



**HAL**  
open science

# **An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents**

Baptiste Bouillot, Sébastien Teychené, Béatrice Biscans

## ► **To cite this version:**

Baptiste Bouillot, Sébastien Teychené, Béatrice Biscans. An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents. *Fluid Phase Equilibria*, 2011, vol. 309 (n° 1), pp. 36-52. <10.1016/J.FLUID.2011.06.032>. <hal-03539264>

**HAL Id: hal-03539264**

**<https://hal.science/hal-03539264v1>**

Submitted on 21 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



## 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: 6017

**To link to this article:** DOI:10.1016/J.FLUID.2011.06.032  
URL: <http://dx.doi.org/10.1016/J.FLUID.2011.06.032>

**To cite this version:** Bouillot, Baptiste and Teychené, Sébastien and Biscans, Béatrice (2011) An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents. *Fluid Phase Equilibria*, vol. 309 (n°1). pp. 36-52. ISSN 0378-3812

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

# An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents

Baptiste Bouillot\*, Sébastien Teychené, Béatrice Biscans

Laboratoire de Génie Chimique, UMR CNRS 5503, BP 84234 Campus INP-ENSIACET, 4 allée Emile Monso, 31030 Toulouse Cedex 4, France

## A B S T R A C T

Prediction of solubility of active pharmaceutical ingredients (API) in different solvents is one of the main issue for crystallization process design. Experimental determination is not always possible because of the small amount of product available in the early stages of a drug development. Thus, one interesting perspective is the use of thermodynamic models, which are usually employed for predicting the activity coefficients in case of Vapour–Liquid equilibria or Liquid–Liquid equilibria (VLE or LLE). The choice of the best thermodynamic model for Solid–Liquid equilibria (SLE) is not an easy task as most of them are not meant particularly for this. In this paper, several models are tested for the solubility prediction of five drugs or drug-like molecules: Ibuprofen, Acetaminophen, Benzoic acid, Salicylic acid and 4-aminobenzoic acid, and another molecule, anthracene, a rather simple molecule. The performance of predictive (UNI-FAC, UNIFAC mod., COSMO-SAC) and semi-predictive (NRTL-SAC) models are compared and discussed according to the functional groups of the molecules and the selected solvents. Moreover, the model errors caused by solid state property uncertainties are taken into account. These errors are indeed not negligible when accurate quantitative predictions want to be performed. It was found that UNIFAC models give the best results and could be an useful method for rapid solubility estimations of an API in various solvents. This model achieves the order of magnitude of the experimental solubility and can predict in which solvents the drug will be very soluble, soluble or not soluble. In addition, predictions obtained with NRTL-SAC model are also in good agreement with the experiments, but in that case the relevance of the results is strongly dependent on the model parameters regressed from solubility data in single and mixed solvents. However, this is a very interesting model for quick estimations like UNIFAC models. Finally, COSMO-SAC needs more developments to increase its accuracy especially when hydrogen bonding is involved. In that case, the predicted solubility is always overestimated from two to three orders of magnitude. Considering the use of the most accurate equilibrium equation involving the  $\Delta C_p$  term, no benefits were found for drug predictions as the models are still too inaccurate. However, in function of the molecules and their solid thermodynamic properties, the  $\Delta C_p$  term can be neglected and will not have a great impact on the results.

## 1. Introduction

Solubility of solid compounds in solvents is obviously one of the most fundamental physicochemical properties, particularly useful in a wide variety of phenomena relevant to biological, pharmaceutical, environmental and other organic, physical and analytical chemistry, and engineering sciences. From a simple thermodynamic point of view, solubility is the fugacity equality between the solid state and the liquid state of the solute molecule. In practice, it is the maximum amount of the crystalline form of the compound that can be dissolved in a given volume of solvent at

a given temperature and pressure. This state is the expression of the Gibbs free energy of mixing, which is the combination of two contributions. The enthalpic contribution which reflects the solute–solute and solute–solvent interactions; and the entropic contribution which takes into account the molecular size and the potential for non-random distribution of molecules in the mixture. Then, the molecular size and the potential for non-random distribution are also important. From a practical point of view, solubility has a considerable influence on the choice of solvent for performing polymorphism, salt and co-crystal screening and it is the key parameter for the process design of a new drug by crystallization, as it will determine the amount of cooling required to yield a given amount of product and will in fact determine if cooling will provide a reasonable product yield and thus define the crystallization process itself (cooling, evaporation, or antisolvent). In

\* Corresponding author.

E-mail address: baptiste.bouillot@ensiacet.fr (B. Bouillot).

addition, at the early stages of a new drug molecule development, small amount of product (typically less than 5 mg) is available, so classical solubility methods cannot be used since they are product consuming and are neither rapid nor cost effective. To overcome these issues, high-throughput solubility-screening protocols have been developed based on systems using either robots [1] or droplet microfluidic [2]. All these methods rely on kinetic determination of dissolution and re-crystallization steps. The process of dissolution of a crystalline compound into a solvent generally shows an asymptotic profile [3]: as the concentration of the compound in solution increases, the process becomes progressively slower, and saturation is theoretically never reached. In addition, polymorphic phase transition may occur during these experiments, leading to misinterpretation of the results.

Even when products are available, the experimental determination of solubility must be prepared and completed by preliminary calculations. Indeed, in classical sealed flask methods, saturated solutions are prepared with an excess of solid at a given temperature. The suspension is then filtrated or centrifuged to ensure a solution completely free from undissolved solute. First, by their nature, this traditional solubility method is time consuming. In addition, the filtration or centrifugation steps are often performed at a lower temperature than the saturation experiments, what entails a shift in solubility values.

Indeed, experimental methods of solubility measurements are time consuming, expensive and often have technical problems, in spite of the obtained values varying in the degree of accuracy and precision. To guide these experiments and to limit misinterpretation, thermodynamic modelling of Solid-Liquid Equilibrium (SLE), using a minimum of experimental data, would be of a great value. Based on the actual SLE thermodynamics, a question arises: can the existing models be used as guide for designing solubility experiments? [4] showed that the original UNIFAC model could predict SLE. But, the results were accurate only for restricted compounds. [5] discussed the limits of the UNIFAC model and the equilibria equations used for SLE, but did not focus on the temperature dependence or compare with other models. [6] showed that COSMO-SAC is an interesting model to calculate a priori solubilities, but it is not very accurate for SLE. In addition, these papers did not investigate neither the complete equilibrium equation nor the influence of temperature on solubility predictions. [7] used NRTL-SAC for SLE prediction and found it promising, but did not compared the results to a much more models. More recently, [8] compared UNIFAC model to COSMO-RS model for some organic compounds.

The main objective of this paper is to test and analyze existing predictive (UNIFAC, UNIFAC mod., COSMO-SAC) and semi-predictive (NRTL-SAC) thermodynamic models for the prediction of solubility of six compounds (five drugs or drug-like molecules and one hydrocarbon) in several solvents. Four classic drugs were chosen because of the large number of experimental solubility data and of the different chemical groups (Ibuprofen, Acetaminophen, Benzoic acid, and Salicylic acid). To complete this set, 4-aminobenzoic acid was also chosen given its amine group, and anthracene, a quite simple molecule (only ACH groups).

In this work, the thermodynamic models were chosen based on their various approaches, and their ability to be used with or without experimental data. UNIFAC and its modifications are group contribution methods. They only require binary interaction parameters between the functional groups. These models rely on the use of a specific database of group-group interaction parameters. When these parameters are not available (for molecules containing cyclic S or N for instance), these models cannot be used. COSMO-SAC, is based on quantum mechanics and requires profiles of the charge density around the studied molecules (called  $\sigma$ -profile). The challenge is the determination of the "good"  $\sigma$ -profile which takes into account the right conformations of the molecule in the solu-

tion. NRTL-SAC is based on polymer NRTL and to be used, four parameters representing the behaviour of the molecules have to be calculated from experimental data regression. Once these parameters have been calculated for one molecule, they may be used for any other solubility predictions in any solvent. That is why there is no need to regress each time, as in the case of most the semi-predictive models like UNIQUAC or Wilson.

These four models will be compared and discussed regarding solubility prediction at several temperatures. In addition, the relative importance of the pure solid state properties (i.e. the melting temperature, melting enthalpy and the heat capacity difference between the supercooled liquid and the solid) on the thermodynamic modelling is discussed.

## 2. Theory

### 2.1. Equilibrium equation

Phase equilibria are described by the equality of chemical potentials,  $\mu$ , in each phase. The governing equation of solubility of the compound in a given solvent is written as follows:

$$\mu^S = \mu^{Sat} \quad (1)$$

The superscript 'S' denotes the solid phase, and the superscript 'Sat' the saturated solution. The chemical potential of the solute in the saturated solution is:

$$\mu^{Sat} = \mu^0 + RT \ln(\gamma^{Sat} x^{Sat}) \quad (2)$$

where  $\mu^0$  is the chemical potential in the reference state,  $T$  the temperature, and  $\gamma$  the activity coefficient. The activity of the solute ( $a = \gamma x$ ) in the solvent is obtained by combining Eqs. 1 and 2:

$$\ln(\gamma^{Sat} x^{Sat}) = \frac{\mu^S - \mu^0}{RT} = \frac{g^S - g^0}{RT} = \frac{\Delta g_m}{RT} \quad (3)$$

where  $\Delta g_m$  is the change of partial molar Gibbs energy of the solute from a solid state to the reference state at constant temperature and pressure. This last equation gives a relationship between solubility and activity coefficients, which can finally be written as [9]:

$$\ln(\gamma^{Sat} x^{Sat}) = \frac{\Delta H_m(T_m)}{R} \left( \frac{1}{T_m} - \frac{1}{T} \right) - \frac{\Delta C_p(T_m)}{R} \left[ \ln \left( \frac{T_m}{T} \right) - \frac{T_m}{T} + 1 \right] \quad (4)$$

where  $\Delta H_m$  is the melting enthalpy,  $T_m$ , the melting temperature and  $\Delta C_p$  the difference between the heat capacity of the supercooled melt and the heat capacity of the solid. In practice, this last term can be very difficult to measure especially when sublimation, decomposition [10] or parallel reaction occurs during melting, and is often neglected because it is usually thought that its contribution is negligible. In that case Eq. (4) is simplified as follows:

$$\ln(\gamma^{Sat} x^{Sat}) = \frac{\Delta H_m(T_m)}{R} \left( \frac{1}{T_m} - \frac{1}{T} \right) \quad (5)$$

### 2.2. Group contribution models (classic and modified UNIFAC)

UNIFAC method, for UNIQUAC functional group activity coefficient [11], is based on the UNIQUAC model. The main idea of this approach is a separation of the activity coefficient of the pure compound into two parts:

$$\ln \gamma = \ln \gamma^C + \ln \gamma^R \quad (6)$$

$\gamma^C$  is the combinatorial term. It represents the entropic contribution to the activity coefficient which takes into account the shape and size of the molecules. This  $\gamma^C$  depends on the mole fraction ( $x_i$ ), on the area ( $\theta_i$ ) and segment fraction ( $\Phi_i$ ), and on the Van der

Waals area ( $r_i$ ) and volume ( $q_i$ ) (the superscript 'i' designates the compound):

$$\ln \gamma_i^C = \ln \frac{\Phi_i}{x_i} + 5q_i \ln \frac{\theta_i}{\Phi_i} + l_i - \frac{\Phi_i}{x_i} \sum_j x_j (5(r_j - q_j) - (r_j - 1)) \quad (7)$$

$\gamma^R$  is the residual part which represents the contribution of inter and intramolecular interactions (enthalpic contribution). It is a sum of the activity coefficients of the functional groups weighted by their number in solution.

$$\ln \gamma_i^R = \sum_k v_k^{(i)} \left[ \ln \Gamma_k - \Gamma_k^{(i)} \right] \quad (8)$$

where  $v_k$  and  $v_k^i$  are the number of groups of kind  $k$  in the mixture and in component  $i$ .  $\Gamma_k$  and  $\Gamma_k^i$  are, respectively, the residual activity coefficient of group  $k$  in the mixture and in a solution of pure component  $i$ . They depend on the area and segment fraction of the compounds and on adjustable binary interaction parameters  $a_{mn}$  (the interactions between the functional groups) that are regressed usually from VLE experimental data:

$$\ln \Gamma_k = Q_k \left[ 1 - \ln \left( \sum_m \Theta_m \Psi_{mk} \right) - \sum_m \Theta_m \Psi_{km} / \sum_n \Theta_n \Psi_{nm} \right] \quad (9)$$

with:

$$\Psi_{mn} = \exp \left( -\frac{a_{mn}}{T} \right) \quad (10)$$

In UNIFAC mod. (Dortmund), by [12], a modified version of UNIFAC, is supposed to take a better account of the temperature in the residual activity contribution. The new combinatorial term and interaction parameter are written as, respectively:

$$\ln \gamma_i^C = 1 - \ln \left( \frac{\Phi_i}{x_i} \right) - 5q_i \left( 1 - \frac{\Phi_i}{\theta_i} + \ln \left( \frac{\Phi_i}{\theta_i} \right) \right) \quad (11)$$

and

$$\Psi_{mn} = \exp \left( -\frac{a_{mn} + b_{mn}T + c_{mn}T^2}{T} \right) \quad (12)$$

UNIFAC and modified UNIFAC are frequently used for the prediction of Vapor-Liquid and Liquid-Liquid equilibria for simple molecules. These models can be applied to all molecules providing that the binary interaction parameters exist, which is not always the case. Moreover, the conformation and isomerism of molecules are not considered.

### 2.3. Segment activity coefficient models

#### 2.3.1. COSMO-SAC

COSMO-SAC is based on COSMO-RS model [13,14]. First published by Lin and Sandler [15], this model is a combination of a quantum chemical treatment of solutes and solvents with statistical thermodynamics procedure for the determination of the molecular surface interactions.

The COSMO-SAC model is based on the calculation of the solvation free energy  $\Delta G^{sol}$  which defines the free energy change comparing to an ideal solution at constant temperature and pressure, when the solute molecules are introduced into the solvent. In COSMO-based models, this term is divided into two parts. First, the solute charges are turned off, and its molecules are inserted into a perfect conductor (with an infinite dielectric constant  $\epsilon$ ). This first step gives the cavitation formation free energy,  $\Delta G^{cav}$ , which depends on the molecule specific dimension and represents the entropic contribution of the activity coefficient. When the charges

are turned on (the solvent becomes a real solvent), the free energy of restoring the charges,  $\Delta G^{res}$ , corresponds to the enthalpic contribution. Combining these two steps, the activity coefficient is expressed as:

$$\ln \gamma_i = \ln \gamma_i^C + \frac{\Delta G_{i/S}^{*res} - \Delta G_{i/i}^{*res}}{RT} \quad (13)$$

To calculate the restoring free energy, a  $\sigma$ -profile of each molecule is needed. This profile corresponds to the probability of finding a surface of the molecule with a charge density  $\sigma$ . They are computed from quantum chemical simulations. Once they have been calculated, the molecules are divided into segments. These conceptual segments are portions of the molecule surface with the same charge density  $\sigma$ . The equation for the activity coefficient is then written:

$$\ln \gamma_i = n_i \sum_{\sigma_m} p_i(\sigma_m) [\ln \Gamma_S(\sigma_S) - \ln \Gamma_i(\sigma_m)] + \gamma_i^C \quad (14)$$

where  $\Gamma_S(\sigma_m)$  and  $\Gamma_i(\sigma_m)$  are the activity coefficients of segment  $\sigma_m$  in solution and in pure liquid, respectively. In COSMO-SAC model, the activity coefficient is obtained by summing the contributions of each segment.

The  $\sigma$ -profiles used in this work were taken from Mullins database [6,16]. Unlike other group-contribution methods, this model is not strictly a pair-wise additive one, since the  $\sigma$ -profiles take into account the geometry of molecules and hydrogen bonding. As the  $\sigma$ -profiles take into account the geometry of the molecules, each molecule and its conformation have its own profile [16].

#### 2.3.2. NRTL-SAC

NRTL-SAC model [17] is based on the original NRTL (Non-Random Two-Liquid) [18] and polymer NRTL [19]. The NRTL-SAC model characterizes the molecules in terms of pre-defined conceptual segments. These segments account for the interactions of each molecule in the solution. [17] have defined four types of conceptual segments: hydrophobic  $X$ , repulsive  $Y^-$  and attractive  $Y^+$  polar, and hydrophilic  $Z$ . Like in UNIFAC, the NRTL-SAC model computes the activity coefficient for component  $i$  from the combinatorial term  $\gamma^C$  and the residual term  $\gamma^R$  (Eq. (6)). The combinatorial term is then calculated from the Flory-Huggins approximation for the entropy of mixing:

$$\ln \gamma_i^C = \ln \frac{\Phi_i}{x_i} + 1 - r_i \sum_j \frac{\Phi_j}{r_j} \quad (15)$$

where  $r_i$  and  $\Phi_i$  are, respectively, the total segment number and the segment mole fraction of component  $i$ .

The residual term  $\gamma^R$  is calculated from the local composition (lc) interaction contribution [19] applied to each segment:

$$\ln \gamma_i^R = \ln \gamma_i^{lc} = \sum_k r_{k,i} \left[ \ln \Gamma_k^{lc} - \ln \Gamma_k^{lc,i} \right] \quad (16)$$

The segment-segment interaction parameters and the conceptual segment values of solvents were determined by the regression of experimental vapor-liquid and liquid-liquid equilibrium data [17].

The segment values for a solute are obtained by the regression of solubility data in at least four solvents: one hydrophilic (i.e. water), one polar attractor (i.e. alcohol), one polar donor (i.e. ketone), and one hydrophobic (i.e. n-alkane). As suggested by [17], it is also possible to estimate pure solid properties using Eq. (4) or (5). However, in this study, in order to compare the model performances with the other, we chose to use experimental solid state properties, since they are available in the literature and are easily accessible experimentally. Once the segment values of the solute are obtained, they can be used to predict solubility in other solvents or solvent mixtures.

## 2.4. How the errors on the thermodynamic properties impact the models predictions?

As these models are used with the SLE equation, the solid state properties and the experimental uncertainties have an impact on the parameter determination or model prediction. Indeed, and as shown in the next section, measurements of the thermodynamic properties of the solids are scarcely accurate and to investigate the model performances, their experimental uncertainties have to be taken into account. Thus, the solid state properties,  $\Delta H_m$ ,  $T_m$  and  $\Delta C_p$ , are defined by mean values with standard deviations. Using Eq. (4), there are three parameters (two for Eq. (5)). The influence of the experimental uncertainties on solubility prediction is evaluated by using a classic Monte Carlo method. Solubility predictions are performed by choosing randomly vectors of mean value and standard deviation corresponding to the experimental data, and described it by a Gaussian law. Then, the predicted solubility standard deviation is calculated (about a thousand random simulations).

## 3. Experimental solubility and pure compound properties

### 3.1. Reference drugs

In this paper, Mainly form I of acetyl p-aminophenol (paracetamol), 2-(p-isobutylphenyl)propionic acid (ibuprofen), benzoic acid, and 2-hydroxybenzoic acid (salicylic acid) were chosen as model drugs. These molecules were chosen mainly because of a significant amount of solubility experimental data available in the literature [20–32] and they contain various usual functional groups (OH, COOH, NH, NH<sub>2</sub>). To complete this solubility database, 4-aminobenzoic acid was considered because of its structure (amine group) despite there is few data available [31–34], and anthracene because of its simple structure (ACH functional groups) and the many solubility data available [35–39] although it is not a drug. The molecules structures are shown in Fig. 1.

### 3.2. Experimental

In this study, in spite of the experimental results being taken from the literature, some data were missing. The solubility of ibuprofen in octanol and heptane, paracetamol in heptane at 30 °C were measured. Then, the melting temperature and enthalpy of ibuprofen, paracetamol, salicylic acid and benzoic acid were measured. Combining with the literature data, the goal was to have mean values and standard deviation for these parameters.

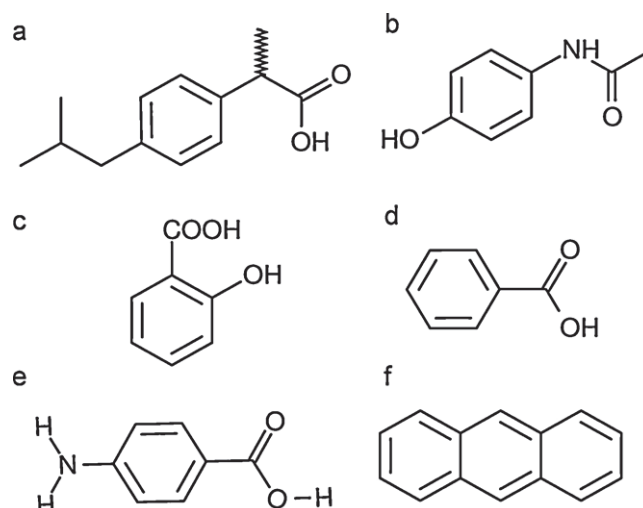
#### 3.2.1. Materials

Salicylic and benzoic acid were bought at Fisher Scientific (>99.5%), Ibuprofen at BASF (Ibuprofen 25 >98.5%), and paracetamol at Sigma-Aldrich (>98%). Heptane and octanol were bought at AnalaR NORMAPUR (>99%). They were used without further purification.

**Table 1**

Mean values and standard deviation of melting temperatures, enthalpies and  $\Delta C_p$  of ibuprofen, paracetamol, salicylic acid, benzoic acid, anthracene and 4-aminobenzoic acid from DSC experiments (except the last two) and data taken from the literature [8,21,22,25,40–53].

Compound	$\Delta H_m$ (J/mol)	$T_m$ (K)	$\Delta C_p$ (J/(mol K))
Ibuprofen [8,21,47,53]	25203.9 (1577)	347.88 (0.65)	48.8
Paracetamol [22,43,45–47,49,50]	27470.6(1720.3)	442.1(0.47)	87.4
Salicylic acid [25,47]	24626.12(2652.4)	431.37(0.12)	None
Benzoic acid [40–42,44,48,51,52]	17350.34(774.7)	395.17(0.29)	58.43(0.81)
Anthracene [47,52]	28790 (585.1)	489.18 (0.89)	6.28
4-Aminobenzoic acid [47]	24030	461.65	–



**Fig. 1.** (a) Ibuprofen, (b) paracetamol, (c) salicylic acid, (d) benzoic acid, (e) 4-aminobenzoic acid and (f) anthracene.

### 3.2.2. Determination of the thermodynamic properties of the solids

The thermodynamic properties of the solids were determined by Differential Scanning Calorimetry (Thermal Analysis Q2000 DSC). The measuring method was a 1 °C per minute ramp from ambient temperature to 110 °C, 220 °C, 210 °C and 165 °C for ibuprofen, paracetamol, salicylic acid and benzoic acid, respectively.

### 3.2.3. Solubility measurements

Solubility measurements were carried out in water with physical mixtures of enantiomers by the analytical shake-flask method, using constant-temperature jacketed glass cells. A constant temperature of 30 °C ( $\pm 0.1$  °C) was maintained with a circulating water bath. The way of preparing the saturated solutions was slightly modified compared to the classical method. Here, solubility was determined by adding weighed quantities of powder mixture to a known volume of distilled water, until crystals cannot be dissolved anymore. Small amounts of solid (about 20 mg) were added each time and no addition was performed before the solution was clear. When the equilibrium was reached, solutions were stirred for 24 h.

The uncertainty of solubility determination,  $\Delta S$ , is lower than 5%.

## 4. Results

### 4.1. Experimental solubility and pure compound properties

Concerning the thermodynamic properties of the solids, the experimental results of this work and found in the literature are shown in Fig. 2, and the mean values and standard deviation are presented in Table 1.

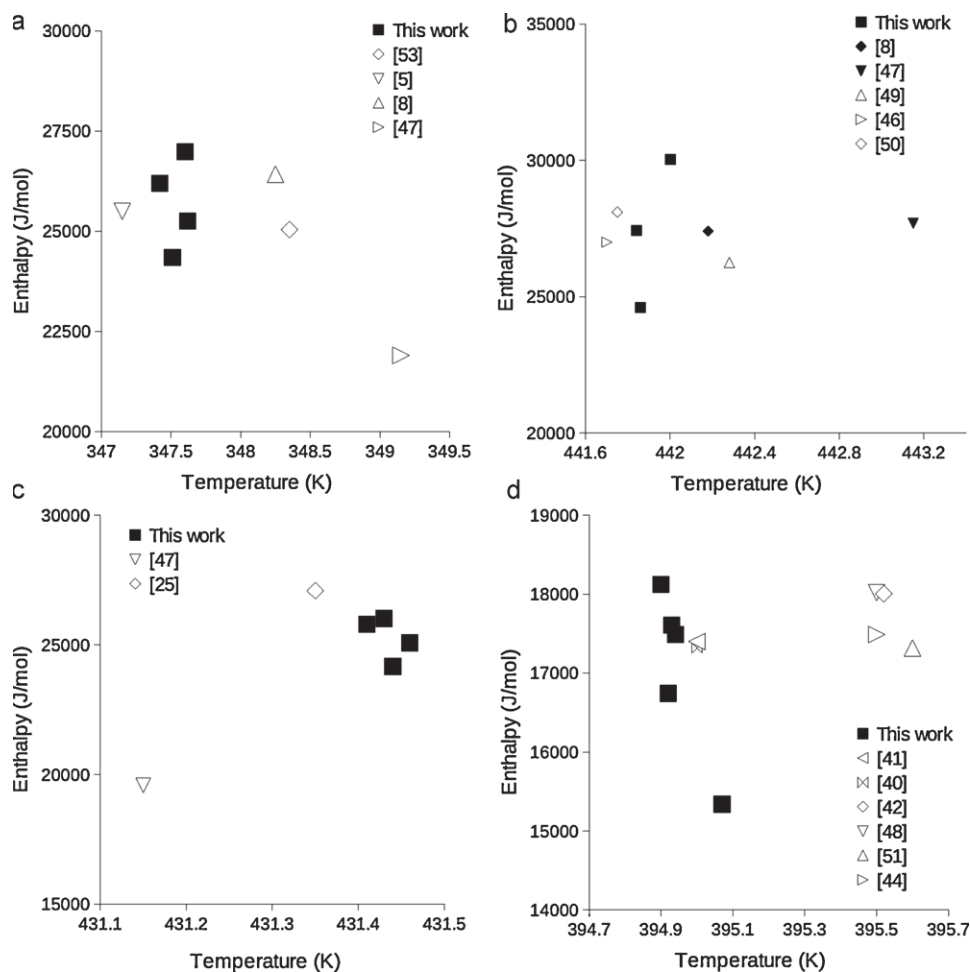


Fig. 2. Melting enthalpy of (a) ibuprofen, (b) paracetamol, (c) salicylic acid and (d) benzoic acid as a function of the melting temperature.

The experimental solubility data obtained in this work and from the literature are given in Fig. 3. The experiment errors on the solubility are from 1% to 4%.

These solubility data will be used to test qualitatively and quantitatively the thermodynamic models accuracy. The chosen criterion chose to qualitatively evaluate the models is the preservation of the solubility ranking of an API in various solvents (Table 2). This scale is defined, in this work, as an ordered list of solvents from the solvent in which the API is the least soluble to most soluble. Quantitatively, the model accuracy is evaluated by the mean square error between the model prediction and the experimental mole fraction solubility. The sensitivity of the models prediction to the solid state properties ( $\Delta H_m$ ,  $T_m$  and  $\Delta C_p$ ) have also been taken into account.

#### 4.2. Original UNIFAC and modified UNIFAC

The mean square relative errors obtained for the solubility prediction using UNIFAC and UNIFAC modified for all the solutes and solvents tested are reported in Table 3. The experimental data used are the same as in Fig. 3. For each solubility prediction, standard deviation will be computed using a Monte Carlo method as explained previously. These mean relative errors have been calculated using the following equation:

$$mse = \frac{1}{n} \sum_{i=1}^n \left( \frac{X_i \text{ predicted} - X_i \text{ experimental}}{X_i \text{ experimental}} \right)^2$$

The mean square relative errors (mse) obtained range from 0.07 to 26.62 for UNIFAC and from 0.137 to 322 for UNIFAC mod. Some predictions in particular solvents induced high mse values, so that the obtained numbers do not truly represent the quality of the models. For example, the upper part of the ranges decreases to 6.79 and 18.04 without chloroform. Despite the relative good results obtained in the case of ibuprofen, benzoic acid or anthracene, it can be stated that UNIFAC seems not able to predict quantitatively the solubility. But, it is really interesting to compare the solubility predicted orders of magnitude to the experimental, and in this way to check if the methods give good approximations or not.

In addition, the predictions obtained at lower temperatures (20–25 °C) are much better than those obtained at higher temperatures. The original UNIFAC solubility predictions of the considered compounds as a function of the experimental are given in Fig. 4. The behaviour observed for UNIFAC mod. is roughly the same. In this paper, we use the term UNIFAC to discuss about both these models. In this figure, for one solute in one solvent, it can be observed that the more the temperature increases, the less the predicted solubility is accurate. In addition, the temperature dependence of the binary parameters is introduced in UNIFAC, by a polynomial  $\tau_{ij} = a_{ij}T^2 + b_{ij}T + c_{ij}$ . Even if this approach gives good results for small molecules and for VLE, it cannot represent the complexity of the evolution of interactions and volumetric properties (density for example) with temperature.

UNIFAC is a pair-wise additive model, thus the temperature evolution of the specific interactions (like hydrogen bonds) cannot be truly represented. Moreover, even if the binary interaction param-

**Table 2**

Experimental solubility logarithm ranking at 30 °C and orders of magnitude (in solubility logarithm) for ibuprofen, paracetamol, salicylic acid and benzoic acid in various organic solvents.

Ibuprofen		Paracetamol	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-3; -2]	Dichloromethane	[-9; -8]
Cyclohexane	[-2; -1.5]	Ethyl acetate	[-6; -5]
Ethanol	[-1.5; -1]	Acetonitrile	[-5; -4]
Toluene		Dioxane	
Ethyl acetate		Chloroform	
Isopropanol		Heptanol	[-4; -3]
Acetone		Methyl ethyl ketone	
Octanol	[-1; -0.5]	Acetone	
Chloroform		Butanol	
		Propanol	[-3; -2]
		Ethanol	
		Methanol	
		Dimethyl sulfoxide	[-1]
Benzoic acid		Salicylic acid	
Solvent	Order of magnitude	Solvent	Order of magnitude
Hexane	[-5; -4]	Cyclohexane	[-8; -7]
Cyclohexane		Hexane	
Heptane		Chloroform	[-7; -6]
Acetonitrile	[-3; -2]	Carbon tetrachloride	[-6; -5]
Caron tetrachloride		Xylene	
Benzene		Acetonitrile	[-4; -3]
Heptanol	[-2; -1.5]	Acetic acid	[-3; -2]
Acetone		Methanol	[-2; -1.5]
Octanol		Ethyl acetate	
Isopropanol		Ethanol	
Butanol		Octanol	
N-Methyl pyrrolidone	[-1.5; -1]	Acetone	
Dioxane		Ethyl methyl ketone	

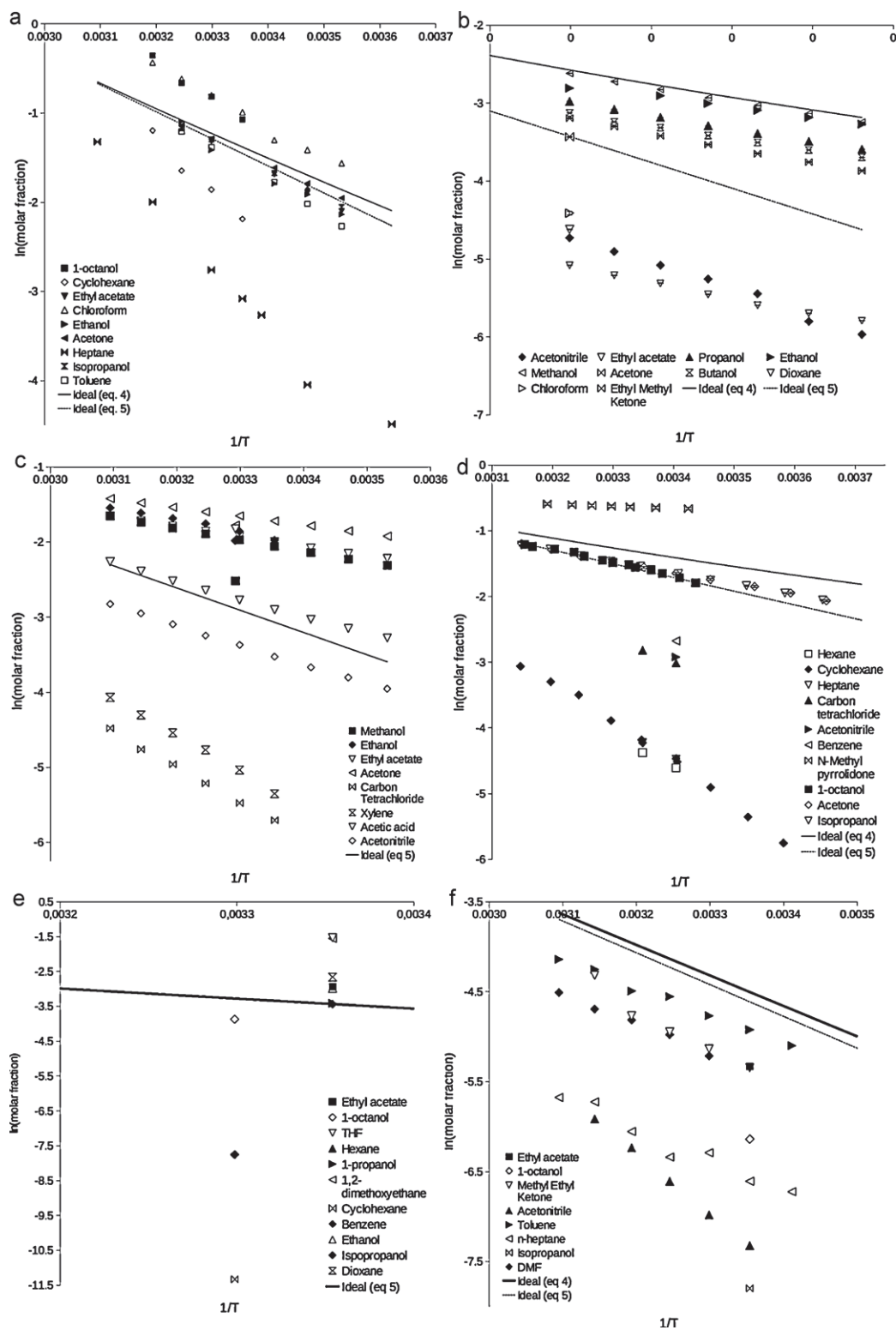
**Table 3**

Prediction errors of UNIFAC, COSMO-SAC and NRTL-SAC for ibuprofen, paracetamol, benzoic acid and salicylic acid using Eqs. (4) and 5, using solubility data in Table 2.

Compound	Model	MS error Eq. (5)	Mean standard deviation	MS error Eq. (4)
Ibuprofen	UNIFAC	0.07	11%	0.104
	UNIFAC mod.	0.137	15%	0.079
	COSMO-SAC	1.150	8%	1.307
	NRTL-SAC lit.	0.201	12%	0.276
	NRTL-SAC this work 1	0.202	10%	-
	NRTL-SAC this work 2	0.165	13%	-
Paracetamol	UNIFAC	2.306	31%	58.67
	UNIFAC mod.	1.165	30%	3.780 (0.754 <sup>b</sup> )
	COSMO-SAC	25.187	19%	309 (166 <sup>b</sup> )
	NRTL-SAC lit.	0.367	26%	0.509
	NRTL-SAC this work 1	76 (0.9 <sup>b</sup> )	-	-
	NRTL-SAC this work 2	0.805 (0.250 <sup>b</sup> )	27%	-
Salicylic acid	NRTL-SAC this work 3	9.034 (2.750 <sup>b</sup> )	-	-
	UNIFAC	26.62 (0.195 <sup>a</sup> )	29%	-
	UNIFAC mod.	322 (3.01 <sup>a</sup> )	36%	-
	COSMO-SAC	15.03 (5.01 <sup>a</sup> )	66%	-
	NRTL-SAC lit.	125 (0.451 <sup>a</sup> )	44%	-
	NRTL-SAC this work	361 (0.436 <sup>a</sup> )	49%	-
Benzoic acid	NRTL-SAC this work	409 (0.590 <sup>a</sup> )	63%	-
	UNIFAC	0.180	9%	1.077
	UNIFAC mod.	0.439	8%	0.572
	COSMO-SAC	1.136	33%	1.932
	NRTL-SAC lit.	0.107	9%	-
	NRTL-SAC this work	0.114	9%	-
4-Aminobenzoic acid	UNIFAC	6.790	-	-
	UNIFAC mod.	18.04	-	-
	COSMO-SAC	8.50	-	-
	NRTL-SAC this work	0.672	-	-
Anthracene	UNIFAC	0.247	11%	0.506
	UNIFAC mod.	0.808	12%	1.388
	COSMO-SAC	3.203	19%	3.80
	NRTL-SAC this work	3.631	19%	-

<sup>a</sup> mse without chloroform.

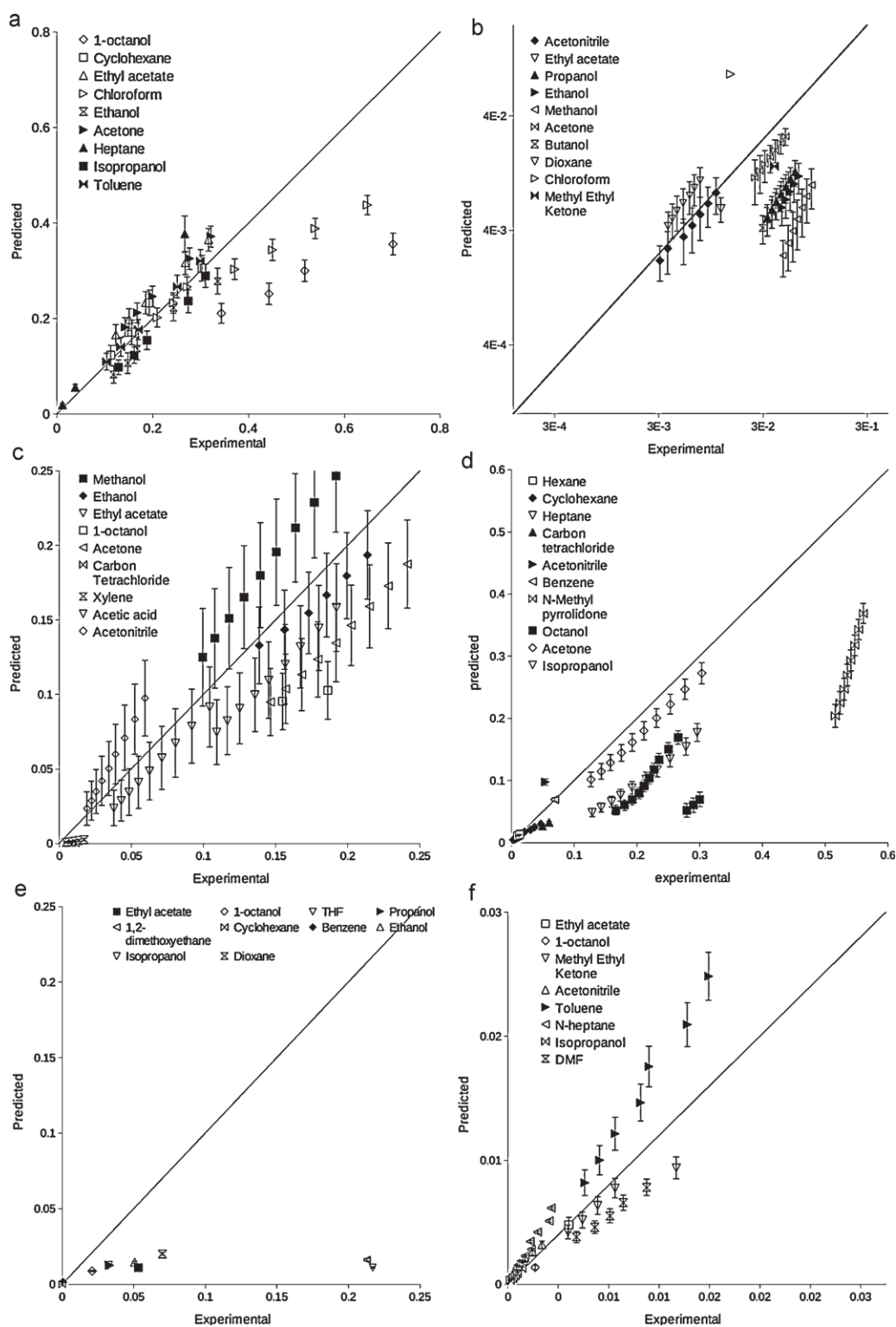
<sup>b</sup> mse without dichloromethane.



**Fig. 3.** Experimental solubility of (a) ibuprofen [20,21], (b) paracetamol [22], (c) salicylic acid [23–27], (d) benzoic acid [28–30], (e) 4-aminobenzoic acid [31–34] and (f) anthracene [35–39] in various solvents as a function of the reverse temperature (van't Hoff plot).

eters are temperature dependent ( $\Psi_{mn}$ ), the molecular geometry and isomerism are not taken into account. As mentioned in the literature [5,8], the predicted solubility of organic molecules in non polar solvents (heptane, cyclohexane, and toluene) are in good agreement with the experimental data. In the case of polar solvents, capable of forming hydrogen bonds, we found the solubility to be always underestimated for the chosen solvents. In addition,

in the case of solubility in alcohols, the orders of magnitude of the predicted solubilities are very close. This probably means that in UNIFAC, the pair-wise interaction coefficients involving the OH group is predominant over the interaction coefficients of the aliphatic chains ( $-\text{CH}_2-\text{CH}_3$ ,  $-\text{CH}_3$  groups). The solubility orders of magnitude, shown in Table 4, seems to confirm this statement. From a qualitative point of view, the obtained solubility ranking



**Fig. 4.** UNIFAC predictions of (a) ibuprofen, (b) paracetamol, (c) salicylic acid, (d) benzoic acid, (e) 4-aminobenzoic acid and (f) anthracene with Eq. (5) as a function of the experimental solubility.

of the studied molecules are given in Table 4. This table shows that for large solubility values ( $x > 0.05$ ), UNIFAC gives satisfactory qualitative results for predicting solubility order of magnitude and the solubility ranking is almost preserved. For paracetamol, original UNIFAC has difficulties to predict solubility lower than 0.05. The predicted orders of magnitude are not in agreement with experimental data.

#### 4.3. COSMO-SAC

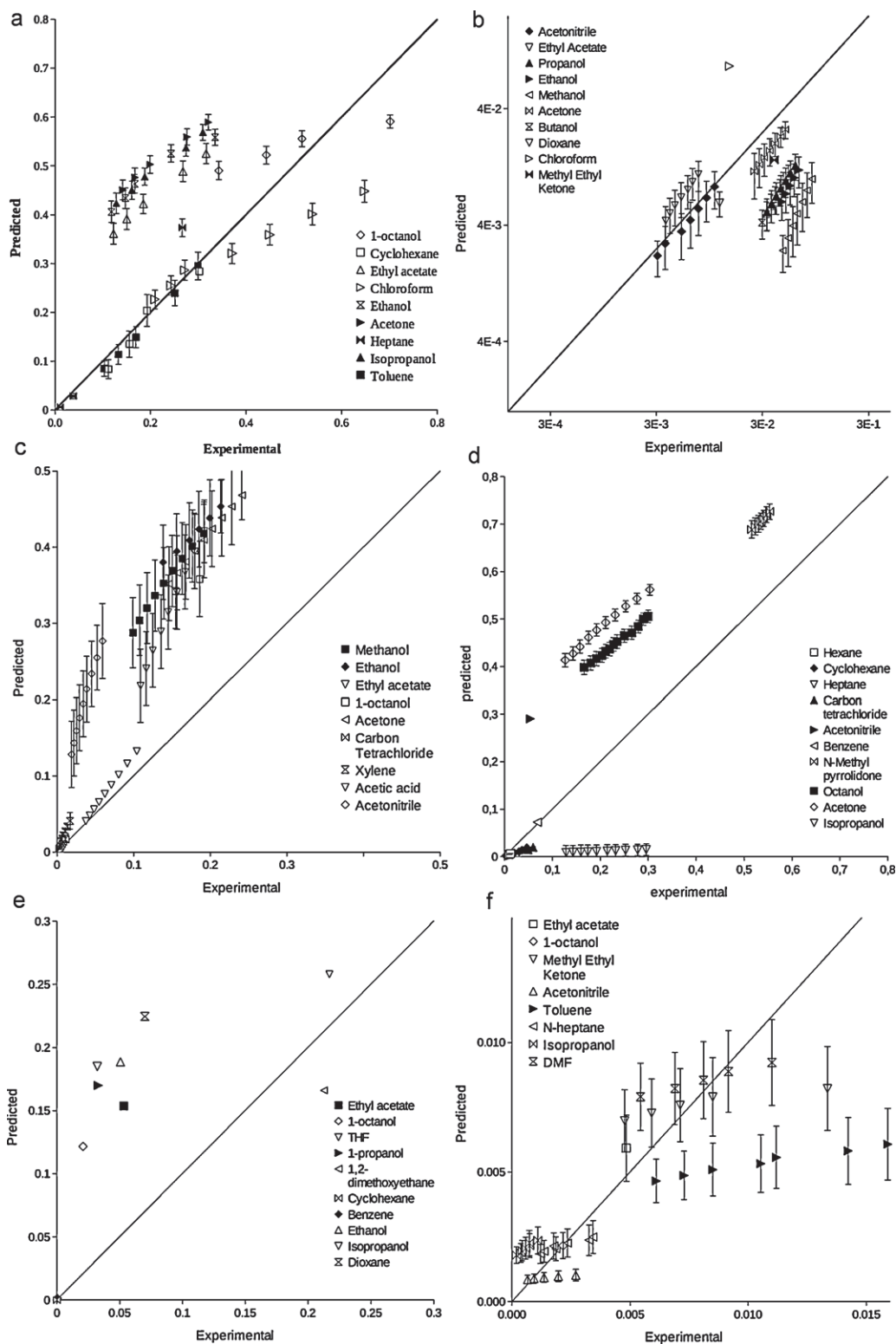
From the results, presented in Fig. 5 and Table 5, it can be stated that COSMO-SAC gives poor results in the solubility prediction of the drugs tested. The mean square errors range from 1.15 to 25, depending on the molecule and on the pure component properties chosen. Not surprisingly, the relative deviation is particularly

**Table 4**  
Original UNIFAC solubility ranking at C or 30 °C and orders of magnitude (in solubility logarithm) for ibuprofen, paracetamol, salicylic acid and benzoic acid in various solvents.

Ibuprofen		Paracetamol	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-3; -2]	Dichloromethane	[-9; -8]
Cyclohexane	[-2; -1.5]	Dioxane	[-5]
Ethanol		Heptanol	
Isopropanol	[-1.5; -1]	Acetonitrile	[-5; -4]
Octanol		DMSO	
Toluene		Methanol	
Ethyl acetate		Butanol	
Acetone		Ethyl acetate	
Chloroform		Propanol	
		Methyl ethyl ketone	
		Ethanol	
		Acetone	[-4; -3]
		Chloroform	[-3; -2]
Benzoic acid		Salicylic acid	
Solvent	Order of magnitude	Solvent	Order of magnitude
Hexane	[-5; -4]	Hexane	[-9; -8]
Heptane		Cyclohexane	
Cyclohexane		Carbon tetrachloride	
Carbon tetrachloride	[-4; -3]	Xylene	[-7; -6]
Benzene	[-3; -2]	Chloroform	[-3]
Octanol		Acetic acid	
Isopropanol		Acetonitrile	
Acetonitrile		Octanol	[-3; -2]
Butanol		Ethyl acetate	
Heptanol		Ethyl methyl ketone	
Dioxane	[-2; -1]	Acetone	[-2; -1.5]
Acetone		Ethanol	
N-Methyl pyrrolidone		Methanol	

**Table 5**  
COSMO-SAC solubility ranking at 30 °C and orders of magnitude (in solubility logarithm) for ibuprofen, paracetamol, salicylic acid and benzoic acid in various solvents.

Ibuprofen		Paracetamol	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-3; -2.5]	Dichloromethane	[-6; -5]
Cyclohexane	[-2]	Chloroform	[-5; -4]
Toluene	[-1.5; -1]	Acetonitrile	[-3; -2.5]
Chloroform		Ethyl acetate	
Ethyl acetate	[-1; -0.5]	Heptanol	
Octanol		Butanol	[-2.5; -2]
Ethanol		Methanol	
Isopropanol		Propanol	
Acetone		Dioxane	[-2; -1.5]
		Ethanol	
		Methyl ethyl mketone	
		Acetone	
		Dimethylsulfoxide	[> -1]
Benzoic acid		Salicylic acid	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-5.5; -5]	Hexane	[-6; -5]
Cyclohexane		Cyclohexane	
Hexane		Carbon tetrachloride	
Isopropanol	[-5; -4]	Xylene	[-4]
Carbon tetrachloride	[-4; -3]	Chloroform	[-4; -3]
Benzene	[-3; -2]	Acetic acid	[-3; -2]
Acetonitrile	[-2; -1]	Acetonitrile	[-2; -1.5]
Octanol	[-1; -0.5]	Ethyl acetate	[-1.5; -1]
Butanol		Methanol	
Dioxane		Octanol	
Acetone		Ethanol	[> -1]
N-Methylpyrrolidone	[> -0.5]	Acetone	
Dimethylsulfoxide		Ethyl methyl ketone	



**Fig. 5.** COSMO-SAC prediction of (a) ibuprofen, (b) paracetamol, (c) salicylic acid, (d) benzoic acid, (e) 4-aminobenzoic acid and (f) anthracene with Eq. (5) as a function of experimental solubility.

high in solvents in which solubility is very low, e.g. paracetamol in ethyl acetate. As in the case of UNIFAC, the model gives satisfactory results with aprotic and apolar solvents (hexane, heptane, and chloroform). On the opposite, in polar solvents capable of forming hydrogen bonds (alcohols, ketones, esters, ...) with the solute, the predicted solubility is systematically overestimated. In addition, it can be noticed that the predicted solubility values of a molecule

in alcohols are always nearly the same, whatever the size of the alcohol molecule (see orders of magnitudes in Table 5).

As previously achieved for the UNIFAC model, the predictions obtained at the lower temperatures (20–25 °C) are much better than the one obtained at higher temperatures (Fig. 5). In this model, the electrostatic and hydrogen bonds parameters are not temperature dependent.

From a qualitative point of view (Table 5 and Fig. 5), COSMO-SAC is quite reliable for the solubility prediction of the most simple molecules, like alkanes. This model overestimates solubility in alcohols and ketones and leads to a non reliable solubility scale. In addition, the orders of magnitude are twice the experimental in most cases. Considering solubility ranking, this model is not as good as UNIFAC. To evaluate the influence of hydrogen bond on the predicting capabilities of COSMO-SAC, the hydrogen bond is excluded from the model, setting the constant for the hydrogen bonding interaction ( $c_{hb}$ ) to zero. This constant appears in the calculation of the segment activity  $\Gamma$ , in the expression of the electrostatic interactions.

The results are shown in Fig. 6 and it can be seen that the predicted solubilities are underestimated in polar solvents. The model predictions are in good agreement with the solubility of polar molecules in polar solvents, which is higher than ideal, mainly results from specific interactions like hydrogen bonds. In addition, the results obtained with and without hydrogen bonds also suggest that in COSMO-SAC the hydrogen bond contribution is overestimated. The  $c_{hb}$  constant is an adjustable parameter which has been optimized upon many VLE data of small organic molecules. However, the organic molecules in our study are much larger, flexible, and contain h-bonding surfaces that are burried or hindered so that h-bonds cannot form. Then, it is likely that such surfaces shall not be included in the h-bond term in the model. It would be very interesting to re-optimize this term for large organic molecules using SLE data (providing that enough experimental data are available). Even if Eq. (4) is considered, this does not improve the results as the  $\Delta C_p$  term increases the predicted solubility (see Section 4.5). In addition, even if stereoisomerism can be taken into account in COSMO-SAC (each isomer has its own sigma profile), molecular self and cross associations are not considered. It is well known that ibuprofen forms hydrogen-bonded dimer in solution: two composed of the same enantiomers of R-R and S-S and the racemate of R-S. The self-association of drug molecules may decrease the number of free molecular sites available to form hydrogen bonds with the solvent, which may explain the overestimation of the contribution of hydrogen bonds to the calculation of the activity coefficient. Such effects are also found for benzoic acid (Self-association and hydration of

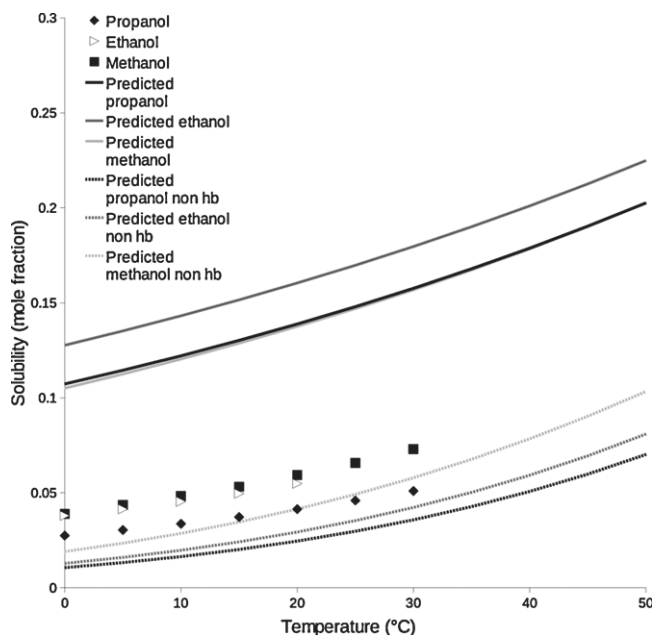


Fig. 6. COSMO-SAC prediction of paracetamol in ethanol and methanol with hydrogen bonds (straight lines), and no hydrogen bonds (dashed lines) as a function of the temperature using Eq. (5).

benzoic acid in benzene [54] or in a less extent for carboxylic acid like aspirin).

At last, even if the predicted solubility is quantitatively far from the experimental data, Fig. 6 shows that the model nicely predicts the solubility temperature dependence. For designing a crystallization process, the one of the most important parameter is the supersaturation: the driving force of the process, defined as the ratio between the initial and final concentration ( $S=C/C^*$ ). For instance, as shown in Fig. 6, the experimental supersaturation for paracetamol in methanol from 30 °C to 0 °C is  $S=1.54$  and the predicted is  $S=1.5$  (similar results are obtained for all the molecules tested in the paper in small alcohol molecules).

Table 6  
NRTL-SAC segments when  $\Delta C_p$  is neglected (Eq. (5)).

Compounds	X	Y-	Y+	Z	mse	Number of solvents
Ibuprofen [55]	1.038	0.051	0.028	0.318	1.055	19
Ibuprofen (this work 1) <sup>a</sup>	0.507	0.297	0.350	0	0.005	6
Ibuprofen (this work 2) <sup>b</sup>	0.484	0	0.267	0.210	0.0188	9
Paracetamol [56]	0.498	0.487	0.162	1.270	-	8
Paracetamol [7]	0.416	0.016	0.168	1.86	-	5
Paracetamol (this work 1) <sup>c</sup>	0.265	0.576	0.668	0.717	0.029	6
Paracetamol (this work 2) <sup>d</sup>	0.369	0.618	0.521	1.095	0.859	14
Paracetamol (this work 3) <sup>e</sup>	0.275	1.036	1.159	0.777	0.033	6
Salicylic acid [55]	0.726	0.176	0	0.749	0.774	18
Salicylic acid (this work 1) <sup>f</sup>	0.911	1.762	0.129	0.329	0.182	6
Salicylic acid (this work 2) <sup>g</sup>	1.006	2.29	0	0.451	0.0617	6
Benzoic acid [17]	0.524	0.089	0.45	0.405	0.160	7
Benzoic acid [55]	0.494	0	0.336	0.468	0.292	26
Benzoic acid (this work) <sup>h</sup>	0.576	0	0.121	0.543	0.144	6
4-aminobenzoic acid <sup>i</sup>	0	0	4.229	2.321	0.222	5
Anthracene <sup>j</sup>	1.486	0	0	0.564	0.180	5

<sup>a</sup> In acetone, cyclohexane, chloroform, ethanol, octanol and ethyl acetate at 30 °C.

<sup>b</sup> In all the solvent considered in this study (see Fig. 3) at 30 °C.

<sup>c</sup> In acetone, ethyl acetate, chloroform, toluene, ethanol and water at 30 °C.

<sup>d</sup> In all the solvents of this work (see Fig. 3) at 30 °C.

<sup>e</sup> In ethanol, propanol, acetone, methyl ethyl ketone, water and toluene at 30 °C.

<sup>f</sup> In ethanol, ethyl acetate, cyclohexane, acetone, acetonitrile and xylene at 30 °C.

<sup>g</sup> In ethanol, ethyl acetate, acetone, xylene, acetonitrile and water at 30 °C.

<sup>h</sup> In cyclohexane, acetonitrile, 1-butanol, benzene, n-methyl pyrrolidone and water, at 30 °C.

<sup>i</sup> In ethyl acetate, ethanol, dimethoxyethane, THF and dioxane, at 30 °C.

<sup>j</sup> In heptane, acetonitrile, 1-octanol, methylethylketone, ethyl acetate, toluene and DMF at 30 °C.

#### 4.4. NRTL-SAC

To test the prediction abilities of NRTL-SAC, the segments values ( $X$ ,  $Y^-$ ,  $Y^+$  and  $Z$ ) were taken from the literature [7,55,56]. It can be noticed that segment values found in the literature can significantly differ for the same compound. For instance, as shown in Table 6, Chen and Crafts [56] and Mota et al. [7] have published different segment values for paracetamol. These differences can be not only due to the values used for the pure component parameters ( $\Delta H_m$ ,  $T_m$ , and  $\Delta C_p$ ) determined experimentally or regressed simultaneously with the segments, but also due to the nature and the number of solvents used in the regression.

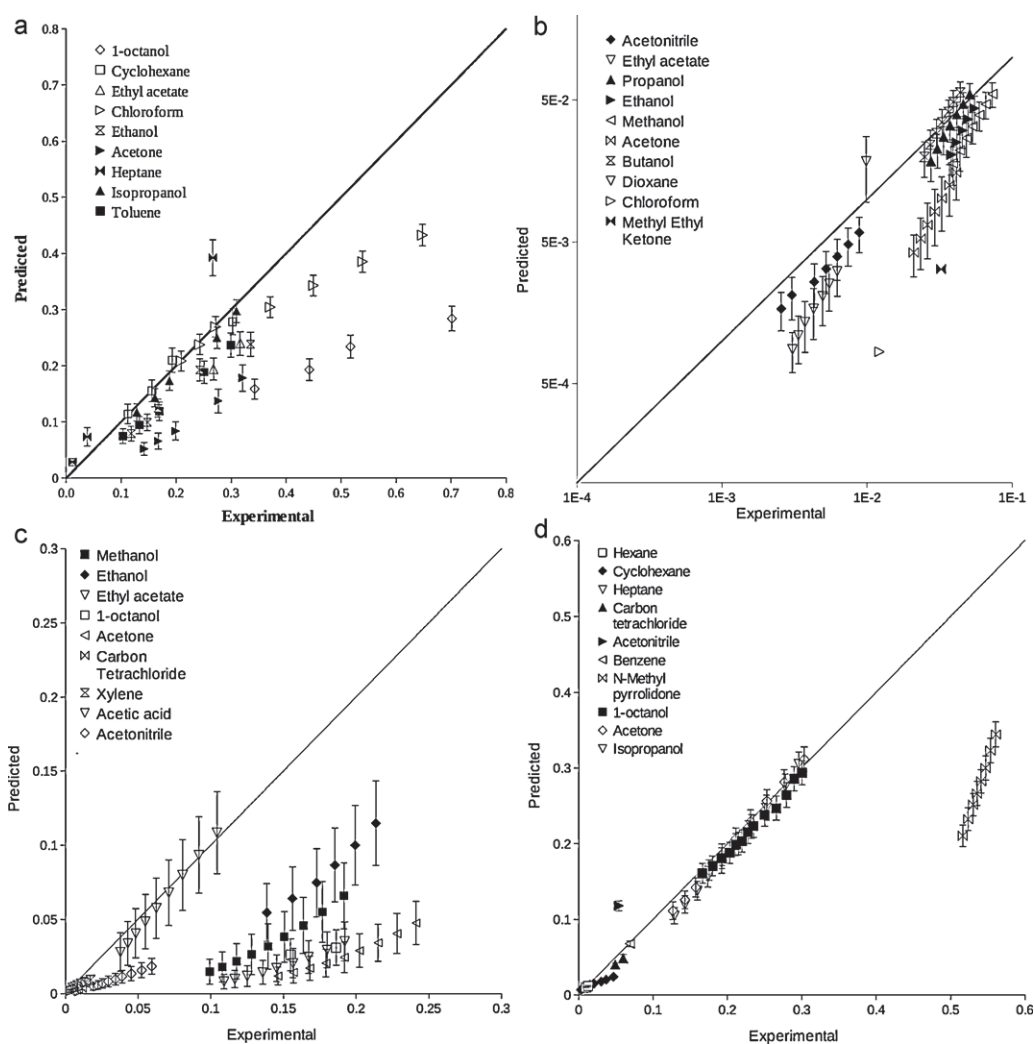
In order to test the model sensitivity to the quadruplet [ $X$   $Y^-$   $Y^+$   $Z$ ], calculations were performed with different solvents for all the solutes in order to obtain new values. All these new values and the corresponding mean square errors are given in Table 6. The sensitivity of the model to the experimental solid state properties is performed using the Monte Carlo method.

From a quantitative point of view, the model gives results with a mse ranging from 0.107 to 409 or 3.63 when neither chloroform nor dichloromethane are considered (see Table 3). NRTL-SAC predictions using [55] segment values versus experimental solubility is represented in Fig. 7, and using segment from this work in Fig. 8 (quadruplet 1 for ibuprofen, quadruplet 2 for paracetamol and

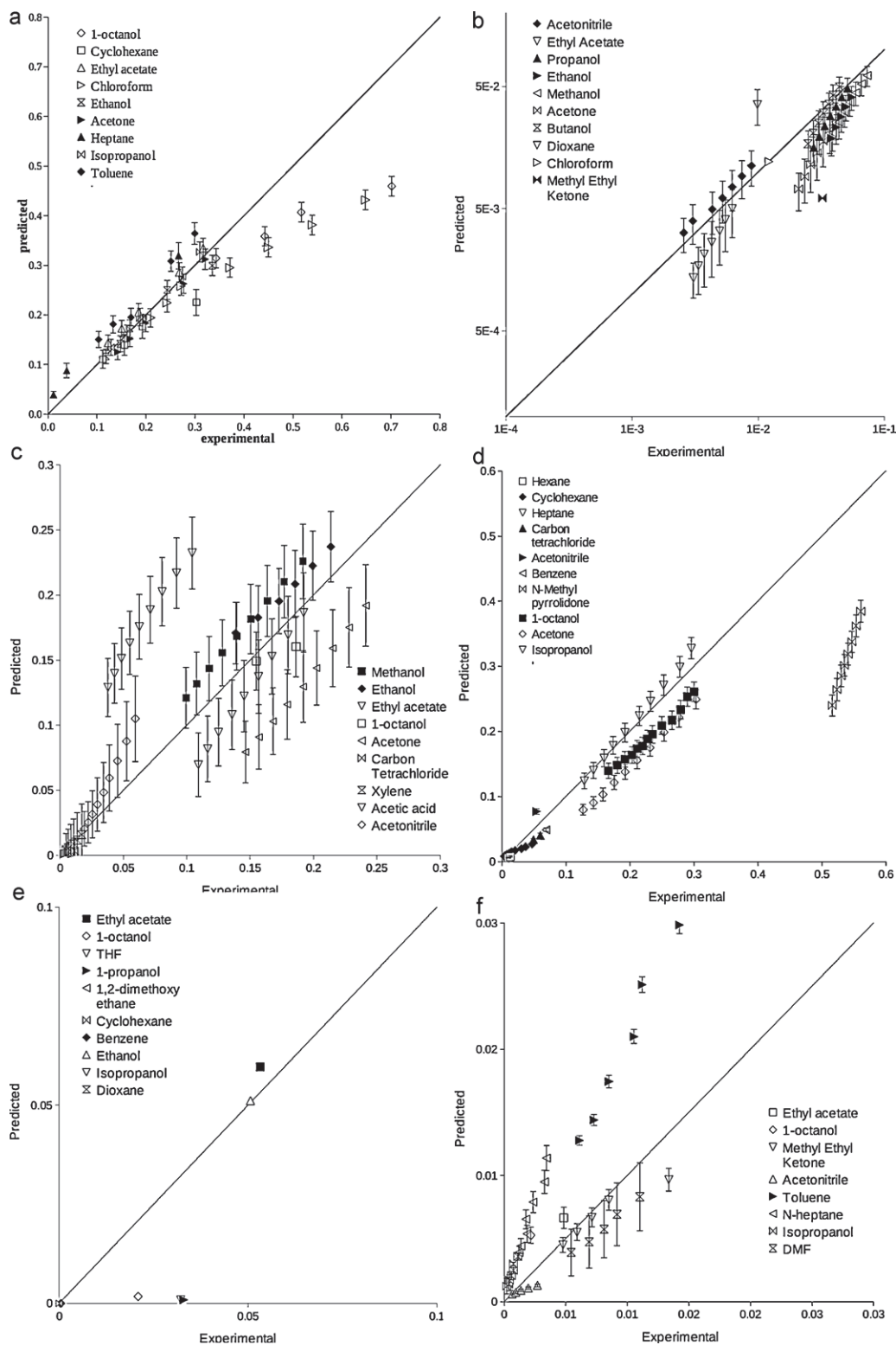
quadruplet 2 for salicylic acid). Due to the empirical nature of the model, NRTL-SAC seems to be less affected by errors on the thermodynamic properties than the other models. The results obtained are in good agreement with the experimental data, and most of the results are beneath a mse of 4. However, the ms errors seem quite high to do quantitative predictions.

The model accuracy of the results strongly depends on the quadruplets [ $X$   $Y^-$   $Y^+$   $Z$ ] used as parameters. In fact, the choice of these parameters is a critical step to use this model.

The segments values can be understood as a weight given to each molecule behaviour: hydrophilic, hydrophobic, polar attractive and repulsive. They represent a "mean behaviour" of the molecule in solution computed using experimental data. These values depend on the quality and quantity of the solubility data used to regress the four parameters. They will not represent the real behaviour of the API in all the solvents if the experimental data are not taken judiciously. However, the choice of the experimental data is not obvious. NRTL-SAC using segments calculated in some alcohols will not necessarily give good predictions in other alcohols and it gives even worse results in other solvent like ketones or alkanes. Some tests performed on paracetamol showed that the predictions obtained using the parameters regressed using four solubility data in propanol and butanol (two temperatures) gave the same results as those obtained with quadruplet 3 of this work. In fact, all alcohols



**Fig. 7.** NRTL-SAC prediction of (a) ibuprofen, (b) paracetamol, (c) salicylic acid, (d) benzoic acid, (e) 4-aminobenzoic acid and (f) anthracene with Eq. (5) using segments from the literature as a function of experimental solubility.



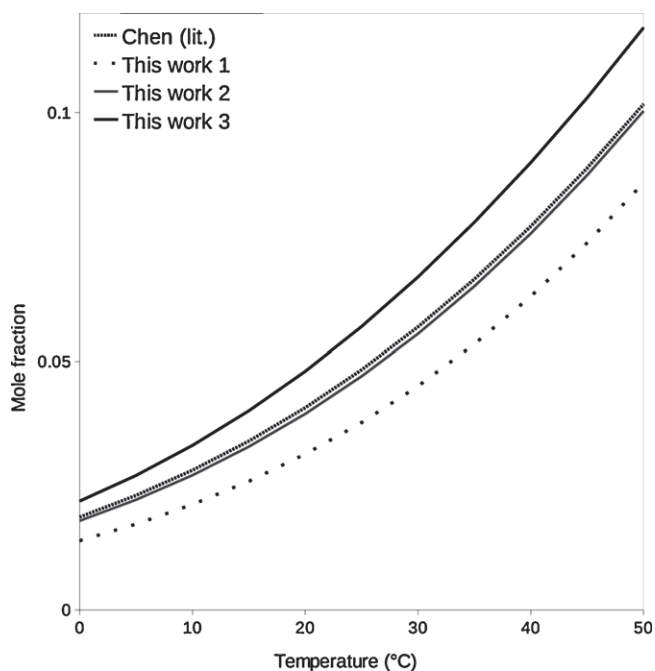
**Fig. 8.** NRTL-SAC prediction of (a) ibuprofen, (b) paracetamol, (c) salicylic acid, (d) benzoic acid, (e) 4-aminobenzoic acid and (f) anthracene with Eq. (5) using segments from this work (this work 1 for ibuprofen, 2 for paracetamol, 1 for salicylic acid and 1 for benzoic acid) as a function of experimental solubility.

do not have the same weight on each segment number. Even if their quadruplets may indicate a close behaviour tendency, they can also be quite different. In addition, the use of a large experimental data set increases the model performance and reliability.

In order to show the influence of the segment values used in this study, four types of ibuprofen solubility predictions in ethanol are reported in Fig. 9.

To end with the solvent selection and in order to get the most reliable regression, the idea is to select the solvent so that the weight of each segment is the same:

$$\sum_i X_i = \sum_i Y_i^+ = \sum_i Y_i^- = \sum_i Z_i \quad (17)$$



**Fig. 9.** NRTL-SAC prediction of paracetamol in ethanol using four different segment quadruplets as a function of the temperature.

with  $i$  the subscript for the solvent. In this way, all the behaviour of the solute molecules will be investigated with the same weight.

This method is still in study, but we tried to find the best solvents for the paracetamol regression. To do that, we take all the solvents investigated by [17] and we forced this sum to be greater than 1 (to have each segment number represented with enough weight). The calculation, made with version 23.6.5 of GAMS software, gave a total of 10 solvents (chloroform, 1,2-dichloroethane, 1,4-dioxane, ethanol, formamide, formic acid, isobutylalcohol, nitromethane, 1-propanol and water) to reach a maximum deviation of 0.002. If we regressed the segment numbers of these solvents, the mean square error would be higher than the mse obtained in Table 6 because of the diversity of the solvent used. But, the obtained quadruplet should be more suitable as the mean behaviour should be better represented. Further studies are still in progress.

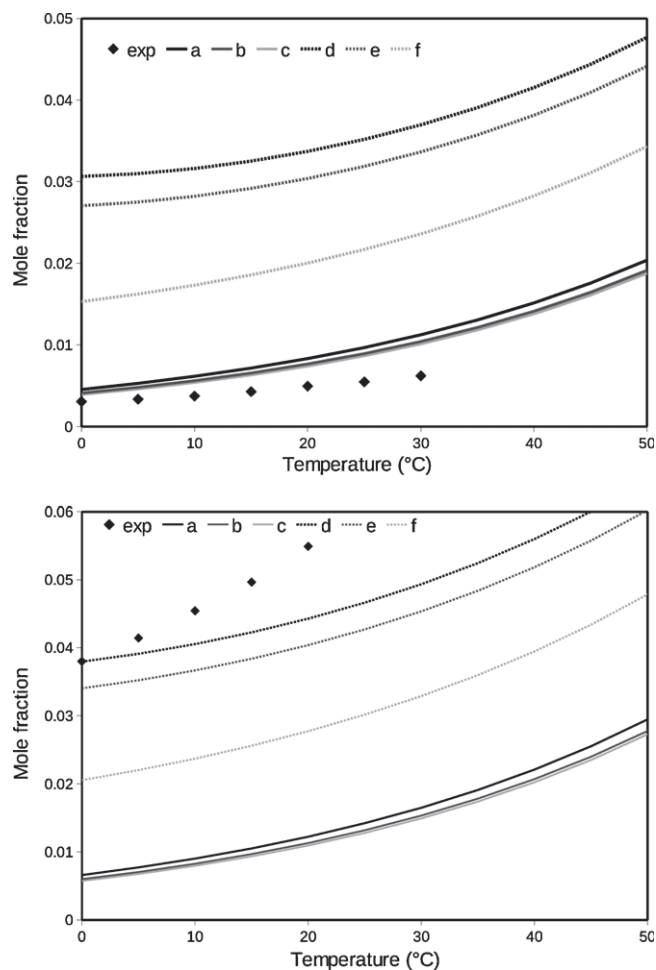
Even if the prediction differs quantitatively from a quadruplet to the other, NRTL-SAC gives good orders of magnitude whatever the segments used (see Tables 7 and 8). Moreover the solubility ranking is preserved.

As shown in Figs. 7 and 8 in the case of ibuprofen, solubility predictions for temperatures above 25 °C are less accurate. In this model, the influence of temperature on solubility is underestimated, especially when hydrogen bonds are involved. This could be explained by the fact that the solubility dependence on temperature being related to the solid state properties, and not on the possible evolution of molecular interactions (hydrogen bond, Van der Waals, dispersion forces, ...) between molecules in solution.

To conclude on NRTL-SAC, this model can be useful to give an estimation of the solubility of APIs in various solvents. But it needs reliable segment values for quantitative purpose and needs experimental data.

#### 4.5. The right melting enthalpy or temperature and the use of $\Delta C_p$

In the case of SLE, equilibrium equation can be Eq. (4) or Eq. (5). Eq. (4) is the most rigorous, but, usually, the  $\Delta C_p$  term



**Fig. 10.** UNIFAC prediction of paracetamol using different thermodynamic properties [23,34,35,37] in ethanol (upper side) and ethyl acetate (down side) with (a)  $\Delta H_m = 27$ , 100 J/mol,  $T_m = 443.6$  K, (b)  $\Delta H_m = 27$ , 100 J/mol,  $T_m = 443.15$  K, (c)  $\Delta H_m = 28$ , 100 J/mol,  $T_m = 441.15$  K, (d) same as 'b' with  $\Delta C_p = 99.8$  J/mol K, (e) same as 'c' with  $\Delta C_p = 99.8$  J/mol K, (f) same as 'a' with  $\Delta C_p = 75$  J/mol K.

is neglected and Eq. (5) is used. Even if Gracin et al. [5] have found a small influence of the  $\Delta C_p$  term with UNIFAC model, the two equations will be compared using all the models presented previously. In addition, the significance of the values of the pure compound properties has to be underlined. Measuring accurately the melting temperature and enthalpy may be difficult, since it relies on product crystallinity, purity and on the analytical method used. As presented in Table 1 many different values have been reported in the literature. In Figs. 4, 5, 7 and 8, standard deviations of the models caused by properties errors have been presented. So, when a model is used for predicting solubility, its error has to be taken into account, and the measured  $T_m$ ,  $\Delta H_m$  and  $\Delta C_p$  have to be as accurate as possible. Nevertheless, prediction errors induced by the models are usually larger than the ones induced by the solid state property uncertainties.

As shown in Fig. 10, it can be seen that a higher melting enthalpy implies a higher predicted solubility. The same behaviour is observed for higher  $\Delta C_p$ . Moreover, in the case of polymorphism (like paracetamol), the use of the wrong melting enthalpy, temperature or  $\Delta C_p$  can cause bad predictions or misinterpretations (paracetamol presents at least two polymorphs with close melting temperature and enthalpy).

**Table 7**  
NRTL-SAC solubility ranking at 30 °C and orders of magnitude (in solubility logarithm) for ibuprofen, paracetamol, salicylic acid and benzoic acid in various solvents using segments from the literature.

Ibuprofen		Paracetamol	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-2]	Dichloromethane	[-8; -7]
Acetone	[-2; -1.5]	Chloroform	
Cyclohexane		Ethyl acetate	[-6; -5]
Toluene		Methylethylketone	
Octanol		Acetonitrile	
Ethanol		Acetone	[-5; -4]
Ethyl acetate		Dioxane	
Isopropanol	[-1.5; -1]	Propanol	[-3; -2.5]
Chloroform	[ $3 \times 10^{-1}$ to $4 \times 10^{-1}$ ]	Methanol	
		Butanol	
		Ethanol	
		Dimethylsulfoxide	[-2.5; -2]
Benzoic acid		Salicylic acid	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-5; -4]	Hexane	[-8; -7]
Hexane		Cyclohexane	
Cyclohexane		Carbon tetrachloride	[-7; -6]
Carbon tetrachloride	-3	Xylene	[-6; -5]
Benzene	[-2.5; -2]	Acetonitrile	[-4; -3]
Acetonitrile		Ethyl methyl ketone	
Butanol	[-2; -1.5]	Ethyl acetate	
Octanol		Acetone	[-4; -3]
Isopropanol		Octanol	
Acetone		Methanol	
Dioxane	[-1.5; -1]	Acetic acid	[-3; -2]
N-Methyl pyrrolidone		Ethanol	
Dimethylsulfoxide	[> -1]	Chloroform	

**Table 8**  
NRTL-SAC solubility ranking at 30 °C and orders of magnitude (in solubility logarithm) for ibuprofen, paracetamol, salicylic acid and benzoic acid in various solvents using segments calculated in this study (this work 1 for ibuprofen, 2 for paracetamol, 1 for salicylic acid and 1 for benzoic acid).

Ibuprofen (this work 2)		Paracetamol (this work 2)	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	-2	Dichloromethane	[-7; -6]
Cyclohexane		Ethyl acetate	[-6; -5]
Ethanol	[-1.5; -1]	Ethyl methyl ketone	
Toluene		Acetonitrile	[-5; -4]
Acetone		Chloroform	
Isopropanol		Acetone	[-4; -3]
Ethyl acetate		Dioxane	
Chloroform		Propanol	-3
Octanol		Butanol	
		Ethanol	[-3; -2.5]
		Methanol	
		DMSO	[-2.5; -2]
Benzoic acid (this work 2)		Salicylic acid	
Solvent	Order of magnitude	Solvent	Order of magnitude
Heptane	[-4.5; -4]	Hexane	[-9; -8]
Cyclohexane		Cyclohexane	[-8; -7]
Hexane		Carbon tetrachloride	[-6; -5]
Carbon tetrachloride	[-4; -3]	Xylene	[-5; -4]
Benzene	[-3; -2]	Acetonitrile	[-4; -3]
Acetonitrile		Ethyl acetate	[-3; -2]
Acetone	[-2; -1.5]	Acetone	
Octanol		Ethyl methyl ketone	
Dioxane		Methanol	[-2; -1.5]
Butanol		Octanol	
Isopropanol		Acetic acid	
N-methyl pyrrolidone	[-1.5; -1]	Ethanol	
Dimethylsulfoxide	[> -1]	Chloroform	

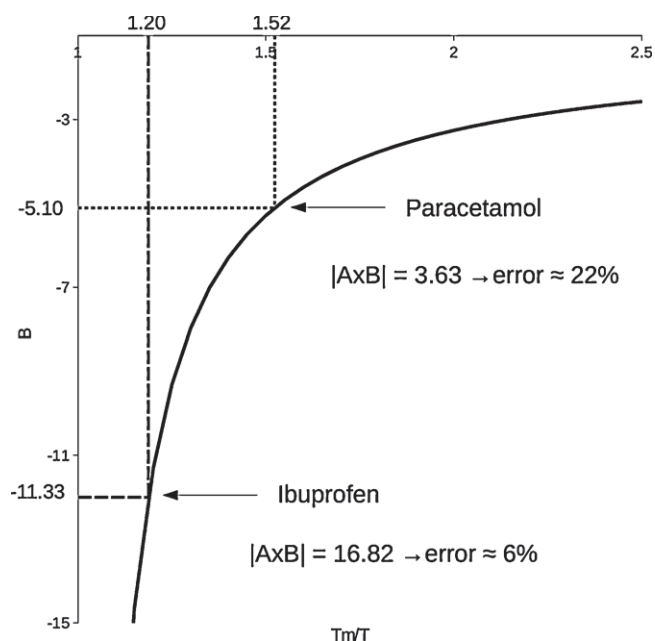


Fig. 11. Coefficient  $B$  as a function of  $T_m/T$ .

To evaluate the impact of neglecting the  $\Delta C_p$  term in the SLE on solubility prediction, relative value of the two parts, constituting Eq. (4), can be compared using the following equation:

$$\frac{\Delta h_m(T_m) \left( \frac{1}{T_m} - \frac{1}{T} \right)}{\Delta C_p(T_m) \left( \ln\left(\frac{T_m}{T}\right) - \frac{T_m}{T} + 1 \right)} \quad (18)$$

Then, we will rewrite:

$$\frac{\Delta H_m}{\Delta C_p T_m} \times \frac{1 - \frac{T_m}{T}}{\left[ \ln\left(\frac{T_m}{T}\right) - \frac{T_m}{T} + 1 \right]} = A \times B \quad (19)$$

If  $|AB| \ll 1$ , then the second part of Eq. (4) cannot be neglected, and if the ratio is high enough, it may be neglected (see example in Fig. 11). The  $A$  term is constant and only depends on the thermodynamic properties of the solid state. The  $B$  term is a function of temperature as shown in Fig. 11.

Considering that for a standard crystallization operation, the running temperature is around 290K. At this temperature,  $B = -11.33$  and  $A = 1.48$  for ibuprofen, and  $B = -5.10$ ,  $A = 0.71$  for paracetamol. The definition of a relative error enables to estimate the effect by neglecting the  $\Delta C_p$  term:

$$\text{relative error (\%)} = \frac{1}{AB - 1} \quad (20)$$

In the case of these two molecules, this error is about 6% for the ibuprofen and 22% for the paracetamol. We can conclude that the  $\Delta C_p$  term is negligible in the case of ibuprofen, and not in the case of paracetamol.

## 5. Conclusion

To conclude, none of the models presented in this paper is able to predict precisely the solubility of the studied molecules (containing various common functional groups like alcohols, ketones, amines) in various solvents. Even considering the model errors most of the times taken by the imprecision on the pure solid properties, the predicted solubility are generally in agreement with experimental data (Ibuprofen with Original UNIFAC, some solutes with NRTL-SAC). COSMO-SAC predictions usually produce higher values than the experimental, especially when hydrogen bonds may occur. UNI-

FAC is more accurate than COSMO-SAC for the API in this study and gives good results for simple molecules (anthracene). However, even if the solubility orders of magnitude are preserved, it is not good enough for performing quantitative predictions. In the case of NRTL-SAC, predictions are very dependent on the parameters used (segment values). If these parameters do not represent the correct behaviour of the API in the solvent chosen, the results will not be satisfying. Finally, UNIFAC and NRTL-SAC, which preserve the solubility orders of magnitude can be a good complement of experimental techniques to guide the choice of a crystallization solvent. But in the end, experiments are still required to obtain quantitative results.

In addition, all the models tested fail to correctly describe the solubility dependence with temperature. The use of an equation of state seems to be more appropriate for predicting the evolution of solubility with temperature [57,58].

Results presented in this paper also show that the use of the most precise equilibrium equation does not significantly influence the quality of the results. As the models are not accurate enough, the use of the most accurate thermodynamic properties values is not necessary. Moreover, the  $\Delta C_p$  term can be ignored as long as the generated error is negligible regarding the model error. But, in the hypothetical case of a perfect model, the most precise values should be used.

## Acknowledgements

The authors want to acknowledge ProSim Company for their collaboration in the use of Simulis Thermodynamics software, and Pr. Florent Bourgeois for enlightening discussions.

## References

- [1] S. Morissette, O. Almarsson, M. Peterson, J. Remenar, M. Read, A. Lemmo, S. Ellis, M. Cima, C. Gardner, High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids, *Advanced Drug Delivery Reviews* 56 (2003) 275–300.
- [2] P. Laval, N. Lisai, J.-B. Salmon, M. Joanicot, A microfluidic device based on droplet storage for screening solubility diagrams, *Lab on a Chip* 7 (2007) 829–834.
- [3] A.A. Noyes, W.R. Whitney, *Physical Chemistry* 23 (1909) 689.
- [4] J.G. Gmehling, T.F. Anderson, J.M. Prausnitz, Solid-liquid equilibria using UNIFAC, *Industrial & Engineering Chemistry Fundamentals* 17 (1978) 269–273.
- [5] S. Gracin, T. Brinck, A. Rasmuson, Prediction of solubility of solid organic compounds in solvents by UNIFAC, *Industrial & Engineering Chemistry Research* 41 (2002) 5114–5124.
- [6] E. Mullins, Y. Liu, A. Ghaderi, S.D. Fast, Sigma profile database for predicting solid solubility in pure and mixed solvent mixtures for organic pharmacological compounds with COSMO-based thermodynamic methods, *Industrial & Engineering Chemistry Research* 47 (2008) 1707–1725.
- [7] F.L. Mota, A.P. Carneiro, S.P. Pinho, E.A. Macedo, Temperature and solvent effect in solubility of some pharmaceutical compounds: measurements and modelling, *European Journal of Pharmaceutical Science* 37 (2009) 499–507.
- [8] I. Hahnenkamp, G. Graubner, J. Gmehling, Measurement and prediction of solubilities of active pharmaceutical ingredients, *International Journal of Pharmaceutics* 388 (2010) 73–81.
- [9] J. Mullin, in: *Crystallization—3 Rev*, Oxford, 1993.
- [10] M. Brown, B. Glass, Decomposition of solids accompanied by melting-Bawn kinetics, *International Journal of Pharmaceutics* 254 (2003) 255–261.
- [11] A. Fredenslund, R. Jones, J. Prausnitz, Group-contribution estimation of activity-coefficients in nonideal liquid-mixtures, *AIChE Journal* 21 (1975) 1086–1099.
- [12] U. Weidlich, J. Gmehling, A modified UNIFAC model. 1. Prediction of VLE, HE, and gamma-infinity, *Industrial & Engineering Chemistry Research* 26 (1987) 1372–1381.
- [13] A. Klamt, G. Schuurmann, COSMO—a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *Journal of the Chemical Society-Perkin Transactions 2* (1993) 799–805.
- [14] A. Klamt, V. Jonas, T. Burger, J. Lohrenz, Refinement and parametrization of COSMO-RS, *Journal of Physical Chemistry A* 102 (1998) 5074–5085.
- [15] S. Lin, S. Sandler, A priori phase equilibrium prediction from a segment contribution solvation model, *Industrial & Engineering Chemistry Research* 41 (2002) 899–913.
- [16] E. Mullins, R. Oldland, Y. Liu, S. Wang, S.I. Sandler, C.-C. Chen, M. Zwolak, K.C. Seavey, Sigma-profile database for using COSMO-based thermodynamic methods, *Industrial & Engineering Chemistry Research* 45 (2006) 4389–4415.

- [17] C.-C. Chen, Y. Song, Solubility modeling with a nonrandom two-liquid segment activity coefficient model, *Industrial & Engineering Chemistry Research* 43 (2004) 8354–8362.
- [18] H. Renon, J. Prausnitz, Local composition in thermodynamic excess functions for liquid mixtures, *AIChE Journal* 14 (1968) 135–144.
- [19] C.-C. Chen, A segment-based local composition model for the Gibbs energy of polymer solutions, *Fluid Phase Equilibria* 83 (1993) 301.
- [20] L.C. Garzon, F. Martinez, Temperature dependence of solubility for ibuprofen in some organic and aqueous solvents, *Journal of Solution Chemistry* 33 (2004) 1379–1395.
- [21] S. Gracin, A. Rasmuson, Solubility of phenylacetic acid, p-hydroxyphenylacetic acid, p-aminophenylacetic acid, p-hydroxybenzoic acid, and ibuprofen in pure solvents, *Journal of Chemical and Engineering Data* 47 (2002) 1379–1383.
- [22] R. Granberg, A. Rasmuson, Solubility of paracetamol in pure solvents, *Journal of Chemical and Engineering Data* 44 (1999) 1391–1395.
- [23] K. De Fina, T. Sharp, L. Roy, W. Acree, Solubility of 8-hydroxybenzoic acid in select organic solvents at 298.15 K, *Journal of Chemical and Engineering Data* 44 (1999) 1262–1264.
- [24] H. Fung, T. Higuchi, Molecular interactions and solubility of polar nonelectrolytes in nonpolar solvents, *Journal of Pharmaceutical Sciences* 60 (1971) 1782–1788.
- [25] F.L. Nordstrom, A.C. Rasmuson, Solubility and melting properties of salicylic acid, *Journal of Chemical and Engineering Data* 51 (2006) 1668–1671.
- [26] A. Paruta, B. Sciarone, N. Lordi, Solubility of salicylic acid as a function of dielectric constant, *Journal of Pharmaceutical Sciences* 53 (1964) 1349–1353.
- [27] A. Shalmashi, A. Eliassi, Solubility of salicylic acid in water, ethanol, carbon tetrachloride, ethyl acetate, and xylene, *Journal of Chemical and Engineering Data* 53 (2008) 199–200.
- [28] W. Acree, G. Bertrand, Thermochemical investigations of nearly ideal binary solvents. 7. Monomer and dimer models for solubility of benzoic-acid in simple binary and ternary solvents, *Journal of Pharmaceutical Sciences* 70 (1981) 1033–1036.
- [29] A. Beerbower, P. Wu, A. Martin, Expanded solubility parameter approach. 1. Naphthalene and benzoic-acid in individual solvents, *Journal of Pharmaceutical Sciences* 73 (1984) 179–188.
- [30] D. Li, D. Liu, F. Wang, Solubility of 4-methylbenzoic acid between 288 K and 370 K, *Journal of Chemical and Engineering Data* 46 (2001) 234–236.
- [31] H. Lin, R. Nash, An experimental method for determining the Hildebrand solubility parameter of organic nonelectrolytes, *Journal of Pharmaceutical Sciences* 82 (1993) 1018–1025.
- [32] S. Yalkowski, S. Valvani, T. Roseman, Solubility and partitioning vi: octanol solubility and octanol–water partition coefficients, *Journal of Pharmaceutical Science* 72 (1983) 866–870.
- [33] C. Hancock, J. Pawlowski, J. Idoux, *Journal of Organic Chemistry* 31 (1966) 3801.
- [34] C. Daniels, A. Charlton, R. Wold, R. Moreno, W. Acree, M. Abraham, Mathematical correlation of 4-aminobenzoic acid solubilities in organic solvents with the Abraham solvation parameter model, *Physics and Chemistry of Liquids* 42 (2004) 633–641.
- [35] E. Cepeda, B. Gomez, Solubility of anthracene and anthraquinone in some pure and mixed solvents, *Journal of Chemical Engineering Data* 34 (1989) 273–275.
- [36] E. Cepeda, M. Diaz, Solubility of anthracene and anthraquinone in acetonitrile, methyl ethyl ketone, isopropyl alcohol and their mixtures, *Fluid Phase Equilibria* 121 (1996) 267–272.
- [37] G. Al-Sharrah, S. Ali, M. Fahim, Solubility of anthracene in two binary solvents containing toluene, *Fluid Phase Equilibria* 193 (2002) 191–201.
- [38] J. Powell, W. Acree, Solubility of anthracene in binary alcohol + dibutyl ether solvent mixtures, *Journal of Chemical Engineering Data* 40 (1995) 914–916.
- [39] A. Zvaigzne, W. Acree, Thermochemical investigations of molecular complexation. Estimation of anthracene–ethyl acetate and anthracene–diethyl adipate association parameters from measured solubility data, *Physics & Chemistry of Liquids* 24 (1991) 31–42.
- [40] D. Andrews, G. Lynn, J. Johnston, The heat capacities and heat of crystallization of some isomeric aromatic compounds, *Journal of American Chemical Society* 48 (1926) 1274–1287.
- [41] D. David, Determination of specific heat and heat of fusion by differential thermal analysis. study of theory and operating parameters, *Analysis Chemistry* 36 (1964) 2162–2166.
- [42] D. Ginnings, G. Furukawa, Heat capacity standards for the range 14 to 1200, *Journal of American Chemical Society* 75 (1953) 522–527.
- [43] D. Grant, M. Mehdizadeh, A. Chow, J. Fairbrother, Nonlinear van't Hoff solubility–temperature plots and their pharmaceutical interpretation, *International Journal of Pharmaceutics* 18 (1984) 25–38.
- [44] D. Li, J. Liu, D. Liu, F. Wang, Solubilities of terephthalaldehydic, p-toluic, benzoic, terephthalic and isophthalic acids in n,n-dimethylformamide from 294.75 to 370.45 K, *Fluid Phase Equilibria* 200 (2002) 69–74.
- [45] R. Manzo, A. Ahumada, Effects of solvent medium on solubility. 5. Enthalpic and entropic contributions to the free-energy changes of disubstituted benzene-derivatives in ethanol water and ethanol cyclohexane mixtures, *Journal of Pharmaceutical Sciences* 79 (1990) 1109–1115.
- [46] S. Neau, S. Bhandarkar, E. Hellmuth, Differential molar heat capacities to test ideal solubility estimations, *Pharmaceutical Research* 14 (1997) 601–605.
- [47] T. Nielsen, J. Abildskov, P. Harper, I. Papaconomou, R. Gani, The CAPEC database, *Journal of Chemical and Engineering Data* 46 (2001) 1041–1044.
- [48] J. Qingzhu, M. Peisheng, Y. Shouzhi, W. Qiang, W. Chan, L. Guiju, Solubilities of benzoic acid, p-methylbenzoic acid, m-methylbenzoic acid, o-methylbenzoic acid, p-hydrobenzoic acid, and o-nitrobenzoic acid in 1-octanol, *Journal of Chemical and Engineering Data* 53 (2008) 1278–1282.
- [49] S. Romero, A. Reillo, B. Escalera, P. Bustamante, The behaviour of paracetamol in mixtures of amphiprotic and amphiprotic–aprotic solvents. Relationship of solubility curves to specific and nonspecific interactions, *Chemical & Pharmaceutical Bulletin* 44 (1996) 1061–1064.
- [50] M. Sacchetti, Thermodynamic analysis of DSC data for acetaminophen polymorphs, *Journal of Thermal Analysis and Calorimetry* 63 (2001) 345–350.
- [51] R. Weast, M. Astle, in: *CRC Handbook of Data on Organic Compounds*, CRC Press, Boca Raton, FL, 1985.
- [52] M. Wu, S. Yalkowsky, Estimation of the molar heat capacity change on melting of organic compounds, *Industrial & Engineering Chemistry Research* 48 (2009) 1063–1066.
- [53] F. Xu, L. Sun, Z. Tan, J. Liang, R. Li, Thermodynamic study of ibuprofen by adiabatic calorimetry and thermal analysis, *Thermochimica Acta* 412 (2004) 33–37.
- [54] R. Van Duyn, Taylor, Self-association and hydration of benzoic acid in benzene, *Journal of Physics Chemistry* 48 (2009) 1063–1066.
- [55] C.-C. Chen, Y. Song, Extension of nonrandom two-liquid segment activity coefficient model for electrolytes, *Industrial & Engineering Chemistry Research* 44 (2005) 8909–8921.
- [56] C.-C. Chen, P.A. Crafts, Correlation and prediction of drug molecule solubility in mixed solvent systems with the nonrandom two-liquid segment activity coefficient (NRTL-SAC) model, *Industrial & Engineering Chemistry Research* 45 (2006) 4816–4824.
- [57] I. Tsvintzelis, G. Dritsas, C. Panayiotou, An alternative approach to nonrandomness in solution thermodynamics, *Industrial & Engineering Chemistry Research* (2006) 7264–7274.
- [58] J. Gross, G. Sadowski, Perturbed-chain saft: an equation of state based on a perturbation theory for chain molecules, *Industrial & Engineering Chemistry Research* 40 (2001) 1244–1260.