

Prediction of solubilities of complex medium-sized chemicals. Solutes in pure solvents, mixed solvents and cosolvent mixtures

J. Abildskov^{a*} and J.P. O'Connell^b

^aComputer-Aided Process Engineering Center, Department of Chemical Engineering,
Technical University of Denmark

Building 229, Søtofts Plads, 2800 Kgs. Lyngby, Denmark

^bDepartment of Chemical Engineering, University of Virginia

102 Engineers' Way, P.O. Box 400741, Charlottesville, VA 22904-4741, USA

Abstract

The processing and formulation of pharmaceutical, medical and agrochemical products often requires the solubility of an active substance in liquids. Being able to provide fast and reasonably reliable property information, either from measurement or prediction, can greatly enhance the competitiveness of formulations. However, due to the complexity of many active ingredients and solvents, reliable measurements are often unavailable while existing property prediction models cannot provide sufficiently accurate predictions. Therefore most design work in this field is done by trial-and-error and empirical descriptions of limited experimental results. The result is that industrial process groups spend as much as 30 % of their time on solubility-related problems. To alleviate the inefficiency and errors of this situation, we have taken a simplifying approach to determining changes in the solubility of sparingly soluble complex chemicals as the solvent type and composition are varied.

Keywords: solubility, prediction, industrial products

1. Introduction

Medium-sized solutes are often characterized by:

- molecular weight greater than 100 g/mol
- elevated melting temperature
- polyfunctionality (two or more functional groups)
- multiple molecular conformations
- large dipoles, high polarizability, hydrogen bonding, charge transfer complexes
- limited amounts of reliable measured data

The thermodynamic approach to such systems uses pure solute properties such as the melting temperature, T_m , and heat of melting, ΔH_m . These numbers are often unknown. If not the data are often uncertain. Therefore we seek a method that avoids using them.

* Author/s to whom correspondence should be addressed: ja@kt.dtu.dk

In addition, the nonideality of the solute/solvent mixture is needed. Given the great range of substances, in the absence of molecular computations, solution-of-groups frameworks are probably the only realistic means for treating such systems. Unfortunately, these models often require many model parameters describing the influence of unlike interactions, not all of which are known. Consider the solute ethinyl estradiol (CAS:57-63-6; Mw = 296.4 g/mol), shown in Figure 1.

- Ethinyl Estradiol
- 22 non-hydrogen atoms
- 6 functional groups
 - CH₂
 - ACH
 - ACCH₂
 - OH
 - ACOH
 - C≡C
- Solubility in
 - Hexane involves 30 interaction parameters
 - Water involves 42 interaction parameters

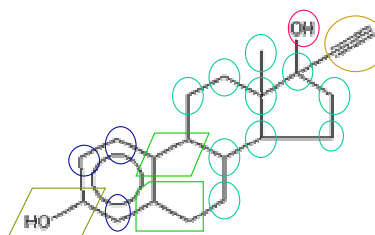


Figure1. Ethinyl Estradiol. Functional Groups.

Characterizing ethinyl estradiol in terms of functional groups requires 6 different functional groups, meaning that even calculating the solubility in n-hexane requires 30 interaction parameters. If the solvent is water, values for 42 interaction parameters should be known. If the solvent is a binary mixed solvent of water and methanol, 56 parameters are required. If any of the solute's functional groups are not covered by UNIFAC or ASOG, there will be many parameters to be evaluated or estimated before calculations can be made. This usually means that such models cannot effectively be applied. Even if all of the groups have been studied for some solutions, it often happens that a subset of the relevant parameters remains unavailable. Based on group contribution methods and a limited amount of data points, we (Abildskov and O'Connell, 2003 and 2004a) have found that a simplified approach can surmount these difficulties by determining changes in the solubility of sparingly soluble complex chemicals as the solvent (mixed or pure) composition is varied and focussing on the most important groups. The method is based on a 'reference solvent' which effectively eliminates the necessity of pure-solute properties and decreases the number of adjustable parameters to be determined by data reduction with a mathematical regularization strategy to find values for only the most significant unknown parameters. The method has been established for mixed and pure solvent systems, and the resulting models have been extended to cosolvent design problems (Abildskov and O'Connell, 2004b). While the method involves no new fundamental solution theory, it shows how existing thermodynamic insights based on molecular interactions can lead to more efficient use of thermodynamic models in complex cases.

2. Method

The basic relationships we use for describing low liquid solubilities of solutes that are solid when pure at the system temperature are derived from the equality of component fugacities in all phases. Then assumptions can be made about solid purity and whether the liquid state is equivalent to infinite dilution. For a component i pure in the solid (S) phase and soluble in the liquid (L) phase

$$\ln x_i \gamma_i(T, \mathbf{x}) = \ln \frac{f_i^{OS}}{f_i^{OL}} \equiv \ln x_i^{id} \approx \frac{\Delta H_{mi}}{RT_{mi}} \left(1 - \frac{T_{mi}}{T}\right) \quad (1)$$

where the 'ideal' solubility, $\ln x_i^{id}$, has been approximated with only melting point properties, T_{mi} and ΔH_{mi} , though more complete expressions can be used (Choi et al., 1983). Here the activity coefficient, γ_i is for the Lewis/Randall standard state (pure component as a liquid at the system T), a function of T and solution composition, $\mathbf{x}$. We estimate γ_i with UNIFAC. Two pure group parameters for each solute, R_k and Q_k , are for size and shape, and are obtained on the basis of atomic structure (Bondi, 1968). Two binary interaction parameters for each pair, identified as $a_{mn} \neq a_{nm}$, are determined by fitting to reliable measured data on systems involving the group(s) of interest. If the solubility is low enough ($x_i \leq 0.01$), the activity coefficient can be considered as at infinite dilution, then

$$\ln \gamma_i(T, \mathbf{x}) \approx \ln \gamma_i^\infty(T, \mathbf{x}') \quad (2)$$

where γ_i^∞ is a function only of T and the (solute-free) solvent composition, $\mathbf{x}'$. At low concentrations, then the solubility can be approximated as

$$\ln x_i = \frac{\Delta H_{mi}}{RT_{mi}} \left(1 - \frac{T_{mi}}{T}\right) - \ln \gamma_i^\infty(T, \mathbf{x}') \quad (3)$$

If one applies equation 3 to a pure or mixed solvent (m) and subtracts the same expression for a reference solvent j , there is a direct relation between the solubility in m with that in j and predicted infinite dilution activity coefficients in the solvents

$$\ln x_{im} = \ln x_{ij} + [\ln \gamma_i^\infty(T, (\mathbf{x}')_j) - \ln \gamma_i^\infty(T, (\mathbf{x}')_m)] \quad (4)$$

Finally, if not all of the binary group parameters are known, it is possible with regularization techniques to determine a large number of parameters from relatively few data by minimizing the objective function

$$S = \sum_{m < j} (\delta \Delta \ln x_{i,mj})^2 + \alpha \sum_{m \neq n} (\delta a_{mn})^2 \quad (5)$$

where

$$\begin{aligned} \delta \Delta \ln x_{i,mj} &\equiv (\ln x_{im} - \ln x_{ij})^{\text{exp}} - (\ln x_{im} - \ln x_{ij})^{\text{calc}} \\ &= (\ln x_{im} - \ln x_{ij})^{\text{exp}} - (\ln \gamma_i^\infty(T, (\mathbf{x}')_j) - \ln \gamma_i^\infty(T, (\mathbf{x}')_m)) \end{aligned} \quad (6)$$

and

$$\alpha \sum_{m \neq n} (\delta a_{mn})^2 = \alpha \sum_{m \neq n} (a_{mn} - a_{mn}^0)^2 \quad (7)$$

Here α is a small constant and a_{mn}^0 an initial guess of the parameter (commonly zero) to keep the data and parameter residuals in proper proportion. If the influence of the parameter is not strong, $a_{mn} \sim a_{mn}^0$ after the fitting; the parameter does not significantly affect the solubility. The selection of an optimal reference solvent, j , is possible if experimental

solubilities of the solute have been measured in N solvents. The optimal reference solvent (j) obeys

$$\ln x_{ij} + \ln \gamma_{ij}^{\infty} - \frac{1}{N} \sum_{k=1}^N (\ln x_{ik} + \ln \gamma_{ik}^{\infty}) = 0 \quad (8)$$

Using equation 4, the procedure is to select the data successively, ultimately choosing the solvent, j , that most closely obeys equation 8 (Abildskov and O'Connell, 2004). One can of course not guarantee optimal predictions for solvents outside of the database used, but we have found good success when minimizing the numerical value of the lhs of equation 8. It should be noted that the equation 4 does not contain solute-solute interaction parameters. This can result in a quite significant reduction of the number of interaction parameters. Consider again ethinyl estradiol. If we categorize the parameters used for calculating its solubility in water or n-hexane we find that the majority of these characterize solute-solute interaction effects.

- Interactions in n-hexane:
 - 10 Solute-Solvent (33%)
 - 20 Solute-Solute (67%)
 - 0 Solvent-Solvent
- Interactions in water:
 - 12 Solute-Solvent (29%)
 - 30 Solute-Solute (71%)
 - 0 Solvent-Solvent

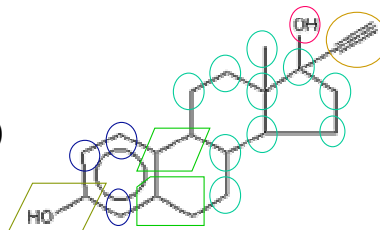


Figure 2. Ethinyl Estradiol. Parameter reduction from 30 or 42 (water) to 10 or 12 (water)

Thus, the scale of the problem can be significantly reduced in such cases, and the likelihood of encountering ‘missing parameters’ can be significantly reduced, if one uses a reference solvent model. Treatments of this type of solutes have been attempted before by other authors. Ruckenstein and Shulgin (2001) have applied strongly simplified equations based on fluctuation solution theory to the solubility of poorly soluble substances of environmental significance in aqueous mixed solvents. Their treatment (like ours) considers a solute at infinite dilution. There is also precedent for correlating differences of solubility in solvents. For example, Acree and Abraham (2002) correlate the equivalent of $\Delta \ln x_{i,mj}$ relative to that of water or in the gas phase with molecular descriptors. However, none of these workers have addressed the issue of selecting an optimal reference solvent or a strategy of data measurements.

3. Examples

The method is applicable to a wide range of cases. For example consider the solubility of solutes in different pure solvents where estimation of ‘missing parameters’ can be made. Diuron has a functional group not covered by UNIFAC, namely the AC-NH-portion. We estimated (Abildskov and O'Connell, 2003) its interaction parameters with a number of other groups, and use the derived values to satisfactorily predict the properties of a chemically related substance monuron, as shown in Figure 3.

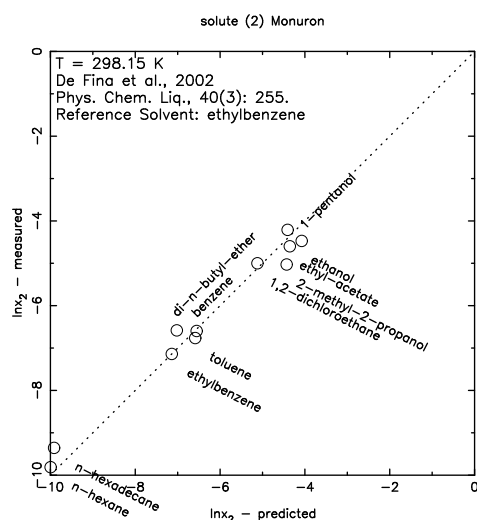


Figure 3. Monuron with ethylbenzene as reference solvent and parameters from diuron

The method is also applicable to solutes in mixed solvents. All UNIFAC groups necessary for calculating the solubility of naphthalene in mixtures of ethanol and water are available. Naphthalene has only a single main group. In Figure 4, the equations (1), (3), and (4) with available parameters capture well the data trends. Here the reference is pure. However, this is not necessary. A mixed solvent reference is used in the case of biphenyl in n-heptane/ CCl_4 mixtures. Figure 5 shows that both equations 1 and 2 are inaccurate due to the high solubility and inaccurate UNFAC parameters. However, the reference solvent approach is quite successful in describing the data.

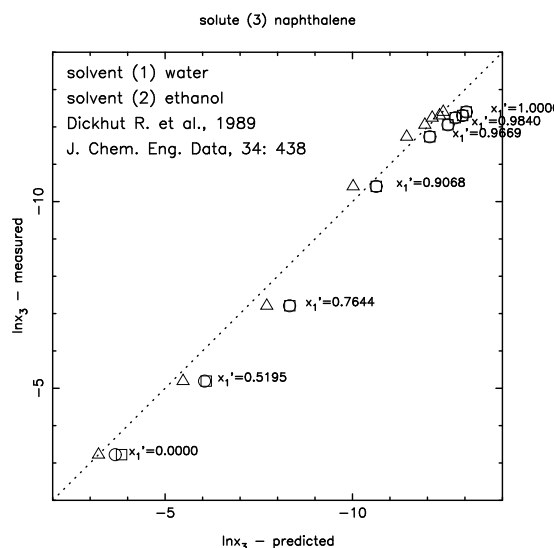


Figure 4. Naphthalene in water/ethanol mixtures ($T=298.15\text{ K}$) with no parameter-fitting. $\circ$ UNIFAC equation 1: $\ln x_3 = \ln x_3^{\text{id}} - \ln \gamma_3$; $\square$ UNIFAC equation 2: $\ln x_3 = \ln x_3^{\text{id}} - \ln \gamma_3^\infty$; Δ UNIFAC Reference Solvent (pure ethanol) equation 4 and equation 8

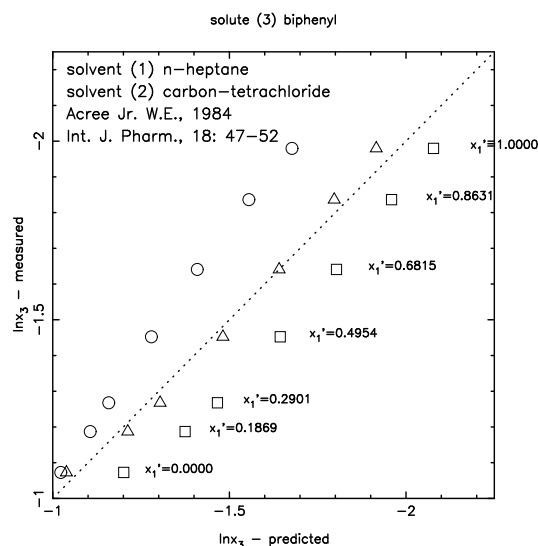


Figure 5 .Biphenyl in *n*-heptane/*CCl*₄ mixtures (*T*=298.15 K) with no parameter fitting. *O* UNIFAC equation 1: $\ln x_3 = \ln x_3^{id} - \ln \gamma_3$; $\square$ UNIFAC equation 2: $\ln x_3 = \ln x_3^{id} - \ln \gamma_3^\infty$; Δ UNIFAC Reference Solvent (x_1' (solute-free) = 0.6815) equation 4 and equation 8.

3. Conclusions

We have used some solute solubility data on a reference solvent system to extend limited solubility data by strategic parameter fitting. Other challenges are:

1. mixed solutes in mixed solvents including complexes and competitive interactions
2. solute structure modifications; can the effect of adding or subtracting substituent groups be predicted? We expect this should be possible. Adding or changing groups may have only limited, and predictable, effects on solubility, even when making significant differences in product properties.
3. solute partitioning between phases in combined polymeric/aqueous phase media may also be expressible in reference terms.

References

- Abildskov J. and O'Connell J.P., 2003. *Ind. Eng. Chem. Res.*, 42, 5622-5634.
 Abildskov J. and O'Connell J.P., 2004a. *Molecular Simulation*, 30(6), 367-378.
 Abildskov J. and O'Connell J.P., 2004b. *Fluid Phase Equilibria* (in press).
 Acree Jr. W.E., 1984. *Int. J. Pharm.*, 18, 47-52.
 Bondi A., 1968, *Properties of Molecular Crystals and Glasses*. Wiley, New York.
 Choi P.B. and McLaughlin E., 1983. *AIChE J.*, 29, 150.
 Dickhut R., Andren A.W. and Armstrong D.E., 1989. *J. Chem. Eng. Data*, 34, 438.
 De Fina et al., 2002. *Phys. Chem. Liq.*, 40(3): 255.
 Ruckenstein E. and Shulgin I., 2001. *Int. J. Pharm.*, 258, 193-201.
 Acree Jr., W.E. and Abraham M.H., 2002. *Fluid Phase Equilibria*, 201, 245.

Acknowledgements

The authors are grateful for the encouragement of Professor R. Gani of CAPEC.