Book published by Royal Society of Chemistry, 2007, ISBN-10: 0-85404-372-1,
ISBN-13: 978-0-85404-372-9, Registered Charity Number 207890, www.rsc.org

Prediction of Solubility with COSMO-RS

Frank Eckert

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

Phone: +49-2171-731682, Email: eckert@cosmologic.de

1 Introduction

Solubility is one of the most fundamental processes in chemistry and biology. The understanding and the prediction of the thermodynamic properties of neutral compounds and of salts in water and organic solvents are of crucial importance in many areas of chemistry and biochemistry. Most physiological and technical processes occur in solution, and the choice of solvent or solvent mixture is very important to the rates and outcome of a process[1]. When considered at thermodynamic equilibrium, solubility can be expressed in terms of the free energy (or chemical potential) difference of a compound X dissolved in a solvent phase S and the pure compound X. The theoretical calculation of a chemical potential is complicated, because we do not only have to calculate the interaction energy of a solute X in a solvent S, but we also have to take into account the change in the entropy and in the interactions of the solvent molecules caused by the solute molecule X.

Molecular dynamics (MD) or Monte Carlo (MC) methods are the most straight-forward procedures to compute the change in free energy of an ensemble of solvent molecules S by insertion of a solute X. To get reasonably accurate values one has to consider a very large ensemble of solvent molecules. Nowadays such calculations can be done routinely based on force-field pair-potentials[2], but one should be aware that all interactions, which are of quantum-chemical nature, are described by a classical force-field approximation. Unfortunately these calculations are rather time-consuming. Jorgensen and Duffy introduced a shortcut of the MD/MC approach, which is based on averaged interaction descriptors derived from rapid simulations in reference solvents and combined with a Quantitative-Structure-Property-Relationship (QSPR) analysis with respect to the solubility[3].

In contrast, the computationally fastest, but chemically least detailed approach to the estimation of partition coefficients is the fragment- or group-based increment approach, also

known as linear free energy relationship (LFER) approach. The basic assumption of these methods is that the change in free energy of a solute X between solvents S and S' can be split into independent contributions of the chemical groups of X. Thus the logarithmic solubility becomes a sum of group contributions, which can be fitted by linear regression from a sufficiently large set of experimental data. The LFER approach has been applied successfully to the prediction of partition properties such as octanol-water partition coefficients[4]. A disadvantage however, is that a data set of several thousand partition coefficients is required in order to fit the large number of group-parameters. Moreover, while the chemical potential of a solute X diluted in a solvent S can be approximated quite well by a sum of fragment contributions, the chemical potential of a pure compound X that acts as both solute and solvent, is not readily available. In the pure compound the assumption of linear additivity of free energy contributions of structural fragments to a target molecule is not valid, because the addition of a certain fragment can either increase or decrease the solubility of a compound, depending on the remainder of the compound X. Thus solubility, being a strongly non-linear property, can hardly be expressed by a linear regression analysis, unless one of the descriptors does include most of the non-linear behaviour.

A different type of group contribution models (GCMs) such as UNIFAC[5], in which solute and solvent are represented by groups, is widely used in chemical engineering applications. In such models the chemical potentials of the compounds are derived from an approximate statistical thermodynamic treatment of pair-wise interacting surface pieces. For each pair of functional groups an interaction parameter has to be fitted to experimental thermodynamic data in a _ non-linear fitting procedure. The advantage of this approach, compared with the LFER approach, is its ability to treat any solvent or solvent mixture as well as complete binary phase diagrams, provided the interaction parameters for all groups involved in the system are known. A disadvantage is the fact that in GCMs the number of required

parameters involved, is related to the square of the number of groups. For the LFER method, however, the number of parameters is linearly related to the number of groups. Therefore, UNIFAC can afford a much less detailed definition of groups compared to typical LFER methods. Nevertheless, GCMs are of limited applicability to typical solid drug-like compounds, because they are not capable of treating heterocycles or multifunctional aromatic rings.

A different approach is the explicit representation of the solute combined with a continuum representation of the solvent. Most of these continuum solvation models[6,7] (CSMs) concentrate on the electrostatic behaviour of the solvent. Either by solution of the dielectric boundary conditions or by solution of the Poisson-Boltzmann equations (both of which represent the same physics in non-ionic solvents) the solute is treated as if it is embedded in a dielectric medium. Usually the macroscopic dielectric constant of the solvent is used. The conductor-like screening model COSMO[8] is a model of this type, which by a slight approximation achieves superior efficiency and robustness compared with others. The advantage of CSMs is that the solute can be treated with great rigor, typically at a quantum chemical level. If supplemented with surface specific descriptors characterising dispersive interactions, cavitation energies, and other non-electrostatic contributions, the results of such CSMs appear to be capable of describing liquid partition properties and solubility[6].

To summarise, there are several approaches to the calculation of free energies of molecules in solution, each of them covering different aspects of the problem. Despite the usually assumed picture of pair-wise and distance-dependent interactions of atoms, the relative success of MD/MC derived interaction parameters, of group interaction models, and of surface parameter supplemented CSMs suggests that many aspects of free energies of molecules in solution can be as well or even better described by a model of surface interactions without explicit knowledge of the atom positions of the solvent. The COSMO-RS method is a model

based on such an assumption, combining the advantages of quantum chemically based CSMs and sound statistical thermodynamics. The method is described in the following sections and is applied to problems of solubility predictions.

2 COSMO-RS

COSMO-RS is a predictive method for the thermodynamic properties of fluids and liquid mixtures that combines a statistical thermodynamic approach with a quantum chemistry method. The theory of COSMO-RS has been described in detail in several articles[9,10,11,12,13]. Therefore only a short survey of the basic concept will be given here.

The starting point for any COSMO-RS calculation is a molecule X in its ideally screened state. This state can be calculated with reasonable effort by a quantum chemical method, the Conductor-like Screening Model COSMO[8], which is an efficient variant of dielectric continuum solvation (DCSM) methods[6,7]. In the COSMO model, a solute molecule is calculated in a virtual conductor environment. In such an environment, the solute molecule induces a polarization charge density σ on the interface between the molecule and the conductor, that is, on the molecular surface. These charges act back on the solute and thus generate a polarized electron density. During the quantum chemical self-consistency algorithm, the solute molecule is thus converged to its energetically optimal state in a conductor with respect to its electron density, including geometry optimisation. The quantum chemical calculation has to be performed once for each molecule of interest. The resulting polarization charge densities on the molecular surface σ are _ good local descriptors of the molecular surface polarity and can be stored in a database. The σ values from the COSMO calculation allow one to extend the model toward “Real Solvents”, which results in the COSMO-RS method. In COSMO-RS, the deviations of a real solvent, compared to an ideal

conductor are taken into account in a model of pair-wise interacting molecular surfaces. For this purpose, the three-dimensional polarization density distribution on the surface of each molecule X is converted into a distribution function, the so-called σ-profile $p^X(\sigma)$, which gives the relative amount of surface with polarity σ on the surface of the molecule. The σ -profile for the entire solvent of interest S, which might be a mixture of several compounds, $p_S(\sigma)$, can be built by adding the $p^X(\sigma)$ values of the components weighted by their mole fractions x_i in the mixture. Now, electrostatic energy differences and hydrogen bonding energies are quantified as functions of the local COSMO polarization charge densities σ and σ' of the two interacting surface pieces. The chemical potential differences arising from these interactions are evaluated using an exact statistical thermodynamic algorithm for independently pair-wise interacting surfaces, which takes into account local deviations from dielectric behaviour as well as hydrogen bonding. In this approach all information about solutes and solvents is extracted from initial QC-COSMO calculations, and only very few parameters have been adjusted to experimental values of partition coefficients and vapour pressures of a wide range of neutral organic compounds. COSMO-RS is capable of predicting partition coefficients, vapour pressures, and solvation free energies of neutral compounds with an error of 0.3 log-units (rms) and better. This corresponds to an accuracy of about 1.5 kJ/mol for large chemical potential differences like those typically involved in octanol-water partition coefficients or in water solubility and slightly less than a factor 2 for equilibrium constants at room temperature. A lot of experience has been gathered during the past years about COSMO-RS' surprising ability to predict thermodynamic properties of mixtures [13].

3 Computational Details

For all molecular species involved, the standard two step procedure for COSMO-RS calculations has been applied:

1) Quantum chemical COSMO calculations: These involve setting up initial molecular geometries and determination of the lowest energy conformations. For each compound a molecular dynamics (MD) calculation has been done with the molecular modelling program package Alchemy[14]. The MM3 force field[15] has been used to obtain the potential energy during the MD calculation, using an overall MD run time of 5 ps, a time step of 0.001 ps and a initial temperature of 293 K. From the geometries created by the MD calculation up to five significant lowest energy conformations have been picked for each molecule. Special care has been taken in choosing conformations of molecules which are able to build internal hydrogen bonds, since the polarization charge densities σ computed in the subsequent QC-COSMO calculations (and thus also COSMO-RS' chemical potentials μ_S^X) critically depend upon the correct representation of such hydrogen bonds. Subsequently the geometry of the chosen conformations has been optimised with the Turbomole quantum chemistry program package[16,17,18] using the B-P density functional[19,20] with TZVP quality basis set and the RI approximation[21,22]. During these calculations the COSMO continuum solvation model was applied in the conductor limit ($\varepsilon = \infty$). Element-specific default radii from the COSMO-RS parameterisations have been used for the COSMO cavity construction[12,13]. Such calculations end up with the self-consistent state of the solute in the presence of a virtual conductor that surrounds the solute outside the cavity.

2) COSMO-RS calculations have been done using the COSMO*therm* program[23]. If more than one conformations were considered to be potentially relevant for the neutral or ionic form of a compound, several conformations have been calculated in step 1 and a thermodynamic

Boltzmann average over the total Gibbs free energies of the conformers was consistently calculated by the COSMO*therm* program in step 2. Details on the COSMO-RS calculation method and all COSMO-RS parameters used are given in Reference 9.

4 Solubility

Considering solubility in thermodynamic equilibrium, the quantity required for its calculation is the chemical potential μ_S^X of a compound X in a solvent S, at a given temperature T and dilution. Using the pseudo-chemical potential $\mu^*_S{}^X$ according to Ben-Naim[24]:

$$\mu^*_S^X = \mu_S^X - kT \ln x_S^X \tag{1}$$

where x_S^X is the molar concentration of compound X in solvent S, the equilibrium condition of equal chemical potentials of X in two phases S and S' reads:

$$\mu^*_S^X - kT \ln x_S^X = \mu^*_{S'}^X - kT \ln x_{S'}^X \tag{2}$$

Thus the solubility S_S^X of a liquid compound X in a solvent S is related to the difference $\Delta_S^X = \mu^*_S{}^X - \mu^*_X{}^X$ of the pseudo chemical potentials of X in solvent S and of pure compound X. If S_S^X is sufficiently small so that the solvent behavior of the X-saturated solvent S is not significantly influenced by the solute X (infinite dilution of X in S), then the decadic logarithm of the solubility is given by:

$$\log S_S^X = \log\left(\frac{MW_X \rho_S}{MW_S}\right) - \frac{1}{kT \ln(10)} \Delta_S^X \tag{3}$$

Note that in the case of high solubility (solubility greater than 10 mol%), Equation (3) becomes approximate and the true solubility would have to be derived from a detailed search

for a thermodynamic equilibrium of a solvent-rich and a solute-rich phase. However, if the zeroth order $S_S^{X(0)}$ as initially provided by Equation (3) using infinite dilution of X in S, is re-substituted into the solubility calculation via $\Delta_S^{X(1)} = \mu^*_{S[x(0)]}{}^X - \mu^*_X{}^X$, a better approximation for S_S^X is achieved. In other words, the solute pseudo chemical potential $\mu^*_{S[x(0)]}{}^X$ is computed for the solvent-solute mixture with the finite mole fraction of X in S that was predicted by the zeroth order $S_S^{X(0)}$. Now, using Equation (3) with the new $\mu^*_{S[x(0)]}{}^X$ and the resulting $\Delta_S^{X(1)}$ values, an improved solubility $S_S^{X(1)}$ is computed. This value again can be used to computed further refined $\mu^*_{S[x(1)]}{}^X$ and again re-substituted into Equation (3), a better guess for S_S^X can be achieved. This procedure can be repeated until the computed value of S_S^X is constant. This iterative refinement procedure is implemented in the COSMO*therm* program[23] and allows the accurate prediction of solubility values even for cases of high solubility (solubility up to 50 mol%). Thus except for rare cases of very high solubility, a complicated search for a multiphase thermodynamic equilibrium of a solvent-rich and a solute-rich phase is not necessary, but instead Equation (3) and the iterative refinement procedure can be used. All of the following examples were calculated with an iteratively refined Equation (3).

Since the molecular weights MW and the solvent density usually ρ are known, Equation (3) is sufficient for the prediction of the solubility of compounds which are liquid at room-temperature. Unfortunately most drugs are solid at room temperature. Since the solid state of a compound X is related to its liquid state by the free energy difference $\Delta G_{fus}{}^X$ which is negative in the case of solids, a more general expression for solubility reads:

$$\log S_S^X = \log\left(\frac{MW_X \rho_S}{MW_S}\right) + \frac{1}{kT\ln(10)}\left[-\Delta_S^X + \min(0, \Delta G_{fus}^X)\right] \qquad (4)$$

Since for liquids $\Delta G_{fus}{}^X$ is positive, Equation (4) reduces to Equation (3) in this case. For the precise calculation of $\Delta G_{fus}{}^X$ it is necessary to evaluate the free energy of a molecule of

compound X in its crystal, i.e. the crystal structure has to be known. In general, crystal structure prediction for complex molecules has to be considered as an unsolved problem[25]. Thus there is no viable way to a fundamental model. However, typically ΔG_{fus}^{X} is small compared to Δ_S^X. Hence it is reasonable to use Δ_S^X of the liquid as a fundamental input for the calculation of $\log S_S^X$ and to search for some plausible empirical approximation for ΔG_{fus}^{X}.

In a study on the aqueous solubility of 150 drug-like organic molecules taken from Reference 3, we found that a simple correlation of $\log S_S^X$ vs. Δ_S^X (as computed by COSMO-RS) yields a correlation coefficient of $r^2 = 0.65$ and a rms-deviation of 1.2 log-units. The slope in this regression is close to the theoretical expectation. This clearly show the great significance of the pseudo chemical potentials as calculated by COSMO-RS. In a second step the theoretical liquid solubility values of $\Delta_S^X/kT\ln(10)$ were subtracted from the experimental values of solubility $\log S_S^X$ in order to obtain reasonable data values for ΔG_{fus}^{X}. Now it was possible to find a simple linear QSPR expression for ΔG_{fus}^{X} based on molecular descriptors provided by COSMO-RS[26]. Thus we derived the equation:

$$\Delta G_{fus}^{X} = 12.2\, V^X + 0.54\,\mu^{*X}_{water} - 0.76\, N^X_{ringatom} \qquad (5)$$

where the units for ΔG_{fus}^{X} and $\mu^*_{water}{}^X$ are [kJ/mol] and for V^X it is [nm^3]. The descriptor V^X is the COSMO volume. Descriptor $\mu^*_{water}{}^X$, the pseudo chemical potential of solute X in solvent water at infinite dilution, is a combined measure of the solutes polarity and hydrogen bonding properties. Hence this descriptor also appears in the solubility calculation of arbitrary non-aqueous solvents S. The number of ring atoms $N^X_{ringatom}$ acts as a descriptor of molecular rigidity. In Equation (5) the regression constant c_0 is omitted, because it was found to be insignificant for this regression. Equation (5) is applicable to room-temperature solubilities of a wide range of solid organic solutes in arbitrary solvents S. The correlation of the experimental aqueous solubilities $\log S_{water}^X$ of the 150 compounds taken from Reference 3

with COSMO-RS properties via Equation (4) and (5) yielded a correlation coefficient of r^2 = 0.90 and a rms deviation of 0.66 $\log(x_{water}^X)$ units[26]. The COSMO-RSol solubility model thus defined was verified using a data set of aqueous solubility values for 548 pesticide compounds, yielding a standard deviation of s = 0.61 $\log(x_{water}^X)$ units[26]. Originally, COSMO-RSol was developed as a tool for the prediction of the solubility of pure, neutral and undissociated drug-like compounds in water[26]. But its applicability is not restricted to these classes of solutes and solvents. Corrections for dissociation or protonation can be trivially made for compounds with known pK-values. If the dissociation constant is not known from experiment, it can be calculated routinely by COSMO-RS, for acids[27] as well as for bases[28]. The application of COSMO-RSol to the solubility of non-neutral compounds is demonstrated in section 5. Although some applications of COSMO-RSol have been reported[29,30], the method has not been systematically verified, using a large data set of solubilities in non-aqueous solvents. In the recent years Acree[31] has built up a large data source for experimental room temperature solubilities of complex organic compounds in various solvents. Acree's set of 706 room temperature solubilities of 21 solutes in a variety of solvents ranging from non-polar alkanes to strongly polar alcohols, amides, carbon acids, and water has been predicted by the COSMO-RSol method as outlined above and is given in Figure 1. Note that a value of $\log(x_S^X) = 0$ means that arbitrary miscibility of solute and solvent was predicted. The 706 solubilities were predicted with an overall rms deviation of 0.74 $\log(x_S^X)$ units[32]. If experimental data for ΔG^X_{fus} as recommended by NIST[33] are used in Equation 4, the rms error of the predictions reduces to 0.43 $\log(x_S^X)$ units[32]. This is well within the expected error ratio of COSMO-RSol and corroborates the broad scope and general applicability of the method. Note, that a major part of the error of the full COSMO-RSol predictions (using the ΔG^X_{fus} estimate) is caused by a single solute, 4-nitrobenzoic acid, where the ΔG^X_{fus} value is underestimated strongly by COSMO-RSol-QSPR. If the 29 data values for 4-nitrobenzoic acid are removed from the data set, the rms error of the full COSMO-RSol predictions

reduces to 0.64 $\log(x_S^X)$ units, which is even below the rms error obtained for the COSMO-RSol fitted data set[26].

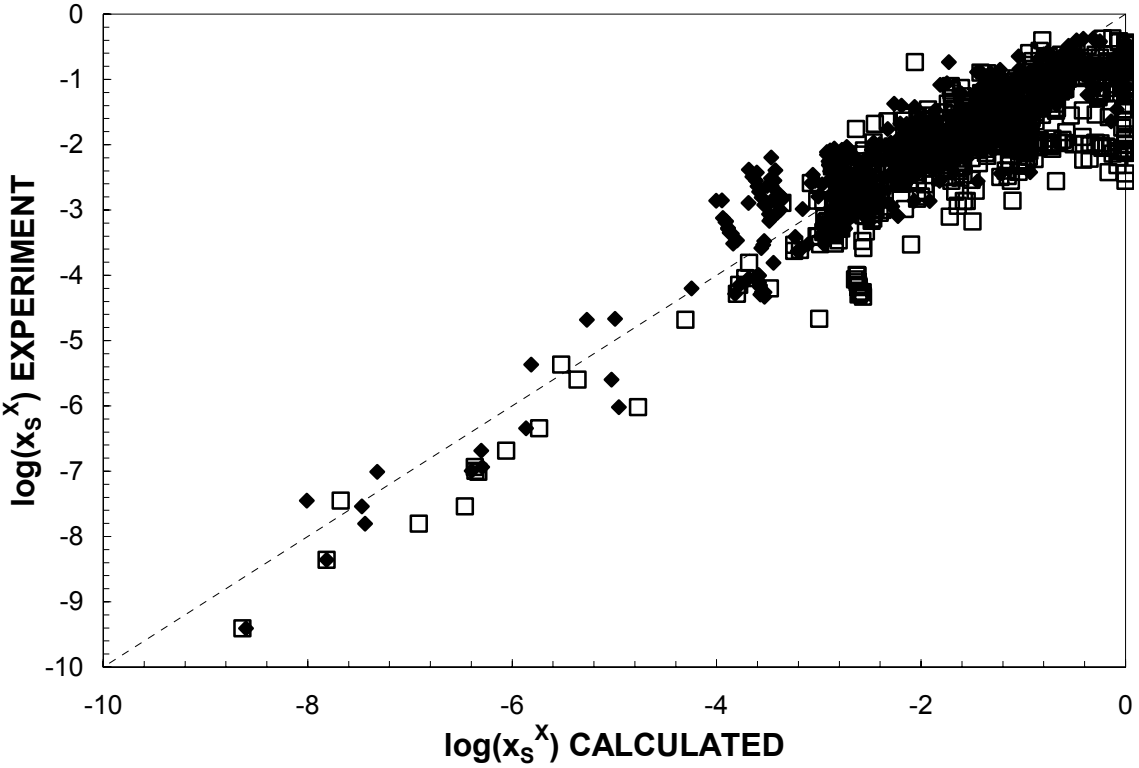

Figure 1 Experimental[31] vs. calculated solubility $\log(x_S^X)$ of organic compounds in different solvents[32]. Filled rhombus: Predictions using ΔG^X_{fus} data estimated by COSMO-RSol. Open squares: Predictions using experimental ΔG^X_{fus} data.

Cinchona alkaloids are an example for compounds that display an interesting solubility behaviour: Their solubilities in solvents of different polarity vary by 5-6 orders of magnitude and they show non-trivial behavior in mixed solvents composed of water and organic solvents[34]. For solute cinchonidine (CAS-RN: 485-71-2) high solubility in solvent 1,4-dioxane is reported, which increases if a small amount of water is added to the solvent[34]. If the

water fraction in the solvent mixture is increased further, the solubility decreases and finally drops to –5.91 $\log(x_S^X)$ units for pure water. This non-trivial behaviour is predicted by COSMO-RSol, qualitatively as well as quantitatively (see Figure 2) achieving an rms error of 0.40 $\log(x_S^X)$ units, if the experimental cinchonidine ΔG^X_{fus} was used[32].

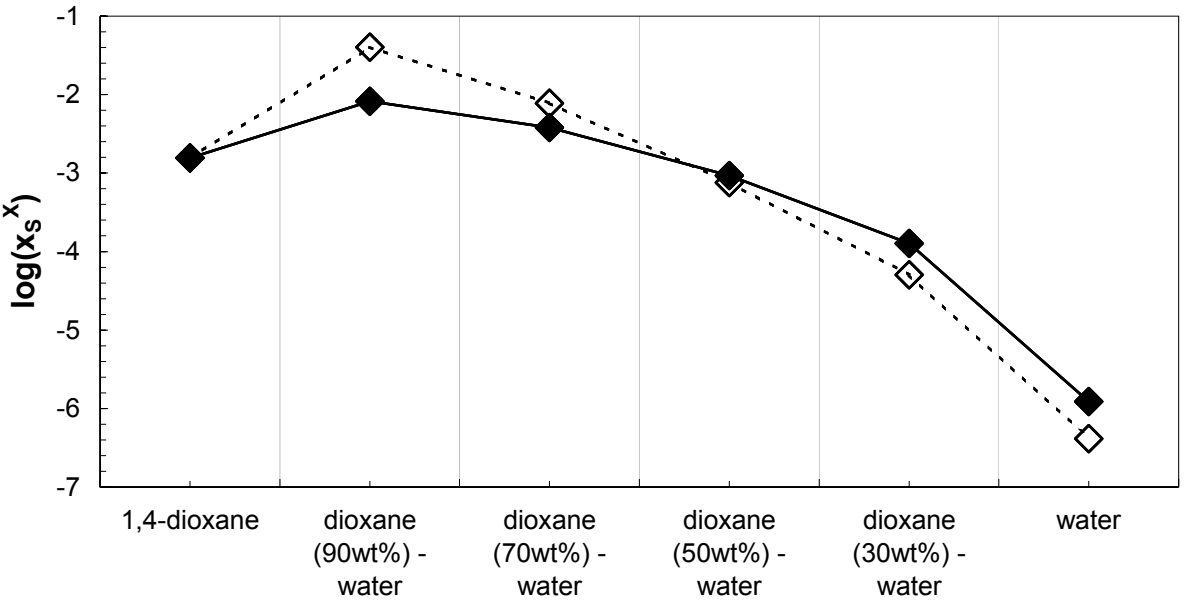

Figure 2 Experimental vs. calculated solubility $\log(x_S^X)$ of cinchonidine in dioxane – water solvent mixtures at T = 25 °C [32]. Filled rhombus: Experimental data of Ma and Zaera[34]. Open rhombus: COSMO-RSol predictions using experimental cinchonidine ΔG^X_{fus} data.

5 Salt Solubility

The prediction of the solubility of salts involves a few complications. First, in COSMO-RS a salt A^-C^+ is always treated by means of its anion A^- and cation C^+ separately. To obtain a salts

solubility, the pseudo chemical potentials and the free energy of fusion ΔG_{fus} have to be determined for the individual anion A^- and cation C^+. Now, the salts solubility is determined from the mean pseudo chemical potentials and the sum of the free energy of fusion of the ions. Thus for salts, Equation 4 is modified to

$$\log S_S^{AC} = \log\left(\frac{MW_{AC}\rho_S}{MW_S}\right) + \frac{1}{kT\ln(10)}\left[-\Delta_S^{AC}/2 + \min(0, \Delta G_{fus}^{AC})\right] \quad (6)$$

wherein $\Delta_S{}^{AC} = \mu^*_S{}^{AC} - \mu^*_{AC}{}^{AC}$. The pseudo chemical potential of the pure salt $\mu^*_{AC}{}^{AC}$ is the sum of the pseudo chemical potentials of the anion A^- and cation C^+ as determined in an equimolar (50:50) mix of anion A^- and cation C^+

$$\mu^*_{AC}{}^{AC} = \mu^*_{(50:50)}{}^{A^-} + \mu^*_{(50:50)}{}^{C^+} \quad (7)$$

The chemical potential of the salt in solution $\mu^*_S{}^{AC}$ is the sum of the chemical potentials of the anion A^- and cation C^+ computed in infinite dilution in solvent S:

$$\mu^*_S{}^{AC} = \mu^*_S{}^{A^-} + \mu^*_S{}^{C^+} \quad (8)$$

In a study on the aqueous solubility of 22 *para*-substituted benzoic acid salts of benzylamine at T = 37 °C[35], it was found that a simple correlation of $\log S_S{}^{AC}$ vs. $\Delta_S{}^{AC}$ as computed from Equations 7 and 8, yielded a correlation coefficient of $r^2 = 0.69$ and a rms-deviation of 1.46 log-units. Again, it turned out that the chemical potential difference between the ions in aqueous solution and the virtual super-cooled melt of the salt, i.e. the virtual ionic liquid, is the most important and significant contribution to the logarithmic solubility. To be able to obtain a QSPR expression for the salts free energy of fusion, the theoretical liquid solubility values of $\Delta_S{}^{AC}/kT\ln(10)$ were subtracted from the experimental values of solubility $\log S_S{}^{AC}$ to obtain pseudo-experimental data values for $\Delta G_{fus}{}^{AC}$. The performance of different

combinations of molecular descriptors provided by COSMO-RS in multilinear regression has been tested. The following descriptors of potential significance for ΔG_{fus}^{AC} have been tested in different combinations: molecular volume V or area A as a measure of the molecules size, the number of ringatoms $N_{ringatom}$ as a measure of molecular rigidity, the dielectric COSMO energy E_{diel} as a descriptor of polarity, the pseudo chemical potential μ^*_{water} of the salt in water as a combined measure of polarity and hydrogen bonding capacity and the generic COSMO-RS descriptors known as σ-moments[13] M_i. Finally, it turned out, that the descriptor combination V^{AC}, $N^{AC}_{ringatom}$, and $\mu^*_{water}{}^{AC}$ is best suited for the regression of ΔG_{fus}^{AC}. The polarity descriptor E_{diel}^{AC} did not achieve any significance. The improvement achieved by introducing the COSMO-RS σ-moments into the regression was negligible. Thus, the QSPR equation for salt free energy of fusion has the same functional form as the equation for neutral compounds, Equation (5). The QSPR parameters however, have to be readjusted for the computation of the salt's free energy of fusion. The QSPR descriptors of the salt are just the sums of the QSPR descriptors of the anion A^- and the cation C^+. Thus, for salts, Equation (5) translates to:

$$\Delta G_{fus}^{AC} = c_0 + c_V V^{AC} + c_\mu \mu^{*AC}_{water} + c_N N^{AC}_{ringatom} \qquad (9)$$

wherein $\mu^*_{water}{}^{AC} = \mu^*_{water}{}^{A-} + \mu^*_{water}{}^{C+}$, $V^{AC} = V^{A-} + V^{C+}$ and $N^{AC}_{ringatom} = N^{A-}_{ringatom} + N^{C+}_{ringatom}$. The correlation of the experimental aqueous solubilities $\log S_{water}{}^{AC}$ of a high quality data set[35] of 22 *para*-substituted benzoic acid salts of benzylamine at T = 37 °C with COSMO-RS properties via Equation (6) –(9) yielded a correlation coefficient of $r^2 = 0.56$ and a rms deviation of only 0.30 $\log(x_{water}{}^{AC})$ units[32]. Regression coefficients c_V = -36.9, c_μ = 0.046, c_N = -0.14, and a regression constant c_0 = 7.83 were determined, where the units for ΔG_{fus}^{AC} and $\mu^*_{water}{}^{AC}$ are [kJ/mol] and for V^{AC} it is [nm^3]. The fitted data set is given in Figure 3. The fitted coefficients were tested on a set of 8 aromatic carbon acid salts of the

local anaesthetic agent bupivacaine (CAS-RN: 2180-92-9) at T = 37 °C. This data was measured with the same experimental methodology as the fit set data[36,37]. Since the test data set is consistent with the measurement method and temperature of the training data set, but chemically much more complex than the fit data set, it can provide an indication for the extrapolative quality of the COSMO-RSol method for salts. The results for the test data set is also given in Figure 3. The experimental solubility values are predicted with a rms deviation of 0.37 $\log(x_{water}^{AC})$ units and a mean deviation of 0.15 $\log(x_{water}^{AC})$ units[32].

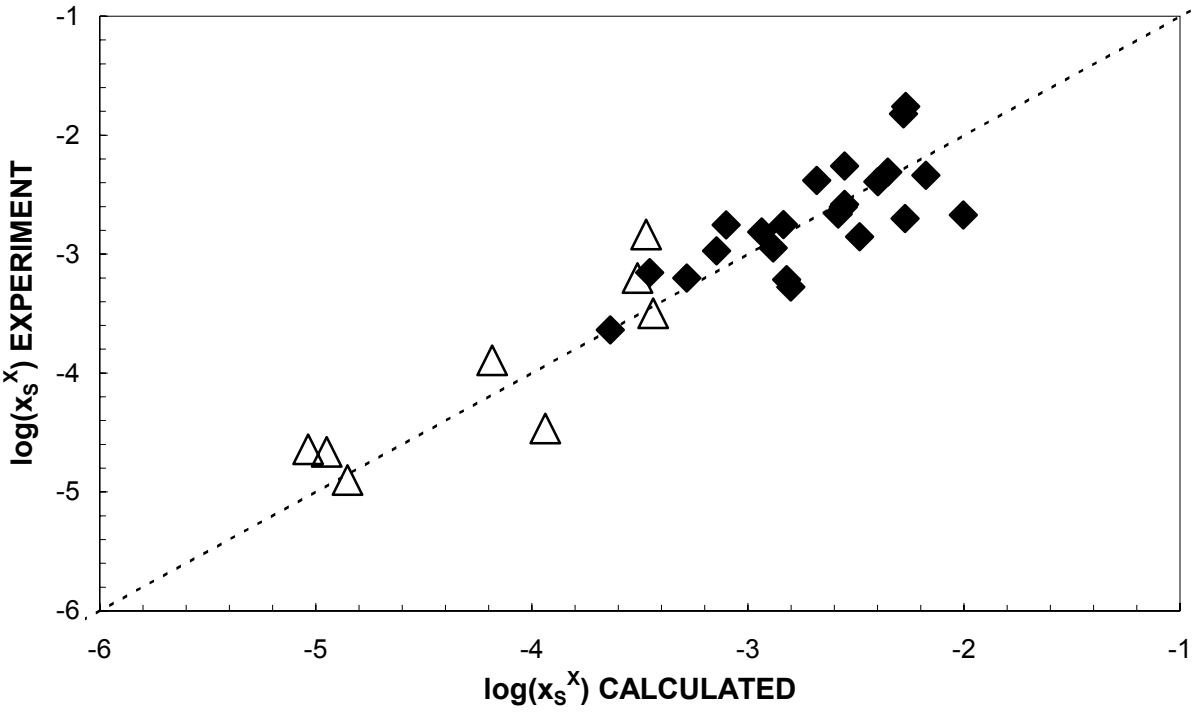

Figure 3 Experimental vs. calculated aqueous solubility $\log(x_{water}^{AC})$ of organic salts at T = 37 °C. Filled rhombus: Fitted data set of benzylamine salts. Open triangles: Test data set of bupivacaine salts.

Considering the increased chemical complexity of the test data set, it can be concluded that the methodology is predictive well beyond the boundary of similar or identical chemical

functionality and structure of the organic salts. It should be noted, however, that the fitted data set presented is too small and not diverse enough to be accounted for as a general and transferable prediction method for salt solubility by COSMO-RSol. Furthermore, the adjusted QSPR parameters are valid for T = 37 °C, the temperature of the solubility measurements, only. This data set was used to demonstrate the principle functionality and practical workability of the method. The generalization of the method to the whole of organic chemistry is straightforward, but ultimately depends on an appropriate set of experimental salt solubilities that allows the fitting of the QSPR parameters. Thus, currently the main problem remaining is the collection of a reliable and validated set of experimental room temperature solubilities of organic salts where the anions and cations show a broad distribution of chemical functionality and structure.

6 Summary and Conclusions

The applicability and capacity of the novel COSMO-RSol method as a prediction tool for the solubility of neutral solid compounds and salts in water and non-aqueous solvent media has been demonstrated. It was found that same empirical QSPR formula for the estimation of ΔG_{fus} can be used for salts and for neutral compounds. The coefficients of the QSPR model, however, have to be readjusted for salts. The readjustment of the QSPR coefficients for salts has been demonstrated at a coherent but small data set of experimental salt solubilities. The data set presented is not sufficiently large and diverse enough to provide general and transferable COSMO-RSol QSPR parameters for salts. Unfortunately the lack of reliable and diverse experimental solubilities for organic salt compounds may be the major drawback for further developments in this area. As a result it can be concluded that despite the empirical

character of the expression for the free energy of fusion, COSMO-RSol has a rather sound physico-chemical basis compared to all presently available prediction methods for solution phenomena and thus provides the most fundamental and transferable prediction tool for the solubility of organic compounds. Another advantage of this new method is that based on the same COSMO calculations used for aqueous solubility many other physico-chemical properties like solubility in non-aqueous solvents, partition coefficients, vapor pressures, Henry constants, etc. are easily available by COSMO-RS[9]. Even physiological partition behavior can be calculated based on COSMO-RS[13].

References

1. P. Kolar, J.-W. Shen, A. Tsuboi, T. Ishikawa, *Fluid Phase Equil*., 2002, **194-197**, 771.

2. W. L. Jorgensen, in P. v. R. Schleyer, L. Allinger, (Eds.) *Encyclopedia of Computational Chemistry*, Vol. 2, Wiley, New York, NY, 1998.

3. E.M. Duffy, W. L. Jorgensen, *J. Am. Chem. Soc.*, 2000, **122**, 2878.

4. C. Hansch, A. J. Leo, *Substituent Parameters for Correlation Analysis in Chemistry and Biology*, Wiley, New York, NY, 1979.

5. A. Fredenslund, J. Gmehling, P. Rasmussen, *Vapor Liquid Equilibria Using UNIFAC*, Elsevier, Amsterdam, 1977.

6. C. J. Cramer, D. G. Truhlar, in K. B. Lipkowitz,, D. B. Boyd, (Eds.), *Reviews in Computational Chemistry*, Vol. 6, VCH Publishers, New York, NY, 1995.

7. J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.*, 2005, **105,** 2999.

8. A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkins Trans.*, 1993, **2**, 799.

9. F. Eckert, A. Klamt, *AIChE J.*, 2002, **48**, 369.

10. A. Klamt, F. Eckert, *Fluid Phase Equilibria*, 2000, **172**, 43.

11. A. Klamt, V. Jonas, T. Buerger, J. C. W. Lohrenz, *J. Phys. Chem.*, 1998, **102**, 5074.

12. A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224.

13. A. Klamt, *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier, Amsterdam, 2005.

14. Alchemy32, Version 2.0.5, Tripos, Inc., St. Louis, MO, 1998.

15. N. L. Allinger, Y. H. Yuh, J.-H. Lii, *J. Am. Chem. Soc.*, 1989, **111**, 8551.

16. A. Schäfer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert, *Phys. Chem. Chem. Phys.*, 2000, **2**, 2187.

17. R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Letters*, 1989, **162**, 165.

18. Turbomole, Version 5.7, Universität Karlsruhe, Germany, 2004.

19. A. D. Becke, *Phys. Rev. A,* 1988, **38**, 3098.

20. J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822.

21. K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Letters*, 1995, **242**, 652.

22. K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119.

23. F. Eckert, A, Klamt, COSMO*therm*, Version C2.1 - Revision 01.05, COSMO*logic* GmbH & Co KG, Leverkusen, Germany, 2005.

24. A. Ben Naim, "Solvation Thermodynamics", Plenum Press, New York, NY, 1987.

25. P. Verwer, F. Leusen, in K. B. Lipkowitz, D. B. Boyd, (Eds.), *Reviews in Computational Chemistry*, Vol. 12, Wiley-VCH, New York, 1998.

26. A. Klamt, F. Eckert, M. Hornig, M. Beck, T. Bürger, *J. Comp. Chem.*, 2002, **23**, 275.

27. A. Klamt, F. Eckert, M. Diedenhofen, M. Beck, *J. Chem. Phys. A.*, 2003, **107**, 9380.

28. F. Eckert, A. Klamt, *J. Comp. Chem.*, 2006, **27**, 11.

29. H. Ikeda, K. Chiba, A. Kanou, N. Hirayama, *Chem. Pharm. Bull.*, 2005, **53**, 253.

30. S. Oleszek-Kudlak, M. Grabda, E. Shibata, F. Eckert and T. Nakamura, *Env. Tox. Chem.*, 2005, **24**, 1368.

31. A. K. Charltona, C. R. Danielsa, R. M. Wolda, E. Pustejovskya,W. E. Acree Jr., M. H. Abraham *J. Mol. Liquids*, 2005, **116**, 19, and references therein.

32. Supporting material with additional information and calculational details is available free of charge from the web-address: http://www.cosmologic.de/IUPAC-Solubility.html

33. NIST Standard Reference Database Number 69, June 2005 Release, http://webbook.nist.gov/chemistry/.

34. Z. Ma, F. Zaera, *J. Phys. Chem. B*, 2005, **109**, 406.

35. H. Parshad, K. Frydenvang, T. Liljefors, C. S. Larsen, *Int. J. Pharmaceutics*, 2002, **237**, 193.

36. J. Østergaard, S. W. Larsen, H. Parshad**,** C. Larsen, *Eur. J. Pharm. Sci.*, 2005, **26**, 280.

37. H. Parshad, *Design of Poorly Soluble Drug Salts*, Ph.D Thesis, The Danish University of Pharmaceutical Sciences, Copenhagen, 2003.