior. The other extreme is represented by the σ-profile of water: It is very broad and the probability for σ is almost zero at the center of the σ-profile. The broad peak around –0.015 $e/Å^2$ arises from the two very polar hydrogen atoms whereas the peak around +0.015 $e/Å^2$ results from the lone pairs of the oxygen. This reflects the excellent ability of water to act as a donor as well as an acceptor for hydrogen bonding. In addition such a symmetric shape of the σ-profile indicates a favorable electrostatic interaction of water with itself, explaining its high boiling point and surface tension. The corresponding σ-potential has a much higher value around zero reflecting an unfavorable interaction with nonpolar surface. This is reflected the much stronger hydrophilicity of water in comparison to hexane. The shape of the outer regions of the σ-potential is due to hydrogen bonding: If a hydrogen bond donor in another compound has a SCD that is greater than 0.01 $e/Å^2$ or if a hydrogen bond acceptor has a SCD that is below -0.01 $e/Å^2$, it can build hydrogen bonds with water. The σ-profile of acetone is not symmetric. The peak at +0.012 $e/Å^2$ resulting from the carbonyl oxygen indicates hydrogen bonding acceptor capacity. However, unlike water there is no corresponding peak in the hydrogen bonding donor area. Therefore the interaction of acetone with itself is very unfavorable, explaining its relatively low boiling point and surface tension. This is also reflected in the σ-potential: While on the positive side it shows almost parabolic behavior (no hydrogen bonding donor capacity), on the negative side it quickly becomes strongly negative. Compared to water, the hydrogen bonding acceptor capacity of acetone is stronger, which is reflected in the smaller σ-values at which the σ-potential becomes negative. The σ-profile of chloroform shows three peaks in the region around zero that derive from the chlorine atoms. The peak at –0.013 $e/Å^2$ correspond to an acidic hydrogen atom. But due to quite small area of this peak, no significant hydrogen bonding donor capacity can be expected from this hydrogen atom. This is clearly visible from the σ-potential, which is not getting negative in the region of large positive σ values. As for the acetone, the asymmetric shape of the σ-profile indicates an unfavorable interaction of chloroform with itself, again resulting in a relatively low boiling point. It should be noted, that the σ-profiles of acetone and chloroform are almost complementary in the region of misfit interactions (i.e. for σ values between –0.008 and +0.008 $e/Å^2$) . This means that they should mix quite favorably. This is in fact the case as can be seen from the strongly negative excess enthalpy of acetone-chloroform mixtures (see also section 3.1). To sum up, one can say that σ-profiles and σ-potentials can be used to qualitatively interpret the interactions in a compound or a mixture, e.g. to assert a certain solvent or co-solvent which has a certain effect on the activities in a solution or mixture.

In addition to the prediction of thermodynamics of liquids and unlike GCMs COSMO-RS is also able to provide a reasonable estimate of a pure compound's chemical potential in the gas phase

$$\mu_{Gas}^{X_i} = E_{Gas}^{X_i} - E_{COSMO}^{X_i} - E_{vdW}^{X_i} + \omega_{Ring} n_{Ring}^{X_i} + \eta_{Gas} \tag{15}$$

where E^{Xi}_{Gas} and E^{Xi}_{COSMO} are the quantum chemical total energies of the molecule in the gas phase and in the COSMO conductor respectively. E^{Xi}_{vdW} is the vdW energy of X_i. The remaining contributions consist of a correction term for ring shaped molecules with n^{Xi}_{Ring} being the number of ring atoms in the molecule and ω_{Ring} an adjustable parameter as well as parameter η_{Gas} providing the link between the reference states of the system's free energy in the gas phase and in the liquid. Using eqs. (13) and (15) it is possible to *a priori* predict vapor pressures of pure compounds (see below, section 3.4). Please note, that eq. (15) is an empirical formulation. It is not part of the rigorous statistical thermodynamics approach that leads to eqs. (11) - (14). Eq. (15) is valid for pure compounds only.

Compared to GCMs, COSMO-RS depends on an extremely small number of adjustable parameters (the seven basic parameters of eq. (8)-(10), (13) and (14) plus nine τ_{vdW} values) some of which are physically predetermined (Klamt, 1995). COSMO-RS parameters are not specific of functional groups or molecule types. The parameters have to be adjusted for the QM-COSMO method that is used as a basis for the COSMO-RS calculations only. Thus the resulting parameterization is completely general and can be used to predict the properties of almost any imaginable compound mixture or system.

COSMO-RS has been parameterized for BP-RI/COSMO-density functional theory with TZVP basis set which is available in the Turbomole program package (Schäfer et al., 2000). The geometries of all molecules involved in the parameterization as well as the validation of COSMO-RS have been optimized at this level of QM theory. The parameter optimization was done with a dataset of 890 room temperature values of activity coefficients in infinite aqueous dilution, vapor pressure and partition coefficients of water with octanol, hexane, benzene and diethyl ether. The parameterization data set consists of 310 compounds of broad chemical functionality based on the elements H, C, N, O, F, S, Cl, Br and I. The parameter optimization resulted in root mean square (rms) deviations of 0.285 log(γ^∞) units (maximum deviation: 0.451 log(γ^∞) units) for activity coefficients, 0.307 (rms) and 0.566 (max.) log(p) units for vapor pressure, 0.471 (rms) and 0.723 (max.) log(K) units for 1-octanol/water partition coefficients, 0.200 (rms) and 0.374 (max.) log(K) units for hexane/water partition coefficients, 0.160 (rms) and 0.392 (max.) log(K) units for benzene/water partition coefficients and 0.433 (rms) and 0.906 (max.) log(K) units for diethyl ether/water partition coefficients. For the six properties considered in the parameterization an overall rms deviation of 1.8 kJ/mol for the chemical potential differences was found, which corresponds to 0.34 log units for the partition properties. The resulting parameterization was validated with three different test sets: (i) a set of over 1000 activity coefficients of various solutes (which are not contained in the parameterization data set) in various solvents at a range of temperatures (Howard and Meylan, 2000; Schiller, 2000), which resulted in an overall rms deviation

of 0.47 log(γ^{∞}) units and a maximum deviation of 1.10 log(γ^{∞}) units; (ii) a set of 150 Henry law constants (which are not contained in the parameterization data set) at various temperatures (Sander, 2000), which resulted in an overall rms deviation of 0.38 log(k_H) units and a maximum deviation of 0.75 log(k_H) units; (iii) a set of 100 excess Gibbs free energy values of binary mixtures at various temperatures (Kang et al., 2000), which resulted in an overall rms deviation of 205 J/mol and a maximum deviation of 1.2 kJ/mol. The rms deviations of the parameterization and test data give a rough estimate of the errors that must be expected for the prediction of a certain property with COSMO-RS.

The following parameter values were optimized for the Turbomole BP/TZVP QM-COSMO method: The effective contact area for a single independent molecular contact resulted in a_{eff} = 6.25 Å². Considering the area of 45Å² of a water molecule this corresponds to about 7.2 independent neighbours for a water molecule. The electrostatic misfit energy coefficient is α' = 5950 kJ/mol/Å². This agrees reasonably well with the estimate of 8300 kJ/mol/Å², which can be roughly derived from electrostatic considerations (Klamt, 1995). The optimized values of the hydrogen bonding parameters are c_{hb} = 36700 kJ/mol/Å² and σ_{hb} = 0.085 e/Å². The ring correction coefficient ω_{ring} was optimized to 0.89 kJ/mol. Using the reference states 1 mol/mol and 1 bar for the fluid phase and for the gas phase, respectively, η_{gas} was optimized to 21.7 kJ/mol. The coefficient λ_C in the combinatorial part of the chemical potential was optimized to 0.07. The optimized values of the element specific dispersion constants τ_{vdW} are given in Table 1.

2.3 COSMO-RS vs. UNIFAC

Although there are similarities in the basic statistical thermodynamics approach, COSMO-RS and UNIFAC are quite different approaches to the prediction of thermodynamic properties, both with their specific strengths and weaknesses.

Due to their longer history and the numerous contributors GCMs presently are in a very elaborated and sophisticated state and thus also widely accepted as state of the art method in industrial and academic research. Especially UNIFAC has been parameterized carefully to a very large set of experimental data. Yet, the accuracy of COSMO-RS cannot compete with UNIFAC in its core region of parameterization, although generally the quality of the COSMO-RS predictions is only slightly worse.

The basic weakness of GCMs lies in their nature of interpolating molecular structure with groups. It restricts the applicability of the GCMs to systems where group interaction parameters are available and thus to systems for which a significant amount of reliable experimental data is available. COSMO-

RS is not ridden by such restrictions – it is generally applicable to any system that can be calculated by quantum chemistry. This claim also holds for systems where no experimental data is available, such as compounds involving rare functional groups, heterocyclic aromatic compounds or complicated biochemical molecules, unstable systems (e.g isocyanates in water) and even reactive intermediates, transition states, metastable complexes, complexes on surfaces and many more.

A basic assumption of GCMs such as UNIFAC is that each contact between two groups m and n is associated with a specific group interaction energy a_{mn}. This has two important consequences: First, any kind of contact between groups is associated with the same energy. The group interactions do not differentiate between vdW interactions (which are spatially non-specific) and hydrogen bonding (which is strongly directed) – only a physically nondescriptive and averaged energy is provided. Second, intramolecular interactions (e.g. electronic push-pull effects on aromatic rings or intramolecular hydrogen bonds) are completely neglected. In contrast, COSMO-RS theory avoids both of these defects of GCMs: Firstly, the interaction energies of surface patches are specific in that they are summed up from contributions of electrostatic misfit-, vdW and hydrogen bonding interactions. In addition, with the concept of σ-profiles and σ-potentials COSMO-RS allows for a vivid and physically sound interpretation of molecular interactions. Secondly, the concept of molecular surfaces naturally includes all kinds of intramolecular interactions. Thus also isomeric effects are fully taken into account allowing the screening of solvents or entrainers for isomer separation problems (Clausen, 1999).

With one and the same parameterization COSMO-RS is able to predict any equilibrium thermodynamic property at any concentration (including infinite dilution) with approximately the same quality whereas UNIFAC had to be reparameterized to be applicable to infinite dilution properties or partition coefficients (Wienke and Gmehling, 1998).

In original UNIFAC, the temperature dependency of the predicted properties was only poorly described. This was fixed in *mod*-UNIFAC however at the cost of a strong increase in the number of adjustable parameters and thus also of the amount of required experimental data (Weidlich and Gmehling, 1987; Gmehling et al., 1993; Gmehling et al., 1998). The temperature dependency of COSMO-RS is settled in its generic equations and thereby dependent on only a very small number of adjustable parameters which are physically predetermined (Klamt et al., 1998).

It is noteworthy that COSMO-RS allows the prediction of vapor pressures of pure compounds (which UNIFAC can not). Although very often the vapor pressure of pure compounds is known experimentally, the COSMO-RS predictions can be used to check the consistency of the measurements or provide a first estimate if no reliable data is available.

The computational effort of GCMs is very low. COSMO-RS calculations themselves are of comparable speed. The computationally demanding part of a COSMO-RS calculation is the

underlying quantum chemical COSMO calculation. However, the time consuming QM-COSMO computation has to be done only once per compound. Its results can be stored in a database. Subsequent COSMO-RS calculations can be done from the database (compare Figure 4). Recent developments in computer technology as well as in the techniques of quantum chemistry (Grotendorst, 2000) make QM computations increasingly inexpensive and fast.

3. Applications

This section presents a variety of COSMO-RS applications to practical thermodynamic problems, namely the prediction of VLE, LLE, SLE data, partition coefficients and vapor pressures. All of the COSMO-RS calculations have been done with the COSMO*therm* program (Klamt and Eckert, 2001) using the parameter set described in section 2.2. The timings of the COSMO*therm* calculations given below were obtained on Linux-PC running on a single CPU (800 MHz, PentiumIII). The underlying quantum chemical calculations of the molecular COSMO surfaces have been done with the Turbomole program package using BP-RI-density functional theory with TZVP quality basis set (Schäfer et al., 2000). The geometry of all molecules was fully optimized at that level of QM theory. On a single CPU (800 MHz, PentiumIII) timings for the QM-COSMO calculations of the molecules were in the range of less than thirty seconds for water and two hours for octylbenzene. UNIFAC calculations have been done with a UNIFAC program by Sandler (1998), which is based on the fourth revision of the original UNIFAC model (Tiegs et al., 1987). The UNIFAC program of Sandler (1998) can be operated in an interactive way only. Thus no timings can be given for UNIFAC calculations.

3.1 Vapor-Liquid Equilibria

VLE thermodynamic properties are routinely demanded in industrial process design. GCMs are heavily parameterized on a very large set of VLE data of various binary mixtures and thus are able to reproduce many of such data with good to excellent quality. In principle, VLE predictions can be thought of as the "core region" of GCM application. Nevertheless there is a significant amount of systems whose VLEs cannot be predicted properly by GCMs simply due to missing group interaction parameters or inadequate groups. In contrast, COSMO-RS has no restriction in the structure of molecules to be predicted. However, the quality of the predictions that can be expected from COSMO-RS usually is slightly lower than that of GCMs in their core region. On the other hand, due to its generic approach COSMO-RS' error can be expected to lie within a certain range – independent of the compounds or compound classes involved. The following examples try to give an overview over COSMO-RS' capacities (and limitations).

Table 2 shows the excess enthalpies (H^E) and excess Gibbs free energies (G^E) of equimolar compositions of chloroform (1) with compound (2) (compound (2) being acetone, butanone or methanol) at a variety of temperatures. As has been deduced from qualitative consideration of the σ-profiles of acetone and chloroform, the excess energy of this system is strongly negative. The correspondence between COSMO-RS and the experimental values of Gonzalez et al. (1997) is good for G^E (rms deviation 49 J/mol) and satisfactory for H^E (rms deviation 101 J/mol). For all systems the COSMO-RS predictions of G^E and H^E show an increase with temperature that is stronger than in experiment, i.e. the temperature dependency of G^E and H^E is overestimated. UNIFAC predictions of G^E were better than COSMO-RS for the acetone – chloroform and methanol – chloroform systems but considerably worse for butanone – chloroform resulting in an overall rms deviation of 202 J/mol. UNIFAC tends to underestimate the temperature dependency of G^E. Thus UNIFAC predictions of H^E (rms deviation 452 J/mol) are significantly worse than the COSMO-RS predictions. Please note that parts of the experimental data for methanol – chloroform are questionable: G^E and H^E do not increase homogeneously with temperature (as is predicted by COSMO-RS and UNIFAC). Such inconsistencies in experimental data can be detected with the help of predictive methods like COSMO-RS and UNIFAC. The sum of COSMO-RS calculational times for all predictions in Table 2 was 8 s (CPU).

Figures 5 and 6 demonstrate the application of COSMO-RS to the problem of compounds with different isomeric structures at the example of the VLE properties of the three *n*-hexyne isomers mixed with *n*-octane. Figure 5 shows the excess enthalpies (H^E) (A) and excess Gibbs free energies (G^E) (B) of binary mixtures of the three isomeric *n*-hexynes (1) in *n*-octane (2) at T=303.15 K. Figure 6 shows the vapor-liquid composition (*x-y*) diagrams (A) as well as the activity coefficients (B) of the 1/2/3-hexyne (1) - *n*-octane (2) mixtures at T=303.15 K. The vapor mole fractions y_i have been calculated from the ratio of partial and total vapor pressure:

$$y_i = p_i^0 x_i \gamma_i / p_{tot} \tag{16}$$

The total pressures p_{tot} have been obtained from

$$p_{tot} = p_1^0 x_1 \gamma_1 + p_2^0 x_2 \gamma_2 \tag{17}$$

where p_i^0 are the pure compound vapor pressures for compounds *i* (*i* = 1, 2). Experimental p_i^0 values of Boukais-Belaribi et al. (2000) have been used. x_i are the mole fractions of the compounds in the liquid and γ_i are the activity coefficients of the compounds as predicted by COSMO-RS or UNIFAC. In any case ideal behavior of the gas phase has been assumed. Gas phase pressures were not corrected by fugacity coefficients. The COSMO-RS calculation of each binary mixture (30 points of varying composition) took less than 2 s (CPU). The good correspondence between experiment and the COSMO-RS predictions for all given thermodynamic properties is obvious. In addition, COSMO-RS is able to reproduce the qualitative differences between the hexyne isomers, which results from the chemically different environment of the triple bond in 1-hexyne (a hydrogen atom terminating the

triple bond, which can be expected to be slightly acidic) and 2- and 3-hexyne (no terminal hydrogen, thus very similar chemical behavior). Such isomeric effects cannot be easily reproduced by GCMs. Since there is only one group for non-terminating triple bonds, UNIFAC cannot distinguish 2-hexyne from 3-hexyne. In additions, as is illustrated in Figures 5 and 6, UNIFAC predictions were also worse quantitatively. The UNIFAC predictions of excess enthalpies, excess Gibbs free energies and activity coefficients were too small.

Figure 7 demonstrates the qualitatively and quantitatively correct prediction of thermodynamic properties at different temperatures for the mixture of 3-hexyne and n-octane. Figure 7 shows the excess Gibbs free energies (G^E) (A) and the activity coefficients $\ln(\gamma_i)$ (B) of the binary system 3-hexyne (1) - n-octane (2) at three different temperatures between T=263.15 K and T=343.15 K. For all properties correspondence between experiment and COSMO-RS calculations is very good. The temperature dependency of the VLE properties is reproduced correctly. Again, the UNIFAC predictions of G^E and γ_i were too small.

3.2 Liquid-Liquid Equilibria

Figure 8 shows the COSMO-RS prediction for the LLE of the ternary liquid system decane (1) – octylbenzene (2) – sulfolane (3) at temperatures T=323.15 K, T=348.15 K and T=373.15 K in comparison with experimental values from a recent measurement (Kao and Lin, 1999). COSMO-RS' overall calculational time was 15 s (CPU). Because of the lack of appropriate groups for sulfolane, no UNIFAC predictions could be done for this system. The experimental tie lines of the LLE are well met by the COSMO-RS predictions. In addition, the temperature dependency of the LLE is reproduced correctly. Table 3 shows the partition coefficients of octylbenzene (2) between n-alkane (decane, dodecane and tetradecane) (1) and sulfolane (3) at temperatures T=323.15 K, T=348.15 K and T=373.15 K. COSMO-RS' overall calculational time for all partition coefficients in Table 3 was 2 s (CPU). The overall rms error of COSMO-RS predictions is 0.33 log(K) units, which is within the accuracy range expected for partition coefficients (compare section 2.2).

3.3 Solid-Liquid Equilibria (Solubility)

The prediction of SLEs and solubility of solid compounds with COSMO-RS involves an additional complication: COSMO-RS is a theory of liquids, i.e. of an ensemble of disordered molecules. COSMO-RS' predictions of chemical potentials of compounds below their melting point are always

predictions of the supercooled melt. The solid state of a compound is related to the liquid state by its heat of crystallization (Gibbs free energy of fusion ΔG_{fus}). A general expression for solubility reads

$$ln\, x_i^{SOL} = \frac{min(0, \Delta G_{fus})}{RT} - ln\, \gamma_i^{SOL} \tag{18}$$

where x_i^{SOL} is the mole fraction of the solid i dissolved in the solvent phase at saturation; γ_i^{SOL} is the activity coefficient for the solute in solution and can be predicted by COSMO-RS or GCMs. ΔG_{fus} is positive for liquids and eq. (18) reduces to $ln x_i^{SOL} = -ln\gamma_i^{SOL}$. ΔG_{fus} can be estimated by COSMO-RS (Klamt, Eckert and Hornig, 2001). Alternatively, ΔG_{fus} can be modeled quite well by the expression

$$\frac{\Delta G_{fus}}{T} = \Delta S_{fus}\left(1 - \frac{T_m}{T}\right) \tag{19}$$

as has been demonstrated by Frank et al. (1999). T_m is the melting point of solid i and $\Delta S_{fus} = \Delta H_{fus}/T_m$ is the solids entropy of fusion. T_m and ΔH_{fus} are properties of the pure solute and can be found in databases.

The prediction of SLE is demonstrated on three examples given by Frank et al. (1999). ΔG_{fus} was estimated via eq. (19) using the T_m and $\Delta S_{fus}/R$ data given in Frank et al. (1999). This allows an unbiased comparison of COSMO-RS predictions with the UNIFAC and Hansen solubility parameter predictions given by Frank et al. (1999). Figure 9 shows the solubility of acenaphthene over a range of temperatures (T_m=365.95 K and $\Delta S_{fus}/R$=6.88). COSMO-RS' overall calculational time for the solubilities presented in Figure 9 was 2 s (CPU). Deviations from experiment are below 0.3 log(x) units. The temperature dependency of the solubility is predicted correctly. Table 4 shows solubilities of naphthalene in a number of different solvents at T=313.15 K (T_m=353.35 K and $\Delta S_{fus}/R$=6.4). COSMO-RS' overall calculational time for the solubilities presented in Table 4 was 11 s (CPU). The COSMO-RS predictions show an rms error of 0.21 log(x) units. The largest deviations were found for carbon disulfide and acetic acid (0.41 and 0.42 log(x) units, respectively). The errors for the remaining solvents are below 0.3 log(x) units. Frank et al. (1999) did mod-UNIFAC(Do) and Hansen solubility model predictions for the naphthalene-solvent systems, which are also presented in Table 4. UNIFAC predictions show an rms error of 0.08 log(x) units and a maximum deviation of 0.28 log(x) units for ethanol. Thus, in this case the quality of the COSMO-RS predictions is lower than that of UNIFAC. This is no surprise if one considers the simplicity of the given naphthalene – solvent systems: mod-UNIFAC(Do) is very well parameterized for all of the compounds involved. The predictions of the Hansen solubility model (which is an non-predictive extrapolative model for solubility – see Frank et al., 1999) show an rms error of 0.20 log(x) units. The maximum error of the Hansen model (0.76 log(x) units for methanol) is much larger than for COSMO-RS or UNIFAC. Table 5 shows solubilities

of cycloserine in a number of different solvents at T=301.15 K (T_m=420.15 K and $\Delta S_{fus}/R$=6.8). COSMO-RS' overall calculational time for the solubilities presented in Table 4 was 10 s (CPU). Cycloserine is structurally more complex than naphthalene and its solubility is very small in unpolar solvents. Thus it is a much harder test case for predictive methods. The COSMO-RS predictions show an rms error of 0.50 log(x) units. The largest deviations were found for the unpolar solvents benzene, toluene and cyclohexane (0.80, 0.73 and 0.79 log(x) units, respectively). Because of the lack of appropriate groups for Cycloserine, UNIFAC could not be applied to this system. The predictions of the Hansen solubility model show an rms error of 0.68 log(x) units and a maximum error of 1.71 log(x) units for formamide. Although the rms error of the COSMO-RS predictions is quite large, it is much better than for the Hansen model (which also shows much larger scattering of the error). Thus, it can be concluded that COSMO-RS is applicable to the screening of solubility - even though the absolute errors of the prediction are quite large for complex molecules like cycloserine. Other methods of solubility prediction either lead to inferior results (Hansen model) or are not applicable (UNIFAC). However, for simple compounds like naphthalene, UNIFAC predictions were superior to COSMO-RS.

3.4 Vapor Pressure Prediction

Unlike most predictive activity coefficient models COSMO-RS also allows the *a priori* prediction of vapor pressures of liquids for a given compound at arbitrary temperatures. The vapor pressure p^X of a pure compound X is estimated via

$$p^X = exp\left\{-(\mu_{gas}^X - \mu_X^X)/RT\right\}$$ (20)

where μ_{gas}^X is the chemical potential of compound X in the gas phase (eq. (15)). μ_X^X is the chemical potential of the compound in itself (eq. (13)). If the compound is solid at the given temperature ΔG_{fus} has to be added to μ_X^X.

Figure 10 shows the COSMO-RS predictions for the vapor pressures of compounds hexafluoroethane **(A)** and octafluorocyclobutane **(B)** at various temperatures. COSMO-RS' overall calculational time for the vapor pressures presented in Figure 10 was 1.8 s (CPU) for hexafluoroethane and 2.1 s (CPU) for octafluorocyclobutane. The deviations from experiment of <0.5 ln(p) units are well within the accuracy range that can be expected for COSMO-RS predictions of vapor pressures (see section 2.2). The temperature dependency of the vapor pressure is very well met for cyclic compound octafluorocyclobutane, however it is less well met for linear compound hexafluoroethane, although it should be noted that the overall deviations from experiment are not higher than for octafluorocyclobutane.

4. Conclusions

As has been demonstrated in the previous sections COSMO-RS is a promising novel approach for the computational prediction of equilibrium thermodynamic properties of pure compounds (e.g. vapor pressures) and arbitrary mixtures (e.g. VLE properties such as activity coefficients, excess properties, phase diagrams and LLE and SLE properties such as partition coefficients and solubility of liquids and solids) and thus is an alternative and/or supplement to group contribution methods which currently are widely used for such calculations. Like GCMs COSMO-RS is a surface interaction model - however not of groups but of molecular surface charge densities that are provided by molecular quantum chemical COSMO calculations. This leads to the main advantage of COSMO-RS compared to GCMs: COSMO-RS is based on a very small number of adjustable parameters, which are completely independent of any molecular or structural information (i.e. no group interaction parameters). COSMO-RS' parameters are established on a physical basis and depend only on the underlying quantum chemical model. Currently COSMO-RS is parameterized for the elements H, C, N, O, F, S, Cl, Br and I. Molecules with other elements like Si or P can be treated as well although a slightly lower quality of the prediction has to be expected in this case. Thus COSMO-RS is generally applicable to any system of compounds that can be thought of. COSMO-RS calculations are very fast (milliseconds on a modern PC). The underlying quantum chemical COSMO calculations which are calculationally more demanding (but are well done over night on single CPU in most cases) have to be done only once per compound and can subsequently be held in a database. Thus in combination with a large database of solvents COSMO-RS allows for fast and efficient large-scale solvent screening.

References and Notes

Andzelm, J., C. Kölmel and A. Klamt "Incorporation of solvent effects into density functional calculations of molecular energies and geometries" *J. Chem. Phys.* **103**, 9312 (1995).

Baldridge, K. and A. Klamt "First principles implementation of solvent effects without outlying charge error" *J. Chem. Phys.* **106**, 6622 (1997).

Ben-Naim, A., Solvation *Thermodynamics*, Plenum Press, New York (1987).

Bondi, A., *Physical Properties of Molecular Crystals, Liquids and Glasses.*, Wiley, New York (1968).

Boukais-Belaribi, G., B. F. Belaribi, A. Ait-Kaci and J. Jose "Thermodynamics of n-octane + hexynes binary mixtures" *Fluid Phase Equilibria* **167**, 83 (2000).

Clausen, I. *PhD Thesis*, Technische Universität Berlin, Berlin, Germany (1999).

Cramer, C. J. and D. G. Truhlar "Continuum Solvation Models: Classical and Quantum Mechanical Implementations" in *Reviews in Computational Chemistry*, Vol. 6, Lipkowitz, K. B. and D.B. Boyd, Editors., VCH Publishers, New York (1995).

Cramer, C.J. and D. G. Truhlar "Implicit Solvation Models: Equilibria, Spectra and Dynamics" *Chem. Rev.* **99**, 2160 (1999).

Eckert, F. and A. Klamt, COSMO*base*, Version C1.1-Revision 01.01, COSMO*logic* GmbH & Co KG, Leverkusen, Germany (2001).

Flick, E. W., *Industrial Solvents Handbook*, 4th Ed., Noyes Data, Park Ridge (1991).

Frank, T. C., J. R. Downey and S. K. Gupta "Quickly Screen Solvents for Organic Solids" *Chemical Engineering Progress* 41 (December 1999).

Fredenslund, A., R. L. Jones and J. M. Prausnitz "Group-contribution estimation of activity coefficients in nonideal liquid mixtures" *AIChE J.* **21**, 1086 (1975).

Fredenslund, A., J. Gmehling and P. Rasmussen, *Vapor Liquid Equilibria Using UNIFAC*, Elsevier, Amsterdam, The Netherlands (1977).

Frisch M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K.

Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. A. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian 98, Gaussian, Inc., Pittsburgh PA, (2001).

Gessner, W., *Industrial Solvents Handbook*, Marcel Dekker, New York (1996).

Gmehling, J., J. Li and M. Schiller "A Modified UNIFAC Model. 2. Present Parameter Matrix and Results for Different Thermodynamic Properties" *Ind. Eng. Chem. Res*. **32**, 178 (1993).

Gmehling, J., J. Lohmann, A. Jakob, J. Li and R. Joh "A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension" *Ind. Eng. Chem. Res*. **37**, 4876 (1998).

Gmehling, J. "Present status of group-contribution methods for the synthesis and design of chemical processes" *Fluid Phase Equilibria* **144**, 37 (1998).

Gonzalez, J. A., I. G. de la Fuente and J. C. Cobos "Thermodynamics of mixtures with strongly negative deviation from Raoult's law" *J. Chem. Soc., Faraday Trans*. **93**, 3773 (1997).

Grotendorst, J., Editor, *Modern Methods and Algorithms of Quantum Chemistry*, NIC-Proceedings, Jülich, Germany (2000).

Hornig, M. and A. Klamt, in preparation.

Kang, J. W., K. -P. Yoo, H. Y. Kim, H. Lee, D. R. Yang and C. S. Lee, Korea Thermophysical Properties Databank (KDB)**,** Department of Chemical Engineering, Korea University, Seoul, Korea (2000).

Kao, C.-F. and W.-C. Lin "Liquid-liquid equilibria of alkane (C10-C14) + octylbenzene + sulfolane" *Fluid Phase Equilibria* **165**, 67 (1999).

Kao, C.-P. C. and R. N. Miller "Vapor Pressures of Hexafluoroethane and Octafluorocyclobutane" *J. Chem. Eng. Data* **45**, 295 (2000).

Klamt, A. "Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena" *J. Phys. Chem*. **99**, 2224 (1995).

Klamt, A., "COSMO and COSMO-RS" in *Encyclopedia of Computational Chemistry*, Schleyer, P. v. R. Editor, Wiley, New York (1998).

Klamt, A. and F. Eckert "COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids" *Fluid Phase Equilibria* **172**, 43 (2000).

Klamt, A., F. Eckert and M. Hornig "COSMO-RS: a novel view to physiological solvation and partition questions" *J. Computer-Aided Mol. Design* **15**, 355 (2001).

Klamt, A., G. J. P. Krooshof and P. Taylor "The Surface Pair Activity Coefficient Equation: An Alternative to the Quasi-Chemical Approximation in the Statistical Thermodynamics of Liquid Systems", *submitted to AIChE Journal*.

Klamt, A. and F. Eckert, *in preparation*.

Klamt, A. and F. Eckert, COSMO*therm*, Version C1.1-Revision 01.01, COSMO*logic* GmbH & Co KG, Leverkusen, Germany (2001).

Klamt, A., V. Jonas, T. Buerger and J. C. W. Lohrenz "Refinement and Parameterization of COSMO-RS" *J. Phys. Chem. A* **102**, 5074 (1998).

Klamt, A. and G. Schüürmann "COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient" *J. Chem. Soc. Perkin Trans.2, 799 (1993).*

Larsen, B. L., P. Rasmussen and A. Fredenslund "A Modified UNIFAC Group-Contribution Model for the Prediction of Phase Equilibria and Heats of Mixing" *Ind. Eng. Chem. Res*. **26**, 2274 (1987).

Lide, D. R., *CRC Handbook of Chemistry and Physics*, 80th Ed., CRC Press, Boca Raton (2000).

Lin, S.-T. and S. I. Sandler "Infinite Dilution Activity Coefficients from Ab Initio Solvation Calculations" *AIChE J*. **45**, 2606 (1999).

Marcus, Y., *The Properties of Solvents*, Wiley, New York (1998).

Howard, P. and W. Meylan, PHYSPROP database, Syracuse Research Corporation, North Syracuse (2000).

Prausnitz, J. M., R. M. Lichtenthaler and E. Gomez de Azevedo *Molecular Thermodynamics of Fluid-Phase Equilibria*, 3rd Ed., Prentice-Hall, Upper Saddle River (1999).

Reid, R. C., J. M. Prausnitz and B. E. Poling *The Properties of Gases and Liquids*, 4th Ed., McGraw-Hill, New York (1986).

Sander, R., Centre for Atmospheric Chemistry, York University, Ontario, Canada, *private communication* (2000).

Sandler, S. I. *Chemical and Engineering Thermodynamics*, 3rd Ed., Wiley, New York (1998).

Schäfer, A., A. Klamt, D. Sattel, J. C. W. Lohrenz and F. Eckert "COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems" *Phys. Chem. Chem. Phys*. **2**, 2187 (2000).

Schiller, M., DuPont, Wilmington, *private communication* (2000).

Tiegs, D., J. Gmehling, P. Rasmussen and A. Fredenslund "Vapour-Liquid Equilibria by UNIFAC Group Contribution: Revision and Extension. 4." *Ind. Eng. Chem. Res*. **26**, 159 (1987).

Tomasi, J. and M. Persico "Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distribution of the Solvents" *Chem. Rev.* **94**, 2027 (1994).

Walas, S. M. *Phase Equilibrium in Chemical Engineering*, 1st Ed., Butterworth-Heinemann, Boston (1985).

Weidlich, U. and J. Gmehling "A Modified UNIFAC Model. 1. Prediction of VLE, h^E and γ^∞" *Ind. Eng. Chem. Res*. **26**, 1372 (1987).

Wienke, G. and J. Gmehling "Prediction of Octanol-Water Partition Coefficients, Henry Coefficients and Water Solubilities" *Toxicol.Environ. Chem.* **65**, 57 (1998); Erratum: **67**, 275 (1998).

List of symbols

α'	COSMO-RS parameter for electrostatic misfit
ΔS_{fus}	entropy of fusion
γ_i	activity coefficient for species i
γ^C_i	UNIFAC combinatorial term for species i
γ^R_i	UNIFAC residual term for species i
λ_C	COSMO-RS parameter for the combinatorial contribution
μ^{Xi}_S	chemical potential for species $X_{i\ in}$ system S
$\mu^{Xi}_{C,S}$	COSMO-RS combinatorial contribution for species X_i in system S
ν_i	UNIFAC number of groups for species i
ω_{Ring}	COSMO-RS parameter for ring correction
Θ_m	UNIFAC group constitution function for group m
Γ_k	UNIFAC group activity factors for group k
Ψ_k	UNIFAC group interaction term for groups m and n
σ_i	COSMO-RS screening charge density for segment i
σ_{HB}	COSMO-RS parameters for hydrogen bonding
τ_{vdW}	COSMO-RS parameter for van der Waals interaction
a_{eff}	COSMO-RS effective contact area
a_{mn}	UNIFAC group interaction parameter between groups m and n
c_{HB}	COSMO-RS parameter for hydrogen bonding
F_i	UNIFAC surface area of species i
G^E	excess Gibbs free energy
H^E	excess enthalpy
$p^{Xi}(\sigma)$	COSMO-RS sigma profile of a compound X_i
$p_S(\sigma)$	COSMO-RS sigma profile of a system/mixture
Q_k	UNIFAC surface area of group k
R_k	UNIFAC vdW volume of group k
rms	root mean square
R	gas constant
S	system/solvent, either pure or mixture
T	temperature
T_m	melting point temperature
V_i	UNIFAC Volume of species i
vdW	van der Waals term
X_i	chemical compound i considered as solute
x_i	mole fraction of compound i in mixture
X_m	UNIFAC mole fraction of group m
y_i	mole fraction of compound i in the vapor above a mixture

Figure Captions

Figure 1:

COSMO-RS view of surface contact interactions of molecular cavities.

Figure 2:

σ-profiles of the solvents water, acetone, chloroform and hexane.

Figure 3:

σ-potentials of the solvents water, acetone, chloroform and hexane at T=298.15 K.

Figure 4:

Schematic illustration of a COSMO-RS calculation of thermodynamic properties.

Figure 5:

Excess enthalpy **(A)** and excess Gibbs free energy **(B)** of binary mixtures of the hexyne-isomers 1/2/3-hexyne (1) with *n*-octane (2) at T=303.15 K. Filled squares, triangles and circles: excess enthalpies for (1) = 1-hexyne, 2-hexyne and 3-hexyne, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Solid lines: calculated values from COSMO-RS. Dotted lines: calculated values from UNIFAC.

Figure 6:

x-y phase diagram **(A)** and activity coefficients **(B)** of binary mixtures of the hexyne-isomers 1/2/3-hexyne (1) with *n*-octane (2) at T=303.15 K. **(A):** Filled squares, triangles and circles: mole fractions in the liquid (*x*) and gas phase (*y*) of 1-hexyne, 2-hexyne and 3-hexyne, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Solid lines: calculated values from COSMO-RS. Dotted lines: calculated values from UNIFAC. **(B):** Filled squares, triangles and circles: activity coefficient of (1) for (1) = 1-hexyne, 2-hexyne and 3-hexyne, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Empty squares, triangles and circles: activity coefficient of (2) for (1) = 1-hexyne, 2-hexyne and 3-hexyne, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Solid lines: calculated values from COSMO-RS. Dotted lines: calculated values from UNIFAC.

Figure 7:

Excess enthalpy **(A)** and activity coefficients **(B)** of binary mixtures of 3-hexyne (1) with *n*-octane (2). **(A):** Filled squares, triangles and circles: excess enthalpies at temperatures T=263.15 K, T=303.15 K and T=343.15 K, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Solid

lines: calculated values from COSMO-RS. Dotted lines: calculated values from UNIFAC. **(B):** Filled squares, triangles and circles: activity coefficient of (1) at temperatures T=263.15 K, T=303.15 K and T=343.15 K, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Empty squares, triangles and circles: activity coefficient of (2) at temperatures T=263.15 K, T=303.15 K and T=343.15 K, respectively, which are experimental values of Boukais-Belaribi et al. (2000). Solid lines: calculated values from COSMO-RS. Dotted lines: calculated values from UNIFAC.

Figure 8:

Tie lines for the LLE of the ternary system decane (1) – octylbenzene (2) – sulfolane (3). Filled squares, triangles and circles: LLE mole fractions at temperatures T=323.15 K, T=348.15 K and T=373.15 K, respectively, which are experimental values of Kao and Lin (1999). Solid lines: calculated values from COSMO-RS.

Figure 9:

Solubility of acenaphthene in cyclohexane at various temperatures. Filled squares: experimental values of Frank et al. (1999). Solid line: calculated values from COSMO-RS.

Figure 10:

Vapor pressures of **(A)** hexafluoroethane and **(B)** octafluorocyclobutane at various temperatures. Filled squares: experimental values of Kao and Miller (2000). Solid lines: calculated values from COSMO-RS.

Table 1:

Element specific COSMO-RS parameters.

Element	τ_{vdW} [kJ/mol/Å²]
H	0.0361
C	0.0401
N	0.0181
O	0.0189
F	0.0265
S	0.0510
Cl	0.0514
Br	0.0550
I	0.0580

Table 2:

Molar excess functions, Gibbs free energies (G^E) and enthalpies (H^E) of chloroform (1) – compound (2) binary mixtures at equimolar composition and various temperatures in [J/mol]. Experimental values: Gonzalez et al. (1997).

Compound(2)	T [K]	G^E Experiment	G^E COSMO-RS	G^E UNIFAC	H^E Experiment	H^E COSMO-RS	H^E UNIFAC
Acetone	283.15		-705	-654	-1972	-1871	-1607
Acetone	287.15		-689	-641	-2173	-1844	-1581
Acetone	298.15	-605	-646	-606	-1907[1]	-1770	-1526
Acetone	303.15	-580	-628	-591	-1880	-1738	-1484
Acetone	308.32	-552	-610	-576		-1705	-1460
Acetone	313.15	-584	-593	-563	-1856	-1674	-1429
Acetone	323.15	-546	-560	-536	-1745[2]	-1612	-1376
Acetone	333.15		-529	-511	-1718	-1552	-1325
Acetone	343.15		-499	-487	-1695	-1494	-1278
Butanone	303.15	-727	-683	-515		-1743	-1168
Butanone	308.15		-666	-504	-2103	-1712	-1142
Butanone	318.15	-709	-634	-484		-1652	-1097
Butanone	328.15	-646	-603	-466		-1594	-1055
Methanol	293.15	781	717	765		-376	-83
Methanol	298.15	757	735	780	-300	-340	-61
Methanol	303.15	811	753	793	-66	-303	-40
Methanol	308.15		770	807	-207	-267	-17
Methanol	313.15	841	786	820		-230	3
Methanol	323.15	873	818	846		-155	44

[1] Experimental data varies -1907±18 kJ/mol (Gonzalez et al., 1997).
[2] Experimental data varies -1745±5 kJ/mol (Gonzalez et al., 1997).

Table 3:

Partition coefficients log(P) of solute octylbenzene (2) between *n*-alkanes (1) and sulfolane (3) at various temperatures. Experimental values: Kao and Lin (1999).

n-Alkane (1)	T [K]	Experiment	COSMO-RS
Decane	323.15	-1.51	-0.95
	348.15	-1.31	-0.91
	373.15	-1.21	-0.87
Dodecane	323.15	-1.51	-1.10
	348.15	-1.37	-1.06
	373.15	-1.24	-1.00
Tetradecane	323.15	-1.51	-1.26
	348.15	-1.33	-1.20
	373.15	-1.19	-1.15

Table 4:

Solubility of naphthalene (decadic logarithm of the mole fraction $\log(x^{SOL})$) in various solvents at T=313.15 K. Experimental values, UNIFAC and Hansen estimates: Frank et al. (1999).

Solvent	Experiment	COSMO-RS	UNIFAC	Hansen
Carbon disulfide	-0.31	-0.72	-0.39	-0.37
Acetone	-0.42	-0.55	-0.45	-0.43
Benzene	-0.37	-0.57	-0.35	-0.40
Ethylene dichloride	-0.35	-0.56	-0.37	-0.36
Toluene	-0.37	-0.57	-0.36	-0.38
1,1-Dichloroethane	-0.36	-0.57	-0.34	-0.39
Chloroform	-0.33	-0.55	-0.33	-0.35
Chlorobenzene	-0.35	-0.57	-0.37	-0.36
Nitrobenzene	-0.36	-0.60	-0.47	-0.39
Aniline	-0.51	-0.82	-0.55	-0.42
1,1-Dibromoethane	-0.34	-0.57	-0.37	-0.37
Tetrachloromethane	-0.40	-0.69	-0.39	-0.42
1,2-Dibromoethane	-0.36	-0.58	-0.38	-0.41
Hexane	-0.65	-0.76	-0.59	-0.65
Cyclohexanol	-0.63	-0.88	-0.81	-0.63
Acetic acid	-0.93	-1.36	-0.91	-0.53
1-Butanol	-0.94	-0.98	-1.03	-1.03
1-Propanol	-1.02	-1.04	-1.18	-1.18
2-Butanol	-0.95	-0.91	-1.03	-0.81
Ethanol	-1.14	-1.14	-1.37	-1.52
Tert-Butanol	-0.99	-0.86	-1.10	-1.07
Methanol	-1.36	-1.41	-1.31	-2.12
Iso-Butanol	-1.03	-0.97	-1.03	-1.18
2-Propanol	-1.12	-0.97	-1.15	-1.12

Table 5:

Solubility of cycloserine (decadic logarithm of the mole fraction $\log(x^{SOL})$) in various solvents at T=301.15 K. Experimental values and Hansen estimates: Frank et al. (1999).

| Solvent	Experiment	COSMO-RS	Hansen
Water	>-2.45	-2.60	>-1.37
Methanol	-3.11	-2.62	-2.76
Formamide	-3.20	-2.49	-1.49
Ethylene glycol	-3.08	-2.66	-2.91
Acetone	-3.21	-2.84	-3.17
Benzyl alcohol	-2.99	-3.11	-2.63
Pyridine	-3.15	-2.58	-2.57
Diethyl ether	-3.33	-3.59	-4.84
Isopropanol	-3.42	-2.99	-3.42
Ethanol	-3.59	-2.87	-2.96
Methyl ethyl ketone	-3.40	-3.08	-3.40
1,4-Dioxane	-3.42	-2.84	-3.60
Ethyl acetate	-3.47	-3.54	-3.81
Isoamyl alcohol	-3.48	-3.36	-3.59
Isoamyl acetate	-3.99	-4.07	-4.35
Benzene	-4.38	-5.19	-4.72
Chloroform	-4.28	-4.90	-3.73
Toluene	-4.50	-5.23	-4.50
Cyclohexane	-4.54	-5.33	-5.38
Carbon Disulfide	Trace	-8.23	-4.43
Isooctane	Trace	-7.79	-6.25

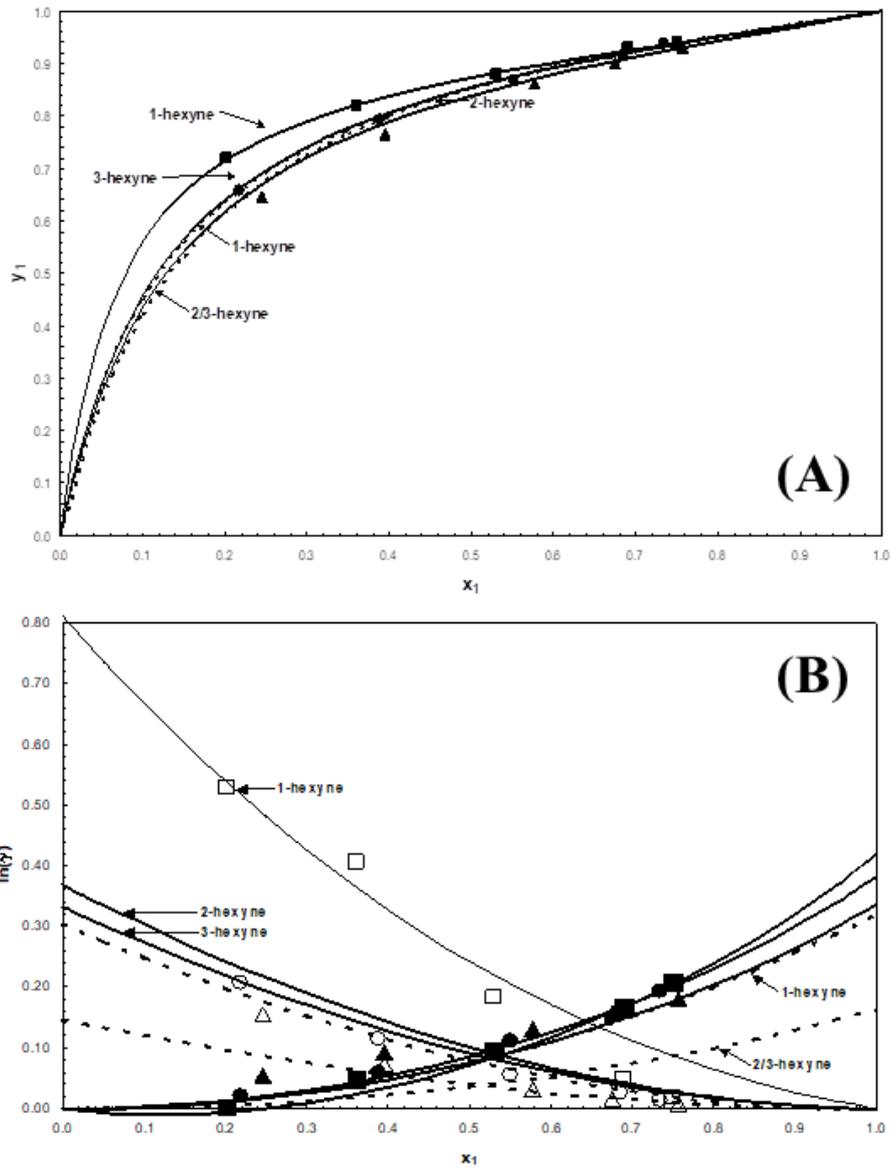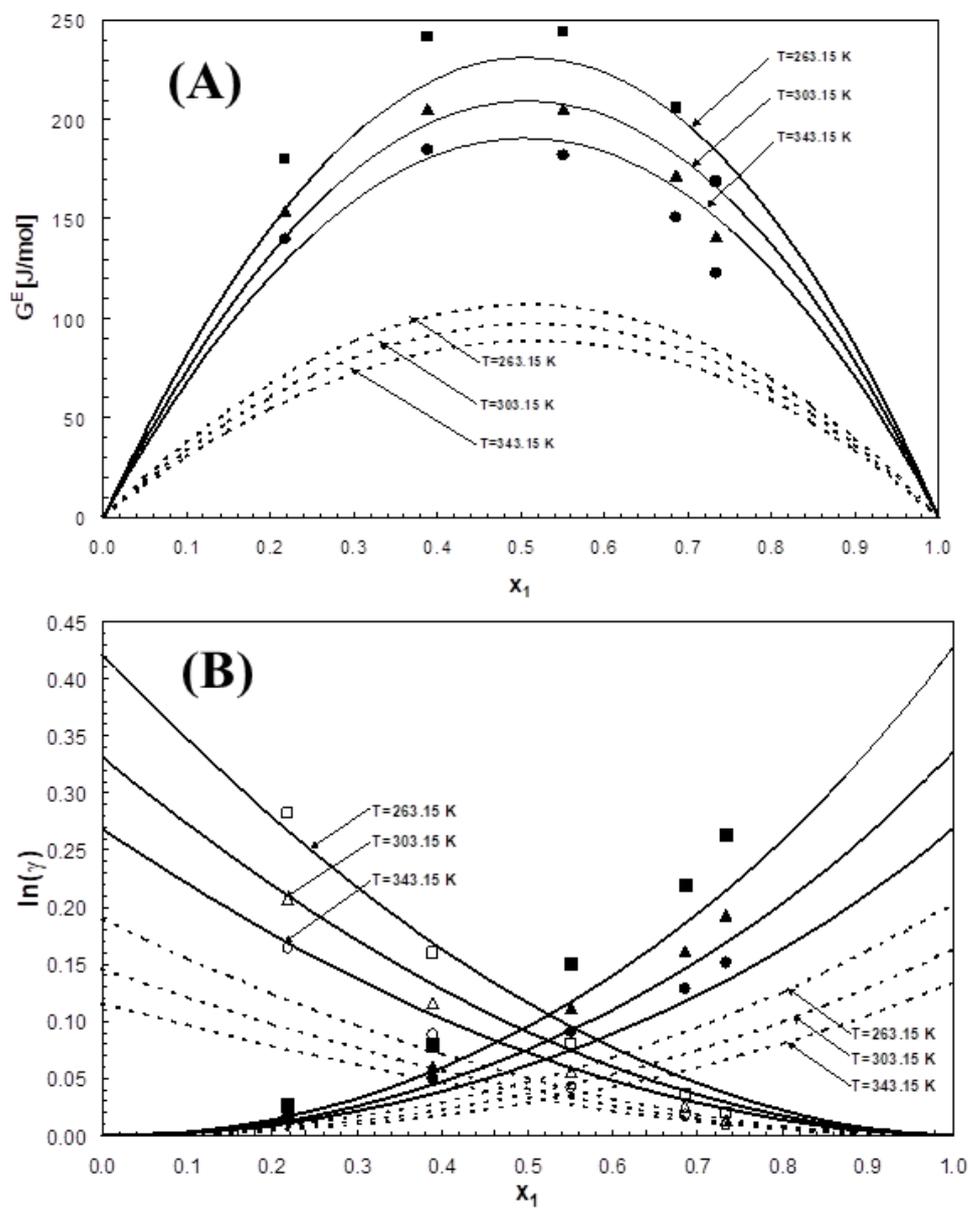**Figures**

Figure 1
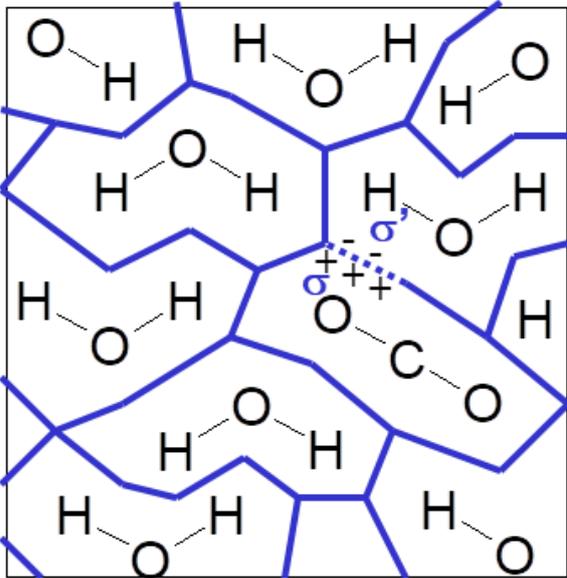
Figure 2
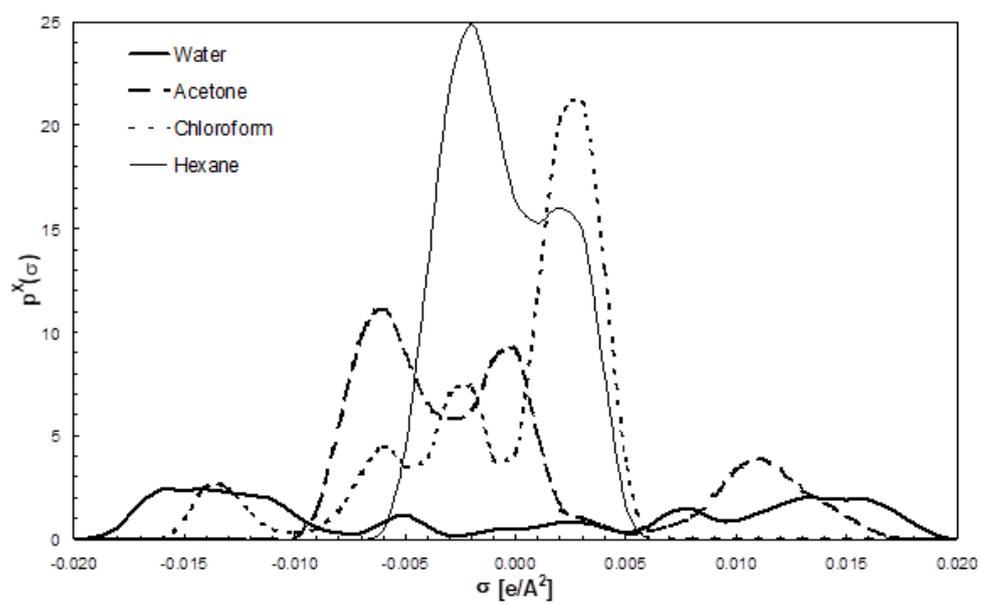
Figure 3
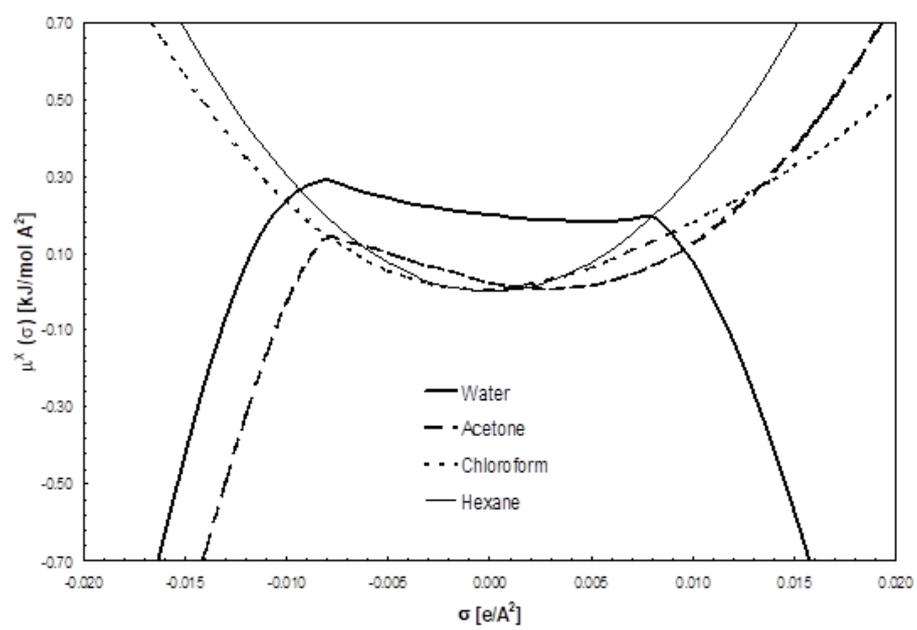
Figure 4
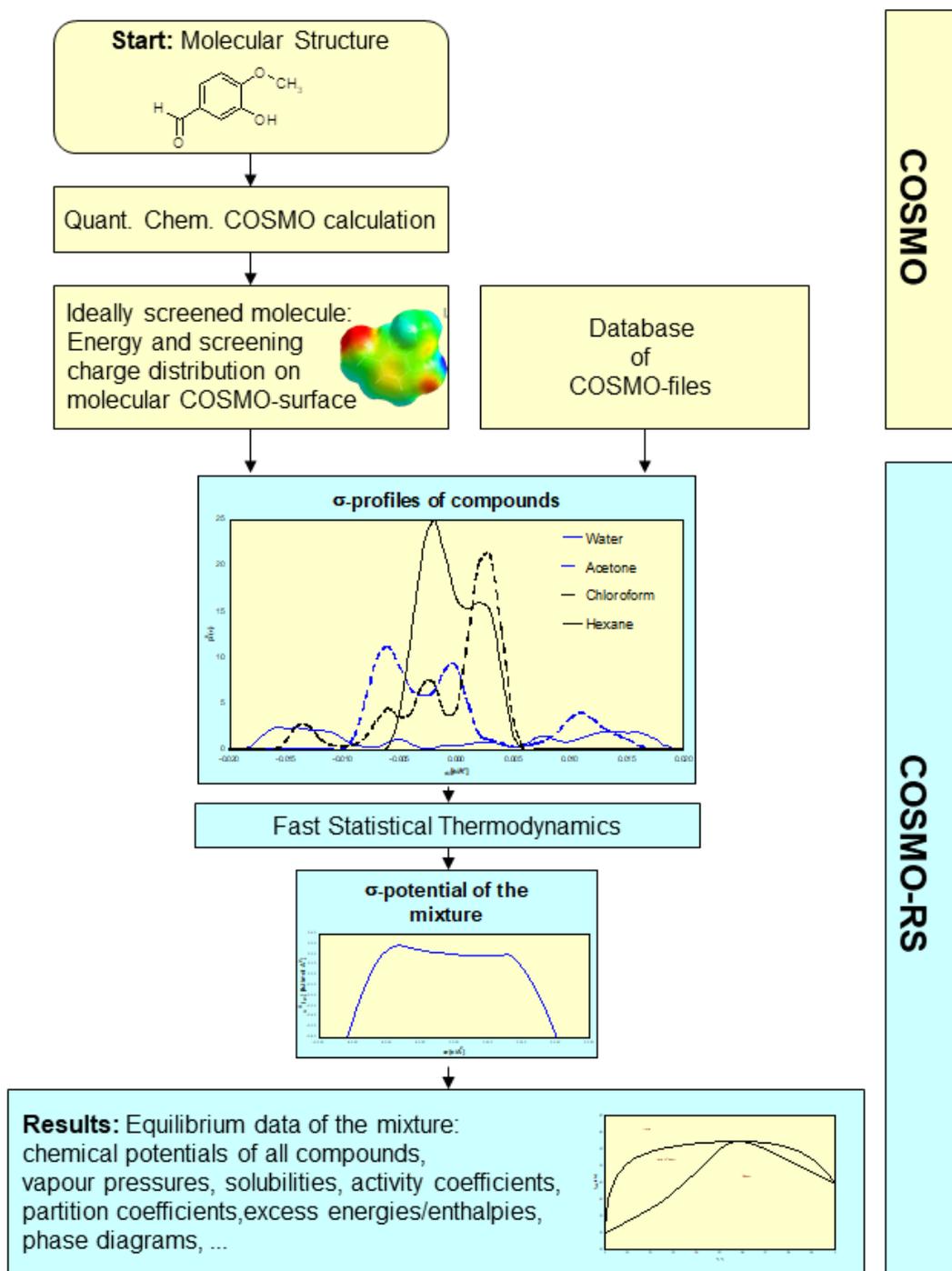
Figure 5
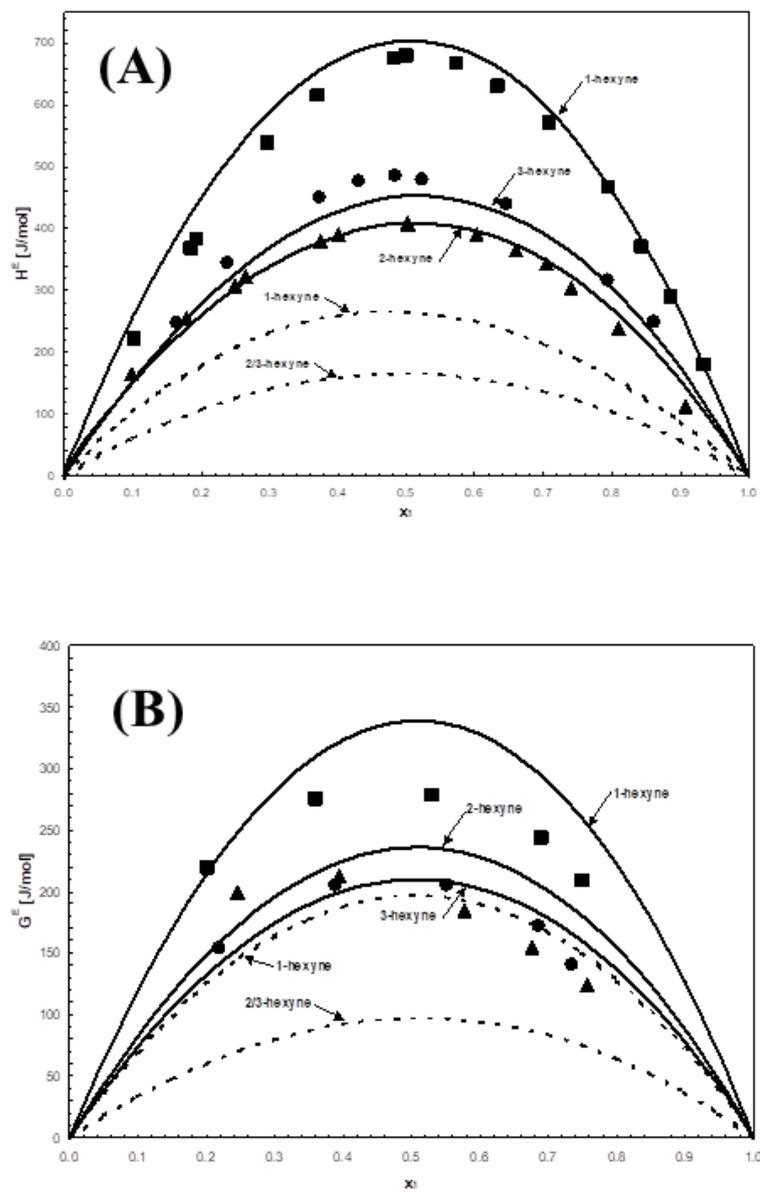
Figure 6

Figure 7

Figure 8

Figure 9

Figure 10