



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 9732

To link to this article : DOI:10.1021/ie3015318
URL : <http://dx.doi.org/10.1021/ie3015318>

To cite this version : Bouillot, Baptiste and Teychené, Sébastien and Biscans, Béatrice *An evaluation of COSMO-SAC model and its evolutions for the prediction of drug-like molecule solubility: part 1.* (2013) Industrial and Engineering Chemistry Research, vol. 52 (n° 26). pp. 9276-9284. ISSN 0888-5885

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

An Evaluation of COSMO-SAC Model and Its Evolutions for the Prediction of Drug-Like Molecule Solubility: Part 1

Baptiste Bouillot,* Sébastien Teychené, and Béatrice Biscans

Laboratoire de Génie Chimique, UMR CNRS 5503, Université de Toulouse, BP 84234 Campus INP-ENSIACET, 4 allée Emile Monso 31432 Toulouse Cedex 4, France

ABSTRACT: COSMO-SAC model and its evolutions have been investigated for solubility prediction of six pharmaceutical ingredients (ibuprofen, paracetamol, benzoic acid, salicylic acid, 4-aminobenzoic acid, and anthracene) in 35 solvents. The aim is to check the relevancy of the last improvements of the COSMO-SAC method for solubility prediction. The performance of each COSMO-SAC version has been evaluated following different criteria: the mean quadratic error (mse), the predicted solubility order of magnitude, and the solubility temperature dependence. In addition, solubility predictions in mixed solvents have also been analyzed. The results obtained show that the model refinements are relevant for solid–liquid equilibrium predictions. Compared to the original model, the solubility prediction in polar solvents is more accurate, and the solubility temperature dependences are well-represented. Moreover, the refined models (especially the 2010 version) are also more able to predict the solubility maxima in mixed solvents. However, the methods still have problems in representing complex interactions (hydrogen bond and dipole–dipole) and weaker electrostatic ones.

INTRODUCTION

The solubility of solid compounds is one of the most important thermodynamic properties in a wide range of processes (crystallization, precipitation of pharmaceutical products, etc.).

Over the past decades, much research has led to useful thermodynamic models for the prediction of liquid–liquid equilibria (LLE) or vapor–liquid equilibria (VLE). However, in the case of solid–liquid equilibria (SLE), such useful and reliable models do not exist. The main cause is probably the involvement of complex molecules, able to form various interactions of various strengths that are more difficult to quantify (electrostatic, hydrogen bonding, dipole–dipole, etc.).

In the past, some attempts to use classical models for SLE predictions have been made.^{1–5} Among these models, some (NRTL-SAC⁶ or PC-SAFT,⁷ for instance) have been recently built specifically for solid–liquid equilibria. However, even if these models give good results, none of them have been able to predict solubility accurately, regardless of the solute and the solvent.

In a previous paper,⁸ four classic models (UNIFAC, UNIFAC mod. Dortmund, COSMO-SAC, NRTL-SAC) were tested and compared to determine their potential for SLE prediction. Since it does not rely on a specific database and because the molecular structure is taken into account, it was found that COSMO-SAC was a promising model. The COSMO-SAC model is an excess Gibbs energy model based on quantum mechanical calculations. It was first developed by Lin and Sandler,⁹ and inspired by the COSMO-RS method developed by Klamt.¹⁰ In this method, each molecule is represented by the electronic charge density around it. This charge density is represented by a function (called the σ -profile) connecting a portion of the molecular surface to a charge density. However, even though COSMO-SAC was found to be promising, it was shown that the model highly overestimates the solubility by several orders of magnitude. It has been identified that the hydrogen bonding contribution to

the activity coefficient in the model was the origin of the model weaknesses.

Recently, COSMO-SAC has undergone several changes^{11–13} that have improved the description of hydrogen bonds. Hsieh et al.¹⁴ have investigated the potential of the last version of the model for the prediction of the solubility of 51 drugs in 37 different solvents. This last study was mainly statistical, and it did not really focus on the solubility temperature dependence.

In this paper, to extend the previous work of Hsieh et al.,¹⁴ solubility predictions were performed by the last two refinements of COSMO-SAC, for six organic molecules in 35 organic solvents. The results are analyzed and discussed employing several criteria: the mean quadratic error (mse), the prediction of the order of magnitude of the solubility, the ability to predict the temperature dependence of the solubility, and the solubility profile in mixed solvents.

THEORY

Equilibrium Equation. Phase equilibria are described by the equality of chemical potentials (μ) in each phase. In the case of SLE, the thermodynamic equilibrium is called the solubility. By a development of the chemical potential equality, the solubility equation for a compound i is

$$\ln(\gamma_i x_i) = \frac{\Delta H_m^{\text{fus}}(T_m)}{R} \left(\frac{1}{T_m} - \frac{1}{T} \right) - \frac{\Delta C_{pm}(T_m)}{R} \left[\ln \left(\frac{T_m}{T} \right) - \frac{T_m}{T} + 1 \right] \quad (1)$$

This equation gives the solubility as a function of the solid-state properties of compound i (ΔH_m is the molar melting enthalpy; T_m is the melting temperature; and ΔC_{pm} is the difference between the heat capacity of the supercooled melt and the heat capacity of the solid), and the activity coefficient γ . The activity coefficient represents the nonideal behavior of the mixture. The second part of eq 1 is more difficult to determine, because of the ΔC_{pm} term, and eq 2 is most commonly used in chemical engineering:

$$\ln(\gamma_i x_i) = \frac{\Delta H_m^{\text{fus}}(T_m)}{R} \left(\frac{1}{T_m} - \frac{1}{T} \right) \quad (2)$$

To be able to calculate the solubility with the previous equations, the thermodynamic properties of the solid form and the activity coefficient of the solute in the liquid phase must be known.

COSMO-SAC Models. *Original COSMO-SAC.* The COSMO-SAC model uses quantum chemical methods for the representation of the charge distribution of the molecules (σ -profiles), and statistical thermodynamics to get the molecular interactions for predicting the chemical potential of a molecule from its structure. To determine the activity coefficient, only the σ -profiles and the van der Waals surface and volume are needed. This makes COSMO-SAC a fully predictive model.

The model gives an expression of the solvation free energy ΔG^{ssolv} , which defines, at constant temperature and pressure, the energy difference between an ideal solution (i.e., in a perfect conductor continuum with an infinite dielectric constant) and the solute in the real solvent.

In this model, the classical combinatory term is expressed using the Staverman–Guggenheim combinatorial term, and the residual is expressed from the σ -profiles of the molecules. These profiles correspond to the probability distribution of the surface charge density of a molecule. The details of the procedure to obtain the σ -profiles (determination of surface charge density and normalization) from a commercially available molecular simulation software can be found in Mullins et al.,¹⁵ and a more-detailed explanation of the COSMO-SAC method can be found in the original paper by Lin and Sandler.⁹

From these σ -profiles, the enthalpic contribution to the activity coefficient (residual term) can be found:

$$\ln \gamma_i^{\text{res}} = \frac{1}{a_{\text{eff}}} \sum_{\sigma_m} p'_i(\sigma_m) [\ln \Gamma_S(\sigma_S) - \ln \Gamma_i(\sigma_m)] \quad (3)$$

where σ_m is the surface charge density of segment m , $p'_i(\sigma_m)$ the surface area of the molecular cavity of surface charge density σ_m , a_{eff} the effective area of the standard surface segment ($a_{\text{eff}} = 7.50 \text{ \AA}^2$), and $\Gamma_S(\sigma_m)$ and $\Gamma_i(\sigma_m)$ are, respectively, the residual activity coefficients of the segment of surface charge density σ_m in the solution and in a pure solution of molecule i .

The interactions, both electrostatic and hydrogen bonding, are taken into account through a ΔW term:

$$\Delta W(\sigma_m, \sigma_n) = \left(\frac{\alpha'}{2} \right) (\sigma_m + \sigma_n)^2 + c_{\text{hb}} \max[0, \sigma_{\text{acc}} - \sigma_{\text{hb}}] \min[0, \sigma_{\text{don}} + \sigma_{\text{hb}}] \quad (4)$$

where $\alpha'/2$ is a constant for the misfit energy, c_{hb} the hydrogen bonding parameter, and σ_{hb} a cutoff value for the hydrogen bonding interactions.

Hydrogen bonds can occur between two segments. To express them, the segments are divided into two categories: acceptor and donor. In fact, segments can be either neutral segments, acceptor segments, or donor segments. In the original COSMO-SAC, σ_{acc} and σ_{don} correspond to the largest and the smallest values of σ_m and σ_n .

COSMO-SAC 2007. This second evolution of the model was published by Wang et al.¹² In that paper, the authors suggested a new averaging method for the calculation of the σ -profiles and a new expression for the contribution of the hydrogen bond. Instead of using a threshold value for the donor or acceptor nature of the segments, the authors suggested using a probability distribution. The possibility of forming hydrogen bonds is based on a continuous probability function of the charge density: the higher the charge density, the higher the probability.

This probability function, which weights the σ -profiles, is P^{hb} :

$$P^{\text{hb}}(\sigma) = 1 - \exp\left(-\frac{\sigma^2}{2\sigma_0^2}\right) \quad (5)$$

This function is a reversed Gaussian curve (σ_0 was found to be equal to 0.007 e/\AA^2 and corresponds to a standard deviation). The higher the value of σ_0 , the fewer hydrogen bonds can occur for the lowest charge density values.

Moreover, this change takes into account the improvement suggested by Lin et al.,¹¹ a division of the σ -profiles, into two parts:

- Surfaces that are not able to form hydrogen bonds (i.e., the previous neutral segments); and
- Surfaces that are able to form hydrogen bonds (surfaces around O, N, and F atoms and the H atoms connected to these atoms).

The aim of this division is to offer the possibility of a better representation of the hydrogen bonds.

COSMO-SAC 2010. More recently, Hsieh et al.¹³ suggested a COSMO-SAC refinement that can be summarized by the introduction of two major contributions: (1) The electrostatic constant is temperature dependent:

$$C_{\text{ES}} = A_{\text{ES}} + \frac{B_{\text{ES}}}{T^2} \quad (6)$$

with $A_{\text{ES}} = 6525.69 \text{ kcal}/(\text{mol } \text{\AA}^4)/e^2$ and $B_{\text{ES}} = 1.4859 \times 10^8 \text{ kcal}/(\text{mol } \text{\AA}^4)/e^2 \text{ K}^2$. (2) A subdivision of the original σ -profile into three parts:

- the p_{OH} profile, corresponding to the surfaces around hydroxyl groups (OH),
- the p_{OT} profile for the surfaces around O, F, and N atoms, with the H atoms connected to them (except OH groups), and
- the p_{nhb} profile (no possible hydrogen bonds).

This change has the same purpose as a previous one:¹¹ a better account of hydrogen bonding interactions. Accordingly, in COSMO-SAC 2010, the c_{hb} parameter in ΔW is written as follows:

$$c_{hb}(\sigma_m^t, \sigma_n^s) = \begin{cases} c_{OH-OH} = 4013.78 \text{ kcal}/(\text{mol } \text{\AA}^4)/e^2 & \text{if } s = t = \text{OH and } \sigma_m^t \sigma_n^s < 0 \\ c_{OT-OT} = 932.31 \text{ kcal}/(\text{mol } \text{\AA}^4)/e^2 & \text{if } s = t = \text{OT and } \sigma_m^t \sigma_n^s < 0 \\ c_{OH-OT} = 3016.43 \text{ kcal}/(\text{mol } \text{\AA}^4)/e^2 & \text{if } s = \text{OH and } t = \text{OT and } \sigma_m^t \sigma_n^s < 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

The values of the parameters have been obtained from regressions of VLE and LLE data. Hsieh et al.¹³ have observed that a division of p_{hb} into more than two parts is not conclusive and does not give more-accurate predictions for LLE and VLE. They also noticed that a c_{hb} temperature-dependent parameter is not useful.

Finally, ΔW is rewritten as follows:

$$\Delta W(\sigma_m^t, \sigma_n^s) = C_{ES}(\sigma_m^t + \sigma_n^s)^2 - c_{hb}(\sigma_m^t, \sigma_n^s)(\sigma_m^t - \sigma_n^s)^2 \quad (8)$$

EXPERIMENTAL SOLUBILITY AND PURE SOLID PROPERTIES

Reference Molecules. In this paper, form I versions of paracetamol, ibuprofen, benzoic acid, and salicylic acid were chosen as the model drugs. These molecules were chosen mainly because (i) a significant amount of experimental solubility data is available in the literature^{16–28} and (ii) they contain various usual functional groups (OH, COOH, NH, NH₂). To complete this solubility database, 4-aminobenzoic acid was also considered, because of its amine group^{27–30}, as well as anthracene, because of its simple structure.^{31–35} The molecular structures of the studied molecules are shown in Figure 1.

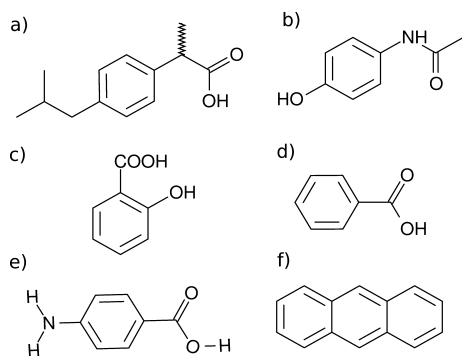


Figure 1. Chemical structures of (a) ibuprofen, (b) paracetamol, (c) salicylic acid, (d) benzoic acid, (e) 4-aminobenzoic acid, and (f) anthracene.

The database used includes the results from the cited literature and the experimental measurements in the previous paper,⁸ plus

some new solid-state property measurements and solubility measurements in mixed solvents.

In order to be able to use the original COSMO-SAC method, σ -profiles were taken from the VT database.^{15,36} The methodology described by Mullins^{15,36} was followed to compute the appropriate σ -profiles for the last COSMO-SAC model versions (2007 and 2010).

Pure Solid-State Properties. In this study, some additional measurements of the solid-state properties were performed using classic differential scanning calorimetry (DSC), including the new values for the ΔC_p terms.

The DSC device used for this study is the Q2000 DSC system from Thermal Analysis. Salicylic and benzoic acid were obtained from Fisher Scientific (>99.5%), and paracetamol as obtained from Sigma–Aldrich (>98%). Ibuprofen was provided by BASF (Ibuprofen 25, >98.5%).

The operating procedure followed was an increase in the temperature of the samples of 1 °C/min from the ambient temperature to 110 °C in the case of ibuprofen, 220 °C in the case of paracetamol, 210 °C in the case of salicylic acid, and 165 °C in the case of benzoic acid. This procedure was repeated three times for ibuprofen, four times for paracetamol and salicylic acid, and five times in the case of benzoic acid. Peak integration from baselines interpolation was used for the calculation of the melting enthalpies. The ΔC_p terms were calculated from baseline differences (before and after the melting).

The experimental results obtained are shown in Table 1. The mean values (experimental and literature) were used for the calculations.

The uncertainties of the pure solid properties have an impact on the model predictions. Their influence on the solubility predictions is investigated using a classic Monte Carlo method.⁸

Solubility in Mixed Solvents. Materials. Salicylic and benzoic acid were obtained from Fisher Scientific (>99.5%), and paracetamol was obtained from Sigma–Aldrich (>98%); ibuprofen was provided by BASF (Ibuprofen 25, >98.5%). Heptane, ethanol, and acetone were obtained from Fisher Scientific (AnalaR NORAMPUR >99%). They were used without further purification.

Solubility Measurements. The solubility measurements were carried out using the classic analytical shake-flask method.³⁷

Table 2 gives the experimental solubility values for ibuprofen and salicylic acid.

Table 1. Mean Values (and Standard Deviations) of Melting Temperatures, Enthalpies, and ΔC_{pm} of Ibuprofen, Paracetamol, Salicylic Acid, and Benzoic Acid from DSC Experiments

compound	ΔH_{mm} (J/mol)		T_m (K)		ΔC_{pm} (J/mol/K)	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
ibuprofen	25692	1146	347.5	0.1	54.3	18.4
paracetamol	27350	2722	441.9	0.1	94.3	16.7
salicylic acid	25269	835	431.4	0	157.4	98.2
benzoic acid	17060	1082	395.0	0.1	47.3	11.8

Table 2. Experimental Solubility in Mixed Solvents at 20 °C

mixture	Volume Fraction (%)		Solubility	
	solvent 1	solvent 2	(mg/g)	mol fraction
ibuprofen	20	80	786.24	1.93×10^{-1}
heptane/acetone	40	60	686.0	1.86×10^{-1}
heptane/acetone	60	40	534.3	1.65×10^{-1}
heptane/acetone	80	20	353.4	1.29×10^{-1}
ibuprofen	20	80	873.52	1.77×10^{-1}
heptane/ethanol	40	60	797.0	1.81×10^{-1}
heptane/ethanol	60	40	672.7	1.78×10^{-1}
heptane/ethanol	80	20	501.5	1.61×10^{-1}
ibuprofen	0	100		1.67×10^{-1}
acetone/ethanol	20	80	913.4	1.75×10^{-1}
acetone/ethanol	40	60	1077.1	2.08×10^{-1}
acetone/ethanol	60	40	1082.8	2.16×10^{-1}
acetone/ethanol	80	20	1074.8	2.23×10^{-1}
salicylic acid	20	80	396.8	1.28×10^{-1}
heptane/ethanol	40	60	316.8	1.16×10^{-1}
heptane/ethanol	60	40	234.0	1.01×10^{-1}
heptane/ethanol	80	20	127.1	6.79×10^{-2}
heptane/ethanol	100	0	2.2	1.68×10^{-3}

CRITERIA FOR EVALUATING THE MODELS

The results were analyzed employing three main criteria:

- The mean square error (mse) between the predicted and the experimental solubility:

$$mse = \frac{1}{n} \sum_i^n \left(\frac{x_{i,predicted} - x_{i,experimental}}{x_{i,experimental}} \right)^2 \quad (9)$$

where n is the number of data points. The experimental solubility data considered can be found in Bouillot et al.,⁸ and the solute/solvent pairs are reported in Table 3.

- The temperature dependence, evaluated as the slope of (dx/dT) , is a very interesting criterion, because it gives information about the supersaturation calculations or crystallization yields. For example, the crystallization yield may be written as $\Delta x/x^* = (x/x^*) - 1$, where x^* is the equilibrium molar fraction. Even if a model is not accurate, this last ratio can be calculated accurately if the solubility temperature dependence is well-handled by the method. Moreover, from the knowledge of the temperature dependence, the entire solubility curve can be determined. If we assume that the solubility logarithm is dependent on $1/T$ (as is often the case), the main criterion to investigate the temperature dependence is the slope of the function $\ln(x) = f^0(1/T)$. The leading coefficient of this line will be written as “ k ” ($\ln(x) = k \times (1/T) + K$). By evaluating the coefficient k of the predicted solubility curves for a given solvent, the temperature dependence will be obtained. If the predicted coefficient k is similar to the experimental one, crystallization yield calculations can be possible, since $x/x^* = \exp(k/T)/\exp(k/T^*) = \exp[k(1/T - 1/T^*)]$ (where T^* is the saturation temperature at concentration x^*). In addition, by using the same procedure, the

Table 3. Experimental and Predicted Solubility Orders of Magnitude, by COSMO-SAC (2007 and 2010) at 30 °C (for Ibuprofen and Benzoic Acid in Logarithm of Solubility)

experimental		2007		2010	
solvent	order of magnitude	solvent	order of magnitude	solvent	order of magnitude
Ibuprofen					
heptane	[−3; −2]	heptane	[−4; −3.5]	heptane	[−4; −3.5]
cyclohexane	[−2; −1.5]	cyclohexane	[−2.5; −2]	cyclohexane	[−2.5]
ethanol	[−1.5; −1]	toluene	[−2; −1.5]	toluene	[−2; −1.5]
toluene		chloroform	[−1.5; −1]	ethyl acetate	[−1.5; −1]
ethyl acetate		octanol	[−1]	chloroform	
isopropanol		ethanol	[−1; −0.5]	octanol	[−1; −0.5]
acetone		isopropanol		ethanol	
octanol	[−1; −0.5]	ethyl acetate		isopropanol	
chloroform		acetone		acetone	
Benzoic Acid					
hexane	[−5; −4]	heptane	[−5.5]	heptane	[−5.5; −5]
cyclohexane		hexane		cyclohexane	
heptane		cyclohexane		hexane	
acetonitrile	[−3; −2]	carbon tetrachloride	[−4.5; −4]	carbon tetrachloride	[−4.5; −4]
carbon tetrachloride		benzene	[−3; −2.5]	benzene	[−3; −2.5]
benzene		DMSO	[−2; −1.5]	DMSO	[−2; −1.5]
heptanol	[−2; −1.5]	octanol	[−1.5; −1]	N-methylpyrrolidone	
acetone		acetonitrile		acetonitrile	[−1.5; −1]
octanol		butanol		dioxane	
isopropanol		isopropanol		octanol	
butanol		dioxane	[−1; −0.5]	butanol	
N-methyl pyrrolidone	[−1.5; −1]	acetone		isopropanol	[−1; −0.5]
dioxane		N-methylpyrrolidone		acetone	

Table 4. Experimental and Predicted Temperature Dependence: Slope of Function $\ln(x) = f^0(1/T)$: Coefficient k and the Regression Coefficient

solvent	Experimental		Original COSMO-SAC		COSMO-SAC 2007		COSMO-SAC 2010	
	k	regression coefficient	k^a	regression coefficient	k^a	regression coefficient	k^a	regression coefficient
Ibuprofen								
toluene	−3700	0.998	−4250 (15)	0.996	−5440 (47)	0.989	−5540 (50)	0.992
acetone	−2870	0.999	−960 (67)	0.999	−1290 (55)	0.999	−1470 (49)	0.999
ethanol	−3460	0.981	−1140 (67)	0.999	−2000 (42)	0.999	−1600 (54)	0.999
Paracetamol								
acetone	−1870	0.999	−1330 (29)	0.999	−1570 (16)	0.999	−2100 (12)	0.999
ethanol	−1467	0.998	−1590 (9)	0.999	−2320 (58)	0.999	−2630 (80)	0.999
propanol	−1700	0.999	−1800 (6)	0.999	−2600 (53)	0.999	−2900 (71)	0.999
Salicylic Acid								
acetone	−1140	0.999	−650 (43)	0.999	−930 (19)	0.999	−1350 (18)	0.999
acetonitrile	−2600	0.999	−1770 (32)	0.999	−1880 (28)	0.999	−2450 (6)	0.999
methanol	−1500	0.999	−870 (42)	0.999	−1790 (19)	0.999	−1700 (14)	0.999
Benzoic Acid								
cyclohexane	−6230	0.995	−5160 (17)	0.977	−5640 (9)	0.976	−5970 (4)	0.977
acetone	−1700	0.999	−610 (64)	0.999	−860 (49)	0.999	−930 (45)	0.999
isopropanol	−1690	0.999	−1160 (32)	0.999	−1340 (21)	0.999	−1010 (40)	0.999
octanol	−2160	0.994	−840 (61)	0.993	−1590 (26)	0.994	−1200 (45)	0.993
Anthracene								
toluene	−3040	0.992	−3480 (15)	0.999	−3480 (15)	0.999	−3470 (14)	0.999
MEK	−4580	0.952	−3210 (30)	0.999	−3180 (31)	0.999	−3090 (33)	0.999
isopropanol	−4880	0.938	−3610 (26)	0.999	−3600 (26)	0.999	−3570 (27)	0.999

^aThe experimental relative error, expressed as a percentage, is given in parentheses.

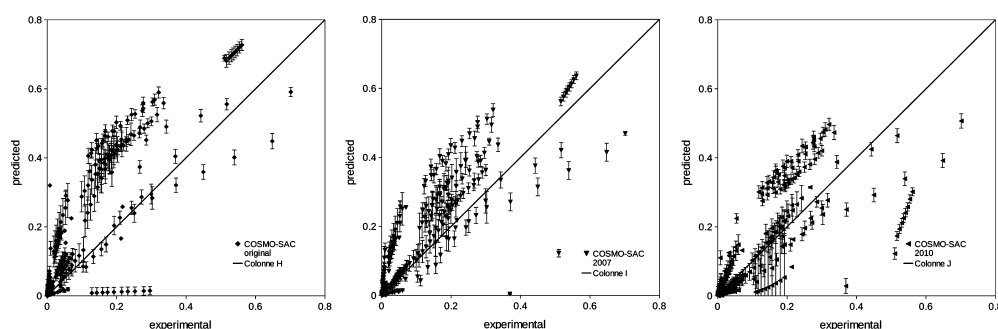


Figure 2. COSMO-SAC (2010) predictions versus experimental solubility for (a) the original COSMO-SAC method, (b) COSMO-SAC 2007, and (c) COSMO-SAC 2010.

knowledge of one experimental point then allows calculation of the solubility at any temperature.

- The solubility orders of magnitude (in logarithm) and solubility ranking preservation (see Table 3).
- The ability of the models to represent the influence of a new solvent on the solubility.

In the next section, in all prediction-versus-experimental figures, the model errors (from the uncertainties in the solid-state properties) are represented by error bars. The experimental errors are not represented, since they are of the same order of magnitude as the size of the symbols.

RESULTS

Review of the Original COSMO-SAC Results, and a New Look at the Solubility Temperature Dependency. The original COSMO-SAC results were presented and discussed in a previous paper.⁸ In order to compare them to the two modified methods, the previous conclusions will be briefly recalled.

It was previously shown that the original COSMO-SAC model overestimates the solubilities. Depending on the nature of the solvent, it could be seen that this model gives good results in aprotic and apolar solvents (ibuprofen in toluene, heptane or cyclohexane, benzoic acid in hexane or benzene, etc.).

However, it gives poor results in polar solvents, especially in solvents able to form hydrogen bonds (similar to alcohols). In this type of solvent, the predicted solubilities are overestimated, from 50% to more than 100%. To explain this, the influence of the hydrogen bonding parameter, c_{hb} , was investigated. It is important to notice that this parameter has been determined using VLE data. In SLE, especially for active pharmaceutical products, the molecules are big and flexible. The interaction possibilities between these types of molecules are large, but they are not easily established, because the interacting sites can be buried or hidden, so that hydrogen bonds cannot form. Consequently, some surfaces should not be considered when evaluating the molecular interactions from the σ -profiles.

Concerning the solubility temperature dependence, the original COSMO-SAC gives interesting results. The temperature dependence is in good agreement with the experimental data in the nonpolar solvents. The errors in k are found to be <17% (see Table 4, presented later in this work). In general, the model tends to underestimate the temperature dependence, because the predicted k coefficients are higher than the experimental ones. In the case of alcohols, the model shows very good results for paracetamol solubility. Less-successful results are obtained in polar solvents. It seems that there is a balance problem between the hydrogen bond interactions and the weakest interactions.

Finally, the model does not preserve the solubility rankings and orders of magnitude.

COSMO-SAC 2007. The COSMO-SAC 2007 results are shown in Figure 2 and Table 3. The mean square errors are about 11.4. They are in general smaller than those obtained from the original model (which was about 24). Moreover, the overestimation of the solubility orders of magnitude is not as important as in the previous version of the model (see Figure 2).

Then, in apolar and aprotic solvents, the model is less accurate than the original one (see ibuprofen in heptane or cyclohexane, benzoic acid in benzene, or hexane in Table 3). For these types of solvents, the σ -profiles are not very different from the original ($p(\sigma) \equiv p_{\text{nhb}}(\sigma)$). More specifically, this recurrent loss of accuracy occurs for complex solutes (not for anthracene, for which the results are the same). Indeed, as the electrostatic parameter is still the same, there are differences only when hydrogen bonding is possible. The accuracy loss could be attributed to a wrong balance between the different types of interactions.

In polar solvents, there is a good improvement. The hydrogen bonding is more accurately taken into account. The solubility predictions in alcohols are better than before, with an error of ~25% (versus an error of 200%, for salicylic acid or paracetamol predictions⁸ with the original COSMO-SAC). However, there is a difference between polar and protic solvents and polar and aprotic solvents. In protic solvents, the improvement is clear. In polar and aprotic solvents, the errors remain quite high.

Table 4 shows that the model improves prediction of solubility temperature dependencies. It can also be seen that in 80% of the cases, this dependence is overestimated (except for paracetamol in alcohols and ibuprofen in toluene). In the case of anthracene, the results show that there is very little evolution when the weakest interactions are predominant. Overall, there is a clear improvement in all the systems involving polar solvents. The only exception is for paracetamol. But paracetamol is generally less soluble than the other molecules, and is probably more subject to error.

Quantitatively, it should also be noticed that the model is more subject to errors due to the uncertainties in the thermodynamic solid state properties (as shown in Table 5: a mean error of 39% against 29% for the original model).

Solubility ranking is still an issue. There is no agreement with the experimental results (see Table 3). But the solubility orders of magnitude in protic solvents are better than the ones obtained from the original model. Despite the clear improvement in protic solvents, there are still difficulties in representing the solubility orders of magnitude when polar and aprotic solvents are used.

This difficulty of predicting solubility in polar and aprotic solvents may be due to the difficulty in quantifying hydrogen bonding and dipole–dipole interactions when a molecule is aprotic.

Table 5. Mean Quadratic Error (mse) Values of Various COSMO-SAC Predictions (Original, COSMO-SAC 2007, COSMO-SAC 2010)

model	Mean Quadratic Error, mse		mean standard deviation (%)
	prediction	prediction without 4-aminobenzoic acid	
Original COSMO-SAC	23.5	24	29
COSMO-SAC 2007	11.4	10.4	39
COSMO-SAC 2010	155	5.4	64

COSMO-SAC 2010. The results for COSMO-SAC 2010 are shown in Figure 2 and Table 3. The mean square error (mse) values are ~155 (5.4 without 4-aminobenzoic acid): the solubility prediction is significantly improved. The overestimation observed is less important than before (see Figure 2).

Concerning the solubility predictions as functions of the solvent nature, some similarities can be observed with the previous results. First, the solubility predictions in apolar and aprotic solvents are less accurate than with the original model (ibuprofen in toluene; benzoic acid in benzene, cyclohexane, or hexane). Once again, the solubility predictions for anthracene are not very accurate. The electrostatic parameter (C_{ES}) may not be appropriate for SLE, with regard to hydrogen bonding parameters (possible wrong balance between them).

The σ -profiles division then results in a better representation of molecular interactions when hydrogen bonds are involved. The solubility predictions in polar and aprotic solvents are generally more accurate than for the previous version of COSMO-SAC (for example, paracetamol in acetonitrile or ethyl acetate; salicylic acid in acetone).

Qualitatively, the model presents solubility orders of magnitude closer to the experimental values. However, the solubility ranking among the solvents is not better than the original model. There are no significant differences between polar protic or polar aprotic solvents.

The solubility temperature dependence preservation is slightly improved compared to COSMO-SAC 2007 (see Table 4). In the case of anthracene, the solubility temperature dependence shows that the electrostatic interactions (C_{ES} parameter) are not really improved with this model.

Finally, the overall quality of the COSMO-SAC 2010 model tends toward valid refinements made for SLE predictions. The subdivision of σ -profiles into three parts helps to take into account the molecular interactions more precisely for solubility predictions. However, the 2010 version gives less good results than the 2007 version for solubility temperature dependence prediction, and it is sensitive to the solid-state properties (see Table 5).

A Closer Look at the Temperature Dependence.

Knowing the solubility temperature dependence, from one solubility experimental point at a given temperature T_1 , the solubility at another temperature T_2 is calculated using the expression

$$x_2 = x_1 \times \exp\left(\frac{k}{T_2} - \frac{k}{T_1}\right) \quad (10)$$

Table 6 gives some examples of predictions using eq 10 and the k coefficient given in Table 4. The results observed agree well

Table 6. Experimental Solubility at Temperature T_1 and Predicted Solubility at Temperature T_2 , According to eq 10, and the Relative Error

product	solvent	T_1 (°C)	T_2 (°C)	experimental at T_1 (mole fraction)	experimental at T_2 (mole fraction)	Original COSMO-SAC		2007		2010	
						prediction at T_2 (mole fraction)	relative error (%)	prediction at T_2 (mole fraction)	relative error (%)	prediction at T_2 (mole fraction)	relative error (%)
benzoic acid	isopropanol	49.57	4.91	0.297	0.129	0.167	29	0.152	18	0.179	39
salicylic acid	acetone	50	10	0.241	0.146	0.181	24	0.160	10	0.134	9
ibuprofen	acetone	35	10	0.321	0.142	0.244	72	0.222	56	0.211	48
paracetamol	ethanol	20	0	0.055	0.038	0.037	3	0.031	20	0.028	25
anthracene	toluene	50	20	0.0159	0.0061	0.0052	13	0.0053	13	0.0053	13

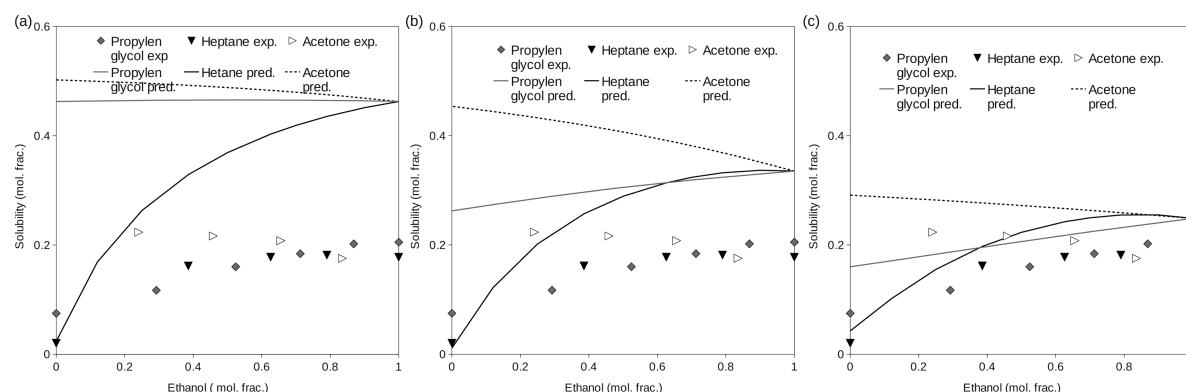


Figure 3. Experimental and predicted solubility with (a) original COSMO-SAC, (b) COSMO-SAC 2007, and (c) COSMO-SAC 2010 of ibuprofen in mixed solvents, as a function of the solvent composition.

with the experimental values, especially for the COSMO-SAC 2007 version. It should also be noticed that the relative errors with the experimental data are of the same orders of magnitude as the relative errors in coefficient k (see Table 4).

Thus, knowledge of the coefficient k and its relative error might lead to determination of the entire solubility curve with just one piece of experimental data.

Solubility in Mixed Solvents. To check the accuracy of COSMO-SAC in solubility prediction in mixed solvents, two cases were investigated.

First, as in the case of ibuprofen in ethanol plus another solvent (i.e., acetone, *n*-heptane, or propylene glycol), the solubility of the API increases monotonically with the addition of a second solvent. The experimental and predicted solubility curves with the different COSMO-SAC models are given in Figure 3. The results shows that, first, the COSMO-SAC 2010 model really improves the prediction, compared to the former version of COSMO-SAC, and, obviously, the better the solubility prediction in pure solvents, the better the solubility prediction in mixed solvents.

The second case concerns the description of the synergetic effect (the presence of an extremum in the solubility curve) of adding a second solvent to the mixture. The solubility of paracetamol in a mixture of water and acetone is chosen as an example. The experimental data and the predictions of the different models are shown in Figure 4. The results obtained show that neither the original COSMO-SAC method nor the COSMO-SAC 2007 method are able to predict the solubility maximum observed experimentally. However, the improvements of the hydrogen bonding descriptions made in COSMO-SAC 2010 allows the model to predict the presence of the solubility maximum, even though the order of magnitude of the paracetamol solubility in the pure solvent is not well-predicted.

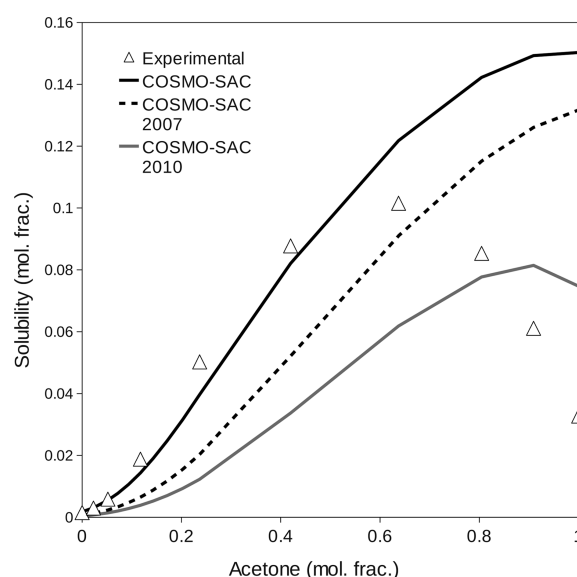


Figure 4. Experimental and COSMO-SAC predicted solubility of paracetamol in acetone/water, as a function of the initial mole fraction of acetone.

CONCLUSION

The two most recent COSMO-SAC refinements were studied with regard to their solubility predictions. The results showed that the successive changes are relevant and are useful for modeling molecular interactions. From the first original model to the last change suggested, quantitative solubility predictions improved. Although the refinements are found to be more

subject to errors from parameter uncertainties, the orders of magnitude of the predicted solubilities are closer to the experimental ones with the last COSMO-SAC version, and the predictions in solvents able to form stronger molecular interactions, such as hydrogen bonds, are much more accurate.

Particularly, an improvement of the results for complex organic molecules in polar solvents was observed. The COSMO-SAC 2010 version is also better able to predict solubility maxima in mixed solvents. Indeed, it seems that the suggested improvements for a better representation of hydrogen bonds are relevant because they refine the interactions of the segments. Indeed, as the interaction parameters have been estimated from VLE and considering small molecules, it is likely that the hydrogen bonding is not properly taken into account when considering bigger and more-complex molecules. Moreover, the electrostatic parameter shows some limitations for solubility predictions in mixtures where the weakest interactions are predominant (van der Waals interactions between instantaneous dipoles).

Nevertheless, this study shows that the COSMO-SAC refinements are more accurate for predicting the solubility temperature dependency, which can lead to supersaturation or crystallization yield calculations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: baptiste.bouillot@ensiacet.fr.

Notes

The authors declare no competing financial interest.

REFERENCES

- Gmehling, J. G.; Anderson, T. F.; Prausnitz, J. M. Solid–Liquid Equilibria Using UNIFAC. *Ind. Eng. Chem. Fundam.* **1978**, *17*, 269–273.
- Gracin, S.; Brinck, T.; Rasmuson, A. Prediction of solubility of solid organic compounds in solvents by UNIFAC. *Ind. Eng. Chem. Res.* **2002**, *41*, 5114–5124.
- Tung, H.; Tabora, J.; Variankaval, N.; Bakken, D.; Chen, C. Prediction of Pharmaceutical Solubility Via NRTL-SAC and COSMO-SAC. *J. Pharm. Sci.* **2008**, *97*, 1813–1820.
- Mota, F. L.; Carneiro, A. P.; Pinho, S. P.; Macedo, E. A. Temperature and solvent effect in solubility of some pharmaceutical compounds: Measurements and modelling. *Eur. J. Pharm. Sci.* **2009**, *37*, 499–507.
- Hahnenkamp, I.; Graubner, G.; Gmehling, J. Measurement and prediction of solubilities of active pharmaceutical ingredients. *Int. J. Pharm.* **2010**, *388*, 73–81.
- Chen, C.-C.; Song, Y. Solubility Modeling with a Nonrandom Two-Liquid Segment Activity Coefficient Model. *Ind. Eng. Chem. Res.* **2004**, *43*, 8354–8362.
- Gross, J.; Sadowski, G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.* **2001**, *40*, 1244–1260.
- Bouillot, B.; Teychené, S.; Biscans, B. An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents. *Fluid Phase Equilib.* **2011**, *309*, 36–52.
- Lin, S.; Sandler, S. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899–913.
- Klamt, A. Conductor-Like Screening Model for Real Solvents—A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- Lin, S.; Chang, J.; Wang, S.; Goddard, W.; Sandler, S. Prediction of vapor pressures and enthalpies of vaporization using a Cosmo solvation model. *J. Phys. Chem. A* **2004**, *108*, 7429–7439.
- Wang, S.; Sandler, S.; Chen, C. Refinement of COSMO-SAC and the Applications. *Ind. Eng. Chem. Res.* **2007**, *46*, 7275–7288.
- Hsieh, C.; Sandler, S.; Lin, S. Improvement of COSMO-SAC for Vapour–Liquid and Liquid–Liquid equilibrium predictions. *Fluid Phase Equilib.* **2010**, *297*, 90–97.
- Hsieh, C.-M.; Wang, S.; Lin, S.-T.; Sandler, S. I. A Predictive Model for the Solubility and Octanol–Water Partition Coefficient of Pharmaceuticals. *J. Chem. Eng. Data* **2011**, *56*, 936–945.
- Mullins, E.; Liu, Y.; Ghaderi, A.; Fast, S. D. Sigma profile database for predicting solid solubility in pure and mixed solvent mixtures for organic pharmacological compounds with COSMO-based thermodynamic methods. *Ind. Eng. Chem. Res.* **2008**, *47*, 1707–1725.
- Garzon, L. C.; Martinez, F. Temperature dependence of solubility for ibuprofen in some organic and aqueous solvents. *J. Solution Chem.* **2004**, *33*, 1379–1395.
- Gracin, S.; Rasmuson, A. Solubility of phenylacetic acid, *p*-hydroxyphenylacetic acid, *p*-aminophenylacetic acid, *p*-hydroxybenzoic acid, and ibuprofen in pure solvents. *J. Chem. Eng. Data* **2002**, *47*, 1379–1383.
- Granberg, R.; Rasmuson, A. Solubility of paracetamol in pure solvents. *J. Chem. Eng. Data* **1999**, *44*, 1391–1395.
- De Fina, K.; Sharp, T.; Roy, L.; Acree, W. Solubility of 8-hydroxybenzoic acid in select organic solvents at 298.15 K. *J. Chem. Eng. Data* **1999**, *44*, 1262–1264.
- Fung, H.; Higuchi, T. Molecular interactions and solubility of polar nonelectrolytes in nonpolar solvents. *J. Pharm. Sci.* **1971**, *60*, 1782–1788.
- Nordstrom, F. L.; Rasmuson, A. C. Solubility and melting properties of salicylic acid. *J. Chem. Eng. Data* **2006**, *51*, 1668–1671.
- Paruta, A.; Sciarrone, B.; Lordi, N. Solubility Of Salicylic Acid as a Function of Dielectric Constant. *J. Pharm. Sci.* **1964**, *53*, 1349–1353.
- Shalmashi, A.; Eliassi, A. Solubility of salicylic acid in water, ethanol, carbon tetrachloride, ethyl acetate, and xylene. *J. Chem. Eng. Data* **2008**, *53*, 199–200.
- Acree, W.; Bertrand, G. Thermochemical Investigations Of Nearly Ideal Binary Solvents 0.7. Monomer and Dimer Models for Solubility of Benzoic-Acid in Simple Binary and Ternary Solvents. *J. Pharm. Sci.* **1981**, *70*, 1033–1036.
- Beerbower, A.; Wu, P.; Martin, A. Expanded Solubility Parameter Approach. 1. Naphthalene and Benzoic-Acid in Individual Solvents. *J. Pharm. Sci.* **1984**, *73*, 179–188.
- Li, D.; Liu, D.; Wang, F. Solubility of 4-methylbenzoic acid between 288 and 370 K. *J. Chem. Eng. Data* **2001**, *46*, 234–236.
- Lin, H.; Nash, R. An Experimental Method for Determining the Hildebrand Solubility Parameter of Organic Nonelectrolytes. *J. Pharm. Sci.* **1993**, *82*, 1018–1025.
- Yalkowski, S.; Valvani, S.; Roseman, T. Solubility and partitioning VI: Octanol solubility and octanol-water partition coefficients. *J. Pharm. Sci.* **1983**, *72*, 866–870.
- Hancock, C.; Pawlowski, J.; Idoux, J. J. *Org. Chem.* **1966**, *31*, 3801.
- Daniels, C.; Charlton, A.; Wold, R.; Moreno, R.; Acree, W.; Abraham, M. Mathematical correlation of 4-aminobenzoic acid solubilities in organic solvents with the Abraham solvation parameter model. *Phys. Chem. Liq.* **2004**, *42*, 633–641.
- Cepeda, E.; Gomez, B. Solubility of Anthracene and Anthraquinone in Some Pure and Mixed Solvents. *J. Chem. Eng. Data* **1989**, *34*, 273–275.
- Cepeda, E.; Diaz, M. Solubility of Anthracene and Anthraquinone in Acetonitrile, methyl ethyl ketone, isopropyl alcohol and their mixtures. *Fluid Phase Equilib.* **1996**, *121*, 267–272.
- Al-Sharrah, G.; Ali, S.; Fahim, M. Solubility of anthracene in two binary solvents containing toluene. *Fluid Phase Equilib.* **2002**, *193*, 191–201.
- Powell, J.; Acree, W. Solubility of Anthracene in Binary Alcohol + Dibutyl Ether Solvent Mixtures. *J. Chem. Eng. Data* **1995**, *40*, 914–916.
- Zvagin, A.; Acree, W. Thermochemical Investigations of Molecular Complexation. Estimation of Anthracene–Ethyl Acetate and Anthracene–Diethyl Adipate Association Parameters from Measured Solubility Data. *Phys. Chem. Liq.* **1991**, *24*, 31–42.

(36) Mullins, E.; Oldland, R.; Liu, Y.; Wang, S.; Sandler, S. I.; Chen, C.-C.; Zwolak, M.; Seavey, K. C. Sigma-profile database for using Cosmo-based thermodynamic methods. *Ind. Eng. Chem. Res.* **2006**, *45*, 4389–4415.

(37) Mullin, J. W. *Crystallization*, 3rd Edition; Butterworth–Heinemann: Oxford, U.K., 1993.