School of Pharmacy[1], Tabriz University of Medical Sciences, Tabriz, Iran, Faculty of Pharmacy[2], The University of Sydney, Sydney, Australia, Department de Farmacia y Tecnologia Farmaceutica[3], Facultad de Farmacia, Universidad de Alcala de Henares, Madrid, Spain, and Drug Applied Research Center[4], Tabriz University of Medical Sciences, Tabriz, Iran

Solubility prediction in water-ethanol mixtures based on the excess free energy approach using a minimum number of experimental data

A. Jouyban[1], H.K. Chan[2], S. Romero[3], M. Khoubnasabjafari[4], P. Bustamante[3]

Received July 9, 2003, accepted July 23, 2003

Dr. A. Jouyban, School of Pharmacy, Tabriz University of Medical Sciences, Tabriz 51664, Iran
ajouyban@hotmail.com

Pharmazie 59: 117–120 (2004)

The solubility of nalidixic acid in water-ethanol mixtures has been determined at 25 °C. The generated solubility data sets have been used to assess the accuracy of different numerical analyses for the excess free energy model and a new numerical method for solubility prediction in water-ethanol mixtures based on four experimental determinations is proposed. The accuracy of a previously presented numerical method to fit all data points is compared with that of a proposed analysis using average percentage deviation (APD). The APD obtained are 14.6 (±8.0) and 8.4 (±4.1), respectively for previous and proposed methods. A minimum number of three and four data points employed to train the original forms of the excess free energy model and the solubility at other solvent compositions were predicted. The APDs obtained were 61.4 and 23.0%, respectively. The APD produced for the proposed numerical method was 16.1%.

1. Introduction

Although a number of papers have been published in the past two decades dealing with the solubility of poorly water-soluble drugs in binary solvents and mathematical modelling of these data (Adjei et al. 1980; Yalkowsky and Roseman 1981; Acree and Rytting 1982; Williams and Amidon 1984; Rubino and Yalkowsky 1987; Acree 1992; Khossravi and Connors 1992; Bustamante et al. 1994; Barzegar-Jalali and Jouyban-Gh 1997; Jouyban-Gh 1998), solubility prediction based on a minimum number of experiments has not been comprehensively considered in the pharmaceutical literature. This has been discussed in a number of publications (Williams and Amidon 1988; Barzegar-Jalali and Jouyban-Gh 1996; Jouyban-Gh et al. 2001a; Jouyban-Gh et al. 2001b). Yalkowsky and his co-workers developed the log-linear model (Yalkowsky and Roseman 1981) which is applicable for solubility data profiles showing no solubility maximum, which is not the case for many solubility drug profiles in water-cosolvent mixtures. Williams and Amidon (1984) developed the excess free energy approach and then showed the possibility of reproducing a solubility profile in a given binary solvent using three experimental data points (Williams and Amidon 1988). Barzegar-Jalali et al. (1996) compared a double log-log model with two others employing a minimum number of three and five data points. All reported cosolvency models suffer from the use of a number of curve-fitting parameters in order to provide a satisfactory solubility prediction, and a model employing a minimum number of curve-fitting parameters would be more useful for this purpose. In an earlier work (Jouyban et al. 2002), the applicability of the excess free energy approach to calculate a solute solubility in mixed solvent at various temperatures using a minimum number of experimental data points was presented.

In this work, solubility prediction for nalidixic acid in water-ethanol mixtures using different numerical analyses based on a minimum number of data points and the excess free energy approach is discussed employing a generated and published data sets. Also, a new numerical method is proposed providing more accurate results in comparison with previous numerical analyses. Predictions using a few experimental data can be employed in drug liquid formulations and also in preformulation studies of a new drug/drug candidate where only a small quantity of the drug is available.

Williams and Amidon (1984) derived relationships between solute activity coefficient, solute's Henry law constants in pure solvents, and solute free cosolvent and water volume fractions at a constant temperature. These models were used to calculate solute solubility in binary solvents by employing Wohl's method for expressing the excess free energies of mixtures in terms of solute free volume fractions of the components (Williams and Amidon 1984). The three-suffix equation for a binary solvent system is:

$$\ln x_m = \phi_c \ln x_c + \phi_w \ln x_w - A_{c-w}\phi_c\phi_w (2\phi_c - 1)\left(\frac{V_2}{V_c}\right) + 2A_{w-c}\phi_c^2\phi_w\left(\frac{V_2}{V_w}\right) + C_2\phi_c\phi_w \quad (1)$$

where x_c, x_w and x_m represent the solubility mole fraction of the solute in the pure cosolvent, water and in the mixed solvent, ϕ_c and ϕ_w are the solute free volume fractions of

cosolvent and water, A_{c-w} and A_{w-c} stand for water-cosolvent interaction terms calculated from vapour-liquid equilibrium data, V_2, V_c and V_w are the molar volumes of the solute, cosolvent and water, respectively, and C_2 is the solute-solvent interaction term. From a theoretical point of view, C_2 is the only unknown constant, which can be calculated by:

$$\ln x_m - \phi_c \ln x_c - \phi_w \ln x_w + A_{c-w}\phi_c\phi_w(2\phi_c - 1)\left(\frac{V_2}{V_c}\right) - 2A_{w-c}\phi_c^2\phi_w\left(\frac{V_2}{V_w}\right) = C_2\phi_c\phi_w \quad (2)$$

The numerical values of C_2 can be computed by a regression analysis with a zero intercept least square method (Williams and Amidon 1984). In another paper, Williams and Amidon (1988) have suggested a single point estimation of the C_2 term. They have used solvent-solvent interaction terms equal to $A_{c-w} = 1.2160$ and $A_{w-c} = 0.9093$ for water-ethanol mixtures. By employing terms with physical meaning in this model, i.e. solvent-solvent and solvent-solute interaction or molar volume terms, it provides more evidence for a better understanding of the cosolvency phenomenon. This covers a part of the aims in solubility data modelling. Another important goal in mathematical modelling is to provide accurate quantitative relationships for correlating/predicting the experimental data points. The fitness (correlation) ability of a model can be used to identify the possible outliers in order to allow re-determination and the capability to provide the accurate predictions to distinguish the good from the poor models. As shown in this work, the numerical methods proposed by Williams and Amidon produced relatively high prediction errors. In order to provide more accurate predictions, since the values of A_{c-w}, A_{w-c}, V_2, V_c, V_w and C_2 for a solute in a given binary solvent are constant, it is possible to rewrite eq. (1) as:

$$\ln x_m = \phi_c \ln x_c + \phi_w \ln x_w + \left[A_{c-w}\left(\frac{V_2}{V_c}\right) + C_2\right]\phi_c\phi_w + \left[2A_{w-c}\left(\frac{V_2}{V_w}\right) - 2A_{c-w}\left(\frac{V_2}{V_c}\right)\right]\phi_c^2\phi_w \quad (3)$$

or

$$\ln x_m = \phi_c \ln x_c + \phi_w \ln x_w + M_1\phi_c\phi_w + M_2\phi_c^2\phi_w \quad (4)$$

where M_1 and M_2 are the model constants. These constants can be computed by regressing $\ln x_m - \phi_c \ln x_c - \phi_w \ln x_w$ against $\phi_c\phi_w$ and $\phi_c^2\phi_w$ using a no intercept least squares method.

The correlation/prediction capability of the models is evaluated by the average percentage deviation (APD):

$$APD = \frac{100}{N}\sum_1^N\left\{\frac{|(x_m)_{Observed} - (x_m)_{Calculated}|}{(x_m)_{Observed}}\right\} \quad (5)$$

where N is the number of data points in each set.

2. Investigations, results and discussion

The heating rate does not significantly influence the melting point or the molar heat of fusion ($T^F = 228.6$ °C, $\Delta H^F = 34.2$ kJ/mol at 2.5 °C/min, $T^F = 228.7$ °C, $\Delta H^F = 35.9$ kJ/mol at 5 °C/min, $T^F = 228.6$ °C, $\Delta H^F = 34.6$ kJ/mol at 10 °C/min, $T^F = 228.6$ °C, $\Delta H^F = 34.9$ kJ/mol at 20 °C/min, and $T^F = 228.2$ °C, $\Delta H^F = 35.8$ kJ/mol at 40 °C/min). These results agree with those obtained from microscopy (228.8 °C). The drug does not decompose during melting. After cooling the drug recrystallizes and during the second heating the melting point does not change. The temperature and heat of fusion do not vary after equilibration with the solvent mixtures. These results suggest the absence of polymorphic changes.

Table 1: Solubility of nalidixic acid at different ϕ_c values in mol/l (M) and mole fraction (x_m) concentration units

ϕ_c	M	x_m
0.00	0.00013524	0.000002417
0.10	0.00022029	0.000004655
0.20	0.00033522	0.000008128
0.30	0.00065144	0.000017948
0.40	0.00122270	0.000037752
0.50	0.00220040	0.000076219
0.60	0.00314840	0.000121030
0.70	0.00460030	0.000196960
0.80	0.00554310	0.000263500
0.85	0.00553980	0.000276030
0.90	0.00512400	0.000269520
0.95	0.00402400	0.000222860
1.00	0.00354800	0.000206340

The molar and mole fraction solubilities of nalidixic acid in water-ethanol mixtures at varying solvent compositions at 25 °C are shown in Table 1. The mole fraction solubility of nalidixic acid increases with an increase in ethanol concentration in the binary solvent system, it reaches a maximum value at $\phi_c = 0.85$ and then decreases with a further increase in ethanol concentration.

The accuracy of the excess free energy model is studied employing solubility of drugs in water-ethanol mixtures generated in this work and collected from the pharmaceutical literature (see Table 2). The two different numerical methods presented by Williams and Amidon (1984; 1988) and the new method proposed here were critically examined. Williams and Amidon calculated the solute-solvent interaction term, C_2, by a least squares analysis employing the all data points (Williams and Amidon 1984). All data points in each set are fitted to eqs. (2) and (4) and the back-calculated solubilities using eqs. (1) and (4) are used to compute APD values. This method is called correlative analysis and shows that how well a model could fit the experimental data. The mean and S.D. for APD of eqs. (1) and (4) are 14.6 (±8.0) and 8.4 (±4.1), respectively. The mean difference between APDs is statistically significant ($p < 0.02$) which means that eq. (4) shows more accurate correlations in comparison with eq. (1).

In another numerical method proposed by Williams and Amidon (1988) the C_2 term is also estimated using a single point solubility measurement in binary mixtures and the solubility at other solvent compositions are predicted using only three solubility data at $\phi_c = 0$, 0.50 and 1 and the corresponding APDs are shown in Table 2. This method is called predictive analysis. Both numerical methods proposed by Williams and Amidon are carried out in this work using twelve solubility data sets. As these methods produced relatively high APD values, an alternative numerical analysis based on eq. (4) was evaluated using four solubility data points with nearly constant ϕ_c intervals. In order to provide similar conditions for both eqs. (1) and (4), APD values were calculated by the 4 data points method. A summary of the results is also shown in Table 2. The mean differences between APD for eqs. (1) and (4) using four training data points is not statistically significant. However, eq. (4) provided less APD values and also there is no need to know vapour-

Table 2: Average percentage deviation (APD) of eqs. (1) and (4) for drug solubility in water-ethanol mixtures at 25 °C using different numerical analyses

	Solute	Reference	N[a]	All data points[b]		4 data points[c]		3 data points[d]
				Eq. (4)	Eq. (1)	Eq. (4)	Eq. (1)	Eq. (1)
1	Acetanilide	Stephen and Stephen 1964	13	4.8	5.2	10.7	5.9	30.5
2	DL-Alanine	Greenstein and Winitz 1961	7	11.8	25.6	27.7	40.0	48.8
3	Caffeine	Bustamante et al. 2002	11	7.9	8.2	12.2	12.2	16.5
4	Furosemide	Jouyban-Gh. et al. 2001b	13	11.2	17.9	14.1	22.1	179.4
5	Glycine	Greenstein and Winitz 1961	7	14.3	25.5	33.4	41.7	45.9
6	Nalidixic acid	This work	13	3.4	8.6	6.9	9.8	67.2
7	Niflumic acid	Bustamante et al. 2002	9	3.0	5.8	5.1	7.5	218.5
8	Oxolinic acid	Jouyban-Gh 2000	11	12.9	15.5	19.9	16.6	24.7
9	Paracetamol	Romero et al. 1996	13	6.6	7.6	9.1	10.6	23.3
10	Sulphamethiazine	Bustamante et al. 1994	11	10.6	25.7	20.9	62.8	41.7
11	Sulphanilamide	Bustamante et al. 1994	12	3.1	10.2	4.8	12.1	16.3
12	DL-Valine	Greenstein and Winitz 1961	7	11.3	19.7	28.1	34.8	23.8

[a] N: is number of data point. The numbers of data points for 4 and 3 data point predictions are (N-4) and (N-3), respectively.
[b] The model constants, i.e. M_1-M_2 and C_2, were computed using all solubility data points in each set and the back-calculated solubilities were used to calculate APD values (correlative analysis).
[c] M_1-M_2 and C_2 were computed using solubility at four volume fractions of ethanol with nearly constant intervals and solubility at other data points were predicted by the trained model and compared with experimental values (predictive analysis).
[d] C_2 was calculated using solubility at 0.00, 0.50 and 1.00 volume fractions of ethanol (predictive analysis).

liquid equilibrium data and molar volumes of the solute and solvents.

To test the effects of concentration units on correlation and prediction capabilities of eq. (4), $g \cdot L^{-1}$ and molar concentrations of nalidixic acid in water-ethanol are used. Although the numerical values of the model constants varies, there is no significant difference in APD values for correlative/predictive analyses considering $g \cdot L^{-1}$, molar or mole fraction solubilities. Therefore, there is no need to determine density of the saturated solutions in industry where the solubilization of a poorly soluble drug is the aim of a project. This could make easier the use of cosolvency models for an inexperienced researcher in industry/academia who needs to find a solution for the solubility problem. In other words, in order to estimate the solubility of a drug in a binary mixture, it is suggested to determine the solubility at four solvent compositions and then predict the solubility (in molar concentration) at other possible solvent compositions and the prediction error expected will approximately be 20%.

Equation (4) produced more accurate predictions than eq. (1) for the methods based on the all data point in a set and for only four data points as training set. In addition, to obtain more accurate predictions with eq. (4), there is no need to employ molar volumes and vapour-liquid equilibrium data. The Fig. shows the mean and standard deviation of the APD values employing twelve solubility data sets in water-ethanol mixtures. For both equations, the more training data points used the higher the accuracy achieved by the predictions. However, one of the aims of solubility data modelling in mixed solvents is to predict the maximum/minimum solubility with as few experimental determinations as possible. Employing higher training data point numbers may breach this aim. Therefore, a predictive model should provide acceptable predictions using a reasonable number of experimental data points. As an example of the limitations of this statement, predictions based on 3 data points using eq. (1) produce relatively high errors, however this occurs with just one more experimental determination in comparison with the four data points method, it is suggested to determine solute solubility at 0, 0.33, 0.66 and 1.00 volume fractions of the cosolvent and then the prediction of the solubility at other solvent compositions was made with eq. (4). It is obvious that by employing the four-suffix excess free energy model (Williams and Amidon 1984) and applying the same modifications as have been done on eq. (1), one is able to predict the solubility with higher accuracy. However, using this four-suffix model, a minimum number of 6 data points are required to calculate four model constants and x_c and x_w values.

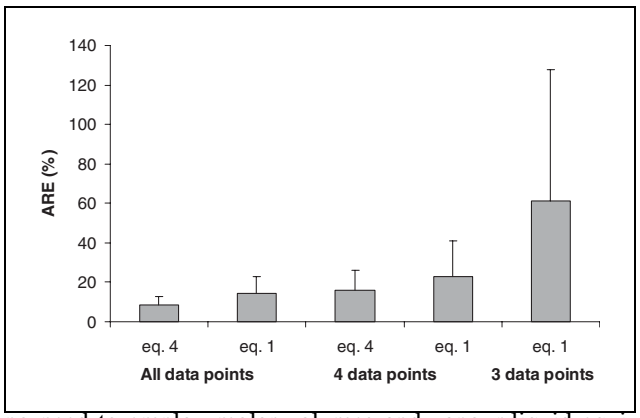

Fig.: Mean and standard deviation of APD values for eqs. (1) and (4) using different numerical analysis

In conclusion, solubility prediction based on a minimum number of experiments, a numerical method based on 4 data points is proposed, as the errors are reasonable. The mean APD is around 14% where the acceptable error range for solubility predictions in pharmaceutical subjects is about 30% (Beerbower et al. 1984; Dickhut et al. 1991; Reillo et al. 1995). Therefore, one can use these types of predictions in order to provide a rational drug formulation strategy rather than a trial and error approach when the optimisation of the cosolvent concentration is required.

3. Experimental

Nalidixic acid was purchased from Sigma. The purity of the lot employed in this work (124H0145) was >99%. Ethanol (spectrophotometric grade, Panreac, Monplet Spain) and double distilled water were used as the solvents.

An excess amount of nalidixic acid was added to sealed flasks containing the pure solvents and solvent mixtures and shaken at 25 °C temperature in

a temperature-controlled bath (±0.1 °C, Heto SH 02/100). The binary mixtures were prepared by volume. When the saturation concentration was attained, the solid phase was removed by filtration (Durapore membranes, 0.2 μm pore size). The drug did not significantly adsorb onto the membranes as shown from the similar results obtained in preliminary experiments (centrifugation and filtration). The clear solutions were diluted with ethanol 96% volume and assayed in a double beam spectrophotometer. The calibration line was obtained by preparing in triplicate 14 concentrations ranging from 1 to 8.5 μg/ml and measuring the absorbance at the wavelength of maximum absorption, 256 nm. The relationship is linear at this concentration range: absorbance = 0.1072 (±0.001) concentration +0.05732 (±0.056), n = 14, $r^2 = 0.999$, S.D. = 0.008, F = 11401. The intercept is not statistically different from zero. The densities of the solutions were determined in a 10-ml pycnometer to convert molarity units into mole fraction units.

The thermal behaviour of nalidixic acid was studied using differential scanning calorimetry (Mettler TA 4000) between room temperature and 400 °C at 5 heating rates ranging from 2.5 to 40 °C/min and from hot stage microscopy (Olimpus BX50, HFS91 heating rate 5 °C/min). A heating-cooling-heating cycle was also used with the DSC method to test possible variations of the melting point and/or polymorphism. The thermal behaviour of the solid phase after equilibration with the solvent system was also investigated to test possible solid phase changes induced by the solvent mixture that may affect the solubility behaviour.

Acknowledgements: The authors would like to thank the Australian Department of Education, Training and Youth Affairs and the University of Sydney for providing the IPRS and IPA scholarships. The experimental part was financed with project PM 99-0127, Ministerio de Ciencia y Tecnologia, Spain

References

Acree Jr, WE (1992) Mathematical representation of thermodynamic properties. Part 2. Derivation of the combined nearly ideal binary solvent (NIBS)/Redlich-Kister mathematical representation from a two-body and three-body interactional mixing model. Thermochim Acta 198: 71–79.

Acree Jr WE, Rytting JH (1982) Solubilities in binary solvent systems II. The importance of non-specific interactions. Int J Pharm 10: 231–238.

Adjei A, Newburger J, Martin A (1980) Extended Hildebrand approach: solubility of caffeine in dioxane-water mixtures. J Pharm Sci 69: 659–661.

Barzegar-Jalali M, Jouyban-Gh A, Hanaee J, Chookhachizadeh-Moghaddam MH (1996) Comparison of double log-log, mixture response-surface and combined NIBS/Redlich-Kister solubility models. Int J Pharm 144: 127–130.

Barzegar-Jalali M, Jouyban-Gh A (1997) A general model from theoretical cosolvency models. Int J Pharm 152: 247–250.

Beerbower A, Wu PL, Martin A (1984) Expanded solubility parameter approach. 1. Naphthalene and benzoic acid in individual solvents. J Pharm Sci 73: 179–188.

Bustamante P, Navarro J, Romero S, Escalera B (2002) Thermodynamic origin of the solubility profile of drugs showing one or two maxima against the polarity of aqueous and nonaqueous mixtures: niflumic acid and caffeine. J Pharm Sci 91: 874–883.

Bustamante P, Ochoa R, Reillo A, Escalera JB (1994) Chameleonic effect of sulfanilamide and sulfamethazine in solvent mixtures. Solubility curves with two maxima. Chem Pharm Bull 42: 1129–1133.

Dickhut RM, Armstrong DE, Andren AW (1991) The solubility of hydrophobic aromatic chemicals in organic solvent/water mixtures. Evaluation of four mixed solvent solubility estimation methods. J Environ Toxicol Chem 10: 881–889.

Greenstein JP, Winitz M (1961) Chemistry of amino acids, Vol. 1, New York, p. 547.

Khossravi D, Connors KA (1992) Solvent effects on chemical processes. I. Solubility of aromatic and heterocyclic compounds in binary aqueous solvents. J Pharm Sci 81: 371–379.

Jouyban-Gh A (1998) The modified wilson model and predicting drug solubility in water-cosolvent mixtures. Chem Pharm Bull 46: 1058–1061.

Jouyban-Gh A, York P, Hanna M, Clark BJ (2001a) Solubility prediction of salmeterol xinafoate in water-dioxane mixtures. Int J Pharm 216: 33–41.

Jouyban-Gh A, Dastmalchi S, Chan HK, Hanaee J, Javanmard A, Barzegar-Jalali M (2001b) Solubility prediction for furosemide in water-cosolvent mixtures using minimum number of experiments. Drug Dev Ind Pharm 27: 577–583.

Jouyban A, Romero S, Chan HK, Clark BJ, Bustamante P (2002) A cosolvency model to predict solubility of drugs at several temperatures from a limited number of solubility measurements. Chem Pharm Bull 50: 594–599.

Jouyban-Gh A, Romero S, Bustamante P, Clark BJ (2000) Multiple solubility maxima of oxolinic acid in mixed solvents and a new extension of Hildebrand solubility approach. Chem Pharm Bull 48: 175–178.

Reillo A, Cordoba M, Escalera B, Selles E, Cordoba Jr M (1995) Prediction of sulfamethiazole solubility in dioxane-water mixtures. Pharmazie 50: 472–475.

Romero S, Reillo A, Escalera B, Bustamante P (1996) The behaviour of paracetamol in mixtures of aprotic and amphiprotic-aprotic solvents. Relationship of solubility curves to specific and nonspecific interactions. Chem Pharm Bull 44: 1061–1064.

Rubino JT, Yalkowsky SH (1987) Cosolvency and cosolvent polarity. Pharm Res 4: 220–230.

Stephen H, Stephen T (1964) Solubilities of inorganic and organic compounds, Vol. 2, New York, p. 1169.

Williams NA, Amidon GL (1984) Excess free energy approach to the estimation of solubility in mixed solvent system. II. Ethanol-water mixtures. J Pharm Sci 73: 14–18.

Williams NA, Amidon GL (1988) The estimation of solubility in binary solvents: Application of the reduced 3-suffix solubility equation to ethanol-water mixtures. Pharm Res 5: 193–195.

Yalkowsky SH, Roseman TJ (1981) Solubilization of drugs by cosolvents. In: Yalkowsky SH (ed.) Techniques of solubilization of drugs, New York, p. 91–134.