Fluid Phase Equilibria 322–323 (2012) 48–55

Contents lists available at SciVerse ScienceDirect

Fluid Phase Equilibria

journal homepage: www.elsevier.com/locate/fluid

Calculation of drug-like molecules solubility using predictive activity coefficient models

Fátima L. Mota[a], António J. Queimada[a], Alfonsina E. Andreatta[b], Simão P. Pinho[c], Eugénia A. Macedo[a,*]

[a] LSRE – Laboratory of Separation and Reaction Engineering, – Associate Laboratory LSRE/LCM, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200 - 465 Porto, Portugal
[b] PLAPIQUI, Universidad Nacional del Sur–CONICET, Camino La Carringanda km 7, 8000 Bahia Blanca, Argentina
[c] LSRE – Laboratory of Separation and Reaction Engineering, – Associate Laboratory LSRE/LCM, Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5301-857 Bragança, Portugal

ARTICLE INFO

Article history:
Received 30 November 2011
Received in revised form 31 January 2012
Accepted 6 February 2012
Available online 10 March 2012

Keywords:
A-UNIFAC
Complex chemicals
Modelling
NRTL-SAC
Solubility
UNIFAC

ABSTRACT

The A-UNIFAC, UNIFAC, and NRTL-SAC models are used to predict solubility in pure solvents of a set of drug-like molecules. To apply A-UNIFAC, a new set of residual interaction parameters between the —ACOH group and six other groups had to be estimated. The solute model parameters of NRTL-SAC were also estimated for this set of molecules. NRTL-SAC showed better performance at 298.15 K, with an average absolute deviation of 37.6%. Solubility dependence with temperature was also studied: all models presented average deviations around 40%. In general, there is an improvement given by the A-UNIFAC over the UNIFAC in aqueous systems, proving the importance of taking association into account.

The reference solvent approach was also applied improving the results. Solubility in pure solvents can now be predicted with an average deviation around 35.2%. This approach reduces differences previously found between the three models, being a powerful methodology.

1. Introduction

Solubility has been recognized as one of the most important properties for designing separation and purification processes of complex molecules, such as biomolecules and active pharmaceutical ingredients [1,2]. In many cases, data are unavailable due to reduced amounts of sample, time limitations, or inherent complexities with experimental measurements. In such cases, the increasing interest and importance of solubility modelling is clear, as can be seen in the very recent works of Ellegaard et al. [3] and Diedrichs and Gmehling [4].

The non-random two-liquid (NRTL) segment activity coefficient (SAC) is a recent and successful model for solid–liquid equilibrium calculations, and the idea behind is to limit the number of intermolecular interaction parameters [5,6]. This model was successfully employed for complex chemicals where acceptable deviations were obtained between experimental and predicted values: average root mean square error in $\ln x$ of 0.37, what corresponds to ±45% of accuracy in solubility predictions [5–8].

The group-contribution methods, like UNIFAC [9,10] and A-UNIFAC [11,12] are adequate techniques to provide reasonably accurate estimates of fluid mixtures non-idealities. Gracin et al. [13] used UNIFAC to predict drug solubilities in pure solvents, but the results were not accurate enough to allow the design of crystallization processes, or the selection of solvents. It is well known that UNIFAC method is not suitable when strong association effects are present. In order to take them into account, an extended version, the A-UNIFAC, was presented by Mengarelli et al. [14], being successfully applied to mixtures containing sugars, alcohols, water, carboxylic acids, esters, aromatic hydrocarbons and alkanes [11,12,15].

Avoiding the knowledge of the solute melting properties, the reference-solvent approach (RSA) [1,16,17] is an interesting alternative for predicting solubilities of solids. It minimizes the impact of the melting data uncertainties, being so interesting when there is a decomposition reaction, or solid–solid or glass transitions, and allows to fit a small number of unknown parameters from a limited set of well-chosen experimental data. However, it involves the selection of a reference solvent, relative to which all solubility calculations are made. This methodology was used to study the solubilities of drugs in pure solvents [1], and of complex medium-sized chemicals in mixed solvents [16], and the results are promising even in the cases of relatively high solubility compounds where it is not expected.

The aim of this work is to evaluate different activity coefficient models, UNIFAC, A-UNIFAC and NRTL-SAC, for solubility

* Corresponding author. Tel.: +351 22 508 1653; fax: +351 22 508 1674.
E-mail address: eamacedo@fe.up.pt (E.A. Macedo).

0378-3812/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.fluid.2012.02.003

Table 1
Pure solute properties.

Solute	T_m (K)	$\Delta_{fus}H$ (kJ/mol)	ΔC_p (J/mol.K)	M (g/mol)
Salicylic acid	432.0[a]	24.6[a]	37.0[b]	138.1
Benzoic acid	395.5[c]	17.5[c]	40.9[b]	122.1
Acetylsalicylic acid	416.2[d]	29.8[d]	26.6[b]	180.2
Ibuprofen	347.2[e]	25.5[e]	50.3[e]	206.3
Hydroquinone	445.5[b]	27.1[b]	26.7[b]	110.1
Estriol	555.0[f]	42.7[f]	–	288.4
Estradiol	445.0[f]	40.6[f]	–	272.4

[a] Ref. [1].
[b] Ref. [19].
[c] Ref. [20].
[d] Ref. [21].
[e] Ref. [22].
[f] Ref. [23].

calculations of drug-like molecules, predicting how solubility vary with solvent type and temperature. With the application of RSA, the impact of calculating solubility without explicitly using pure component properties was checked. For that purpose, an evaluation database constituted by compounds with different functional groups and molecular sizes, showing a multiplicity of interactions in different solvents was compiled. They are benzoic acid, salicylic acid, acetylsalicylic acid, ibuprofen, hydroquinone, estriol and estradiol (Fig. 1). An extension of the group interaction parameters available for A-UNIFAC was needed, and the solute parameters of the NRTL-SAC model were also estimated. The number of solutes and solvents involved extends considerably previous studies, allowing to understand their usefulness to predict solubility of pharmaceuticals.

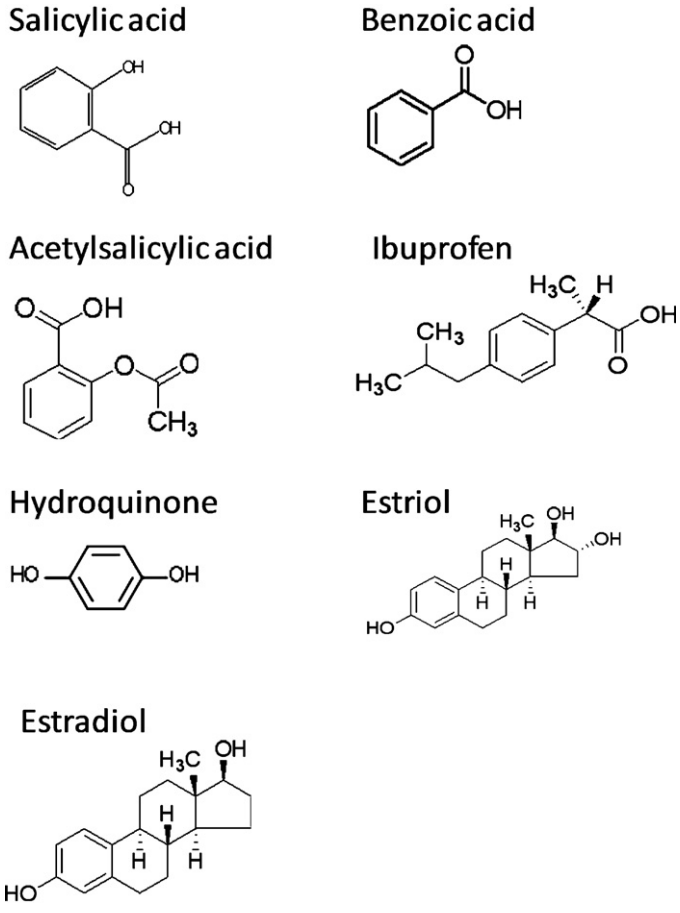

Fig. 1. Chemical structures of the compounds under study.

2. Modelling

The solubility of a solid solute in a liquid solvent can be calculated by solving the thermodynamic equations of equilibrium [18]. In Appendix A, the most important equations of the three models used in this work are briefly presented. The required pure solute properties are given in Table 1.

2.1. New A-UNIFAC group interaction parameters

For the group-contribution methods, the volume and area parameters were based on the Gmehling et al. [24] revised UNIFAC tables. The residual group interaction ($a_{m,n}$) and association parameters used in A-UNIFAC method were published by Ferreira et al. [12], while those used in UNIFAC method were based on Poling et al. [25] tables. The chemicals have the following associative groups: salicylic acid has carboxylic, hydroxyl and aromatic ring; benzoic acid and ibuprofen have carboxylic and aromatic ring; acetylsalicylic acid has carboxylic, ester and aromatic ring; hydroquinone, estriol and estradiol have hydroxyl and aromatic ring. In the majority of the studied systems, both solute and solvent can associate, like in alcohols and aqueous systems. Benzene derivatives and esters can only associate if the solvent has an electropositive site.

However, to apply the A-UNIFAC method, it was necessary to extend the residual group interaction parameters available [12]: the unknown parameters involving the —ACOH group were estimated to represent molecules such as salicylic acid, hydroquinone, estriol and estradiol. The association in this group is assumed to be the same as the association in the —OH group, given by Ferreira et al. [12], but the residual group interaction parameters were estimated using experimental data on low-pressure vapor–liquid equilibria (VLE) of binary mixtures and liquid–liquid equilibria (LLE) of ternary mixtures. Table 2 shows the database used in the fitting procedure. The residual parameters for ACOH/ACH and ACOH/CH_2 were first calculated using VLE data of phenol–benzene and phenol–alkanes, respectively, while the residual parameters for ACOH/H_2O, ACOH/OH, ACOH/COOR, ACOH/COOH interactions were determined simultaneously using all other data expressed in Table 2. Table 3 reports the new estimated group interaction parameters for —ACOH between paraffinic, aromatic, alcohol, water, ester and acid groups.

Table 4 lists the average deviations obtained in the correlation of binary VLE data using the A-UNIFAC method. The results are satisfactory, even if for water–phenol system higher deviations were found. In fact, in order to obtain a good description of both VLE and LLE equilibria a compromise had to be considered when estimating the ACOH/H_2O parameter. As can be seen in Fig. 2, an accurate description of the LLE for the water–phenol–benzene system was achieved using the same set of parameters for both types of equilibria.