Review

The objective minimization function for the mathematical representation of solubility data for solutes dissolved in binary solvent mixtures

William E. Acree Jr. *, Melissa Y. Horton

Department of Chemistry, University of North Texas, 1155 Union Circle Drive #305070, Denton, TX 76203, USA

A B S T R A C T

Two minimization objective functions are used in calculating the curve-fit parameters of the Simplified Combined Nearly Ideal Binary Solvent (NIBS)/Redlich-Kister model for describing the experimental solubility of both linonin in binary aqueous-acetone solvent mixtures and betulinic acid in binary aqueous-tetrahydrofuran solvent mixtures. Curve-fit parameters based on the minimization of the summed squared differences between the experimental mole fraction solubilities and back-calculated values were found to provide a very good mathematical description of the observed solubility behavior in the organic solvent-rich portions of the binary solvent system. The back-calculated values, however, differed significantly from the observed values at low organic solvent compositions. Curve-fit parameters obtained using the summed squared difference between the natural logarithms of the experimental and back-calculated mole fraction solubilities provided the better mathematical description over the entire range of solvent composition.

© 2016 Elsevier Ltd.

Contents

1. Introduction

In the last five years there has been a significant increase in the number of research groups reporting solubility data for crystalline nonelectrolyte solutes dissolved both in neat organic solvents and in binary solvent mixtures. Modern chemical instrumentation and computational software have greatly reduced the time needed to perform accurate solubility determinations. The specific systems studied were often selected to address an industrial or manufac-turing concern resulting from the selection of organic solvents or solvent mixtures to use in solute purifications for recrystallization or to use in solubilizing the solute in possible drug formulations. Medical compounds and pharmaceutical ingredients [1–6], pesticides, herbicides and fungicides [7–10], and starting materials or intermediates for important chemical products [11–14] were often the subject of the published solubility studies. In the case of neat organic solvents the solubility data were measured at several temperatures. For binary solvent mixtures the solubility was generally measured over the entire composition at either a single temperature or at multiple temperatures, depending upon whether the authors wanted to calculate thermodynamic properties of dissolution from their measured solubility data.

---

* Corresponding author.
  E-mail address: acree@unt.edu (W.E. Acree Jr.).

Most of the published solubility studies include some form of mathematical representation that allows journal readers to calculate solubilities at other temperatures and/or at other binary solvent compositions. Some of the publications even include mathematical correlations for estimating the solubility of the solute in additional organic solvents that were not part of the experimental study. Mathematical representations for describing experimental solubility as a functional of temperature include the Modified Apelblat equation [15], the van't Hoff equation [16,17] and the Buchowski-Ksiazczak λh equation [18]. Experimental solubility data for binary solvent mixtures can be described with the Wilson model [19], the NRTL model [20], the Combined Nearly Ideal Binary Solvent/Redlich-Kister (CNIBS/R-K) model [21,22] or any of the polynomial simplifications of the CNIBS/R-K model. The Jouyban-Acree model [23,24], the combined Jouyban-Acree-van't Hoff model [25], the combined Jouyban-Acree-Apelblat model [26], or any of the modified polynomial versions [26–30] of the fore-mentioned models provide a convenient means to mathematical describe solubility data as a function of both temperature and binary solvent composition. The fore-mentioned listing is not intended to be inclusive, but rather indicate a few of the equations that are being used to mathematical describe the measured solubility data published in the last five years.

Each of the models contain adjustable equation coefficients whose numerical values are obtained by curve-fitting the experimental data in accordance with the mathematical form of the chosen model. The curve-fitting procedure generally involves a regression analysis that calculates the curve-fit parameters that minimize an objective function. Various research groups have used different objective minimization functions in calculating the adjustable curve-fit parameters. For example, several researchers [28,31,9,32,33] have used minimization functions based on activity coefficients, γ_i:

$$\text{Minimization function} = \sum_{i=1}^{N} \left[\ln \gamma_i^{calc} - \ln \gamma_i^{exp} \right]^2 \tag{1}$$

when using the Wilson and NRTL models as mathematical representations. Moodley and coworkers [34], Xu et al. [35], Shi et al. [36], and Han and coworkers [37] minimized the sum of the squared differences:

$$\text{Minimization function} = (1/N) \sum_{i=1}^{N} \left[x_i^{calc} - x_i^{exp} \right]^2 \tag{2}$$

between the experimental mole fraction solubilities and calculated values from the NRTL, Wilson and/or Tsuboka-Katayama modified Wilson activity coefficient models. Still others have elected to calculate the curve-fit parameters using the Average Absolute Relative Deviation (AARD) [38,39]:

$$\text{Minimization function} = \text{AARD} = (1/N) \sum_{i=1}^{N} \left| \frac{x_i^{calc} - x_i^{exp}}{x_i^{exp}} \right| \tag{3}$$

or the summed squared difference in the natural logarithms of the experimental and calculated mole fraction solubilities [3]:

$$\text{Minimization function} = \sum_{i=1}^{N} \left[\ln x_i^{calc} - \ln x_i^{exp} \right]^2 \tag{4}$$

or the summed squared relative deviation in the activity coefficient [40]:

$$\text{Minimization function} = \sum_{i=1}^{N} \left[\frac{\gamma_i^{calc} - \gamma_c^{exp}}{\gamma_i^{exp}} \right]^2 \tag{5}$$

as their minimization functions. Needless-to-say, given the number of minimization functions used there is no consensus among the

different researchers measuring and reporting solubility data regarding which function is best.

In the present commentary we are going to compare two of the minimization functions in terms of their ability to generate curve-fit equation coefficients that best describe an actual experimental solubility data set. We have selected for our two illustrative examples the solubility of limonin in binary aqueous-acetone solvent mixtures which was published in *This Journal* [41], and the solubility of betulinic acid binary aqueous-tetrahydrofuran solvent mixtures which was published in another *Elsevier Journal* [42]. These specific systems were picked because the solutes are considerably more soluble in the organic cosolvent than in water. Limonin is approximately 35,000 times more soluble in acetone than in water on a mole fraction basis. Similarly, betulinic acid is nearly 21,000 times more soluble in tetrahydrofuran than in water. Based on our preliminary calculations we found that the difference in curve-fit equation coefficients using different minimization functions becomes more significant in systems where there is a sizeable disparity in solubility in going from one mixture co-solvent to the other.

2. Mathematical representation of limonin solubility data in binary aqueous-acetone mixtures

We have given in Table 1 the pertinent information pertaining to the limonin solubility study. Tabulated in the first two columns are the mole fraction compositions of acetone in the binary solvent mixture calculated as if the solute were not present, $x_{acetone}^o$, and the measured mole fraction solubility of limonin determined at 283.2 K. Solubilities were measured at 5 K intervals between 283.2 and 318.2 K; however, one temperature is sufficient to illustrate the points that we care to make. Included with the mole fraction solubility is the standard uncertainty that the authors reported for each of the respective mole fraction solubilities. As part of their experimental study the authors did mathematically describe the observed mole fraction solubility data in terms of the Combined NIBS/Redlich-Kister, Simplified Combined NIBS/Redlich-Kister, Jouyban-Acree, Jouyban-Acree-van't Hoff, Jouyban-Acree-Apelblat, Ma, and Sun models. We did not find any information in the manuscript indicating which minimization function the authors might have used in calculating their curve-fit parameters.

For illustration purposes we will use as our mathematical representation the model that the authors referred to in their paper as the Simplified Combined NIBS/Redlich-Kister model:

$$\ln x_1 = B_0 + B_1 x_{organic}^o + B_2 \left(x_{organic}^o \right)^2 + B_3 \left(x_{organic}^o \right)^3 + B_4 \left(x_{organic}^o \right)^4 \tag{6}$$

which is a polynomial expansion of the Combined NIBS/Redlich-Kister model

$$\ln x_1 = x_{organic}^o \ln (x_1)_{organic} + x_{water}^o \ln (x_1)_{water} + x_{water}^o x_{organic}^o \sum_{i=1}^{n} B_i \left(x_{organic}^o - x_{water}^o \right)^i \tag{7}$$

obtained by substituting $x_{water}^o = 1 - x_{organic}^o$ into Eq. (7) and then performing several algebraic manipulations to combine like mole fraction terms. In Eqs. (6) and (7) $(x_1)_{organic}$ and $(x_1)_{water}$ denote the solubility of the solute in neat organic cosolvent (acetone) and water, respectively, x_{water}^o and $x_{organic}^o$ refer to the mole fraction compositions of water and organic cosolvent in the binary solvent mixture calculated as if the solute were not present, and the various B_is represent the adjustable curve-fit parameters obtained by analyzing