Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Pharmaceutics 283 (2004) 83–88
international journal of pharmaceutics
www.elsevier.com/locate/ijpharm

Deviation from linearity of drug solubility in ethanol/water mixtures

Stephen G. Machatha[a,*], Pilar Bustamante[b], Samuel H. Yalkowsky[a]

[a] *College of Pharmacy, The University of Arizona, 1703 E, Mabel Street, Tucson, AZ 85721-0207, USA*
[b] *Departamento de Farmacia y Tecnología Farmacéutica, Universidad de Alcalá, Alcalá de Henares E-28871, Madrid, Spain*

Received 9 January 2004; received in revised form 25 June 2004; accepted 25 June 2004
Available online 20 August 2004

Abstract

A new empirical function that describes the deviation from linearity of solubility of a drug in an ethanol/water matrix is applied to the experimental data for 51 compounds. The proposed model is a more accurate predictor of the co-solvent solubility profile than a general third order polynomial with the same number of parameters. Both the root mean square error and average absolute error for the proposed model are significantly lower than those of existing models. The model also accurately predicts the fraction of co-solvent that gives maximum solubility (f_{max}).

Keywords: Deviation from linearity; Solubility in an ethanol/water matrix; Maximum solubility

1. Introduction

Organic co-solvents especially ethanol are among the most powerful solubilizing agents. The prediction of solubility profiles in ethanol/water mixtures is of paramount interest and it facilitates understanding all co-solvent systems.

Parabolic models of co-solvent solubilization of the form in Eq. (1) have been used for predicting solubility in binary mixtures.

$$\log S_{mix} = \log S_w + af_c + bf_c{}^2 \quad (1)$$

where, S_{mix} and S_w are the total solubilities in the co-solvent mixture and water, respectively, a and b are constants and f_c the volume fraction of co-solvent in the mixture.

Paruta et al. (1964) correlated solubility with a parabolic function of the dielectric constant of the

* Corresponding author. Tel: +1 520 626 4308; fax: +1 520 626 4063.
E-mail address: machatha@pharmacy.arizona.edu (S.G. Machatha).

0378-5173/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2004.06.020

solvent mixture. Martin et al. (1979, 1981) also proposed a parabolic relationship between solute solubility and the solubility parameter of a solvent mixture. Ruckenstein et al. (2003) applied fluctuation theory to generate a new parabolic model to predict solubility in aqueous mixed solvents.

Yalkowsky and Roseman, (1981) proposed a log-linear model in the form of Eq. (2), which describes the exponential increase in aqueous solubility for non-polar organic compounds as the co-solvent concentration is increased.

$$\log S_{\text{mix}} = \log S_{\text{w}} + \sigma f_{\text{c}} \tag{2}$$

The term σ defines the co-solvent solubilization power for a particular co-solvent–solute system whose value can be obtained experimentally from the slope of a plot of $\log S_{\text{mix}}$ versus f_{c}. Li An and Yalkowsky (1994) showed that for a given solvent there is a linear relationship between σ and the partition coefficient ($\log K_{\text{ow}}$) of the solute. They also observed that in semi-polar solutes the solubilization curves are linear upto $f_{\text{c}} = 0.5$, after which they sometimes become parabolic. This parabolic behavior is dependent on how close the polarity of the solute is to that of the mixture. They also showed that the use of end to half slope ($\sigma_{0.5}$) instead of the end to end slope (σ) is more appropriate for such compounds, therefore the initial solubility by ethanol is described by:

$$\log S_{\text{mix}} = \log S_{\text{w}} + \sigma_{0.5} f_{\text{c}} \tag{3}$$

In this paper we will show that the following model is consistent with both the parabolic and the log-linear models and is also a better predictor of solubility in ethanol/water mixtures than previously published models.

$$\log S_{\text{mix}} = \frac{\log S_{\text{w}} + a f_{\text{c}}}{1 + b f_{\text{c}} + c {f_{\text{c}}}^2} \tag{4}$$

where a, b and c are constants. When the fraction of co-solvent (f_{c}) is small, Eq. (4) can be approximated to the log-linear model described by Eq. (3). Note that the a term in Eq. (4) is the initial slope and is synonymous with $\sigma_{0.5}$ in Eq. (3). The empirical terms b and c characterize the change in solute–solvent interactions produced by increasing co-solvent concentration. The b term tends to affect the maximum solute solubility while c affects the terminal slope as f_{c} approaches unity.

This proposed model is compared to a general third order polynomial of the form:

$$\log S_{\text{mix}} = \log S_{\text{w}} + a' f_{\text{c}} + b' {f_{\text{c}}}^2 + c' {f_{\text{c}}}^3 \tag{5}$$

where a', b' and c' are empirically derived constants.

2. Method

2.1. Acquisition of data

The 51 compounds were arbitrarily selected and the published solubility data of Li An and Yalkowsky (1994) and Millard et al. (2002).

2.2. Statistical analysis

Non-linear regression was performed on the logarithmic solubility datausing WinCurve Fit Version 1.1.8, 2002, Kevin Rainer Software (Vict., Australia).

The root mean square errors (RMSE) were determined using the following relationship:

$$\text{RMSE} = \sqrt{\frac{\sum (\text{observed} - \text{predicted})^2}{n_{\text{points}}}} \tag{6}$$

where n_{points} is the number of experimental points in each data set. The average absolute error (AAE) was also determined using the relationship in Eq. (7).

$$\text{AAE} = \frac{\sum |\text{observed} - \text{predicted}|}{n_{\text{points}}} \tag{7}$$

t-Tests were performed using Microsoft Excel 1997 (Los Angeles, CA). The P value was determined using a paired *t*-test with a two-tailed distribution. The significance level was set at 0.05 hence, if the P value is <0.05 than the two data sets are considered to be significantly different. The partition coefficients were determined using C log $P^{®}$ (BioByte Corp., 1999), and references herein.

3. Results and discussion

Non-linear regression was run on the data for 51 compounds with 460 data points, using the models described by Eqs. (4) and (5) and the absolute average errors and the root mean square errors calculated for

Table 1
Absolute average errors and root mean square errors calculated from the two models, [Eqs. (4) and (5)]

Compounds	C log P[a]	n[b]	AAE		RMSE	
			(4)	(5)	(4)	(5)
Histidine	−3.73	8	0.015	0.027	0.018	0.033
Asparagine	−3.54	5	0.007	0.016	0.010	0.019
Glutamine	−3.37	5	0.005	0.006	0.006	0.007
Glycine	−3.21	10	0.014	0.065	0.016	0.072
Alanine	−3.12	10	0.008	0.049	0.010	0.055
Glycyglycine	−2.92	7	0.013	0.066	0.016	0.074
Tartaric acid	−2.78	12	0.002	0.005	0.002	0.005
Glutamic acid	−2.69	6	0.033	0.086	0.038	0.102
Amino-isobutyric acid	−2.62	5	0.003	0.006	0.004	0.007
Amino-*n*-butyric acid	−2.53	6	0.004	0.040	0.005	0.044
Aspartic acid	−2.41	9	0.046	0.087	0.062	0.104
DL-Valine	−2.29	7	0.023	0.054	0.026	0.058
Aminocaproic acid	−2.24	10	0.025	0.094	0.029	0.105
Hydantoin	−1.69	7	0.018	0.022	0.020	0.027
Leucine	−1.67	5	0.017	0.012	0.022	0.014
Tryptophan	−1.57	8	0.038	0.016	0.042	0.019
Phenylalanine	−1.56	8	0.015	0.034	0.018	0.040
Hydantoic acid	−1.38	6	0.013	0.031	0.016	0.020
Norleucine	−1.38	10	0.030	0.049	0.035	0.055
Zalcitabine	−1.29	11	0.017	0.021	0.022	0.026
Didanosine	−1.24	11	0.042	0.095	0.050	0.062
Formylglycine	−1.19	9	0.006	0.024	0.008	0.027
Methylhydantoic acid	−1.18	6	0.015	0.018	0.020	0.020
Triglycine	−0.94	7	0.045	0.070	0.057	0.079
5-Ethylhydantoin	−0.64	7	0.061	0.075	0.071	0.019
Formyl-aminobutyric acid	−0.35	7	0.041	0.009	0.051	0.011
Caffeine	−0.06	6	0.021	0.025	0.026	0.032
Zidovudine	0.04	11	0.016	0.016	0.022	0.024
Paracetamol	0.49	13	0.059	0.028	0.083	0.034
Formylleucine	0.58	8	0.030	0.053	0.043	0.063
Benzamide	0.65	14	0.015	0.009	0.019	0.012
Barbital	0.66	11	0.017	0.012	0.018	0.014
p-Aminobenzoic acid	0.98	6	0.019	0.031	0.023	0.036
Metharbital	1.14	11	0.015	0.026	0.018	0.028
Acetanilide	1.16	13	0.016	0.022	0.021	0.027
Phenobarbital	1.37	12	0.015	0.012	0.017	0.015
Oxolinic acid	1.55	11	0.041	0.057	0.049	0.063
Strychnine	1.66	7	0.035	0.019	0.038	0.054
Camphoric acid	1.75	12	0.036	0.035	0.047	0.052
Furosemide	1.87	13	0.130	0.108	0.148	0.133
Benzoic Acid	1.88	11	0.026	0.077	0.029	0.083
Benzocaine	1.92	11	0.030	0.045	0.038	0.048
Phenytoin	2.08	11	0.040	0.049	0.046	0.053
Alprazolam	2.19	9	0.019	0.027	0.022	0.035
Salicylic acid	2.19	6	0.007	0.062	0.007	0.067
Diazepam	2.99	11	0.043	0.044	0.054	0.057
Ibuprofen	3.68	8	0.102	0.100	0.120	0.123
β-estradiol	3.78	6	0.031	0.075	0.034	0.085
Biphenyl	4.03	11	0.067	0.086	0.069	0.092
Indomethacine	4.18	10	0.050	0.073	0.058	0.083
Anthracene	4.49	11	0.049	0.080	0.061	0.091

[a] C log P is the octanol/water partition coefficient.
[b] n is the number of experimental points in each data set.

Table 2
Average of the errors (AAE and RMSE) and the percent difference between the two models [Eqs. (4) and (5)]

	AAE	RMSE
Eq. (4)	0.029	0.035
Eq. (5)	0.044	0.049
Difference in error (%)	34.2	28.9

each of the models. The AAE and RMSE for the proposed model (4) and the third order polynomial (5) are listed in Table 1. The average errors and the percent difference in the errors from each model are shown in Table 2.

The percent difference in the AAE (34.2%) and the RMSE (28.9%) was large enough (>20%) to postulate that the two data sets are significantly different. A paired *t*-test gives *P* values of 0.00002 and 0.00036, respectively confirming the hypothesis. It can therefore, be concluded that the suggested model is a more accurate predictor of the ethanol/water solubility profile than a third order polynomial.

The model was also used to predict the fraction of ethanol that gives maximum solubility (f_{max}) (Table 3). The average absolute difference in the predicted and experimental value of f_{max} for all the compounds is only 0.0376. A *P* value of 0.4095 implies that the predicted f_{max} values are not significantly different from the experimental values. It should be noted that the fraction of ethanol producing maximum solubility tends to increase with increasing solute C log *P*.

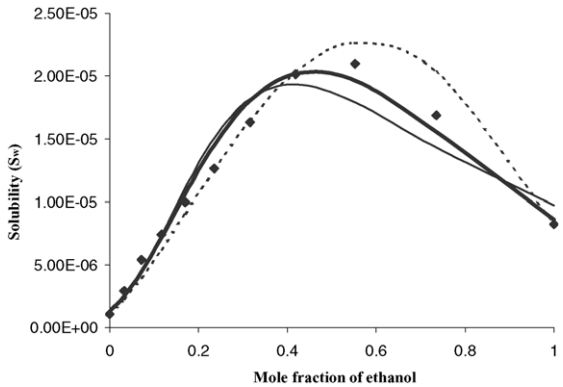

Fig. 1. Comparison between experimental and the predicted solubilities (mg/ml) [(· · ·) Ruckenstein, (—) proposed model and (–) third order polynomial] of oxolinic acid plotted against mole fraction of ethanol.

In Fig. 1 and Fig. 2 , the experimental ethanol/water solubility profile for oxolinic acid Jouyban et al. (2002), was compared to predicted solubilities using the third order polynomial and the suggested model. It was further compared to previously published data for the same solute performed by Ruckenstein and Shulgin (2003), who used the following equation based on fluctuation theory.

$$\ln X_2^t = \frac{(\ln V - \ln V_3)\ln X_2^{b1} + (\ln V_1 - \ln V)\ln X_2^{b3}}{\ln V_1 - \ln V_3} \quad (8a)$$

where, V, V_3, V_1, are the molar volumes of the solute, water and co-solvent, respectively. The terms $X_2^{b3}$and

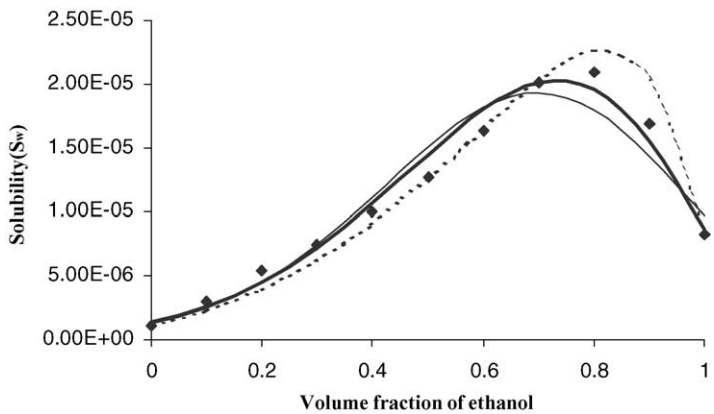

Fig. 2. Comparison between experimental and the predicted solubilities (mg/ml) [(· · · ·) Ruckenstein, (—) proposed model and (–) third order polynomial] of oxolinic acid plotted against volume fraction of ethanol.

Table 3
Comparison between predicted and experimental f_{max} values

Compounds	n[a]	Predicted	Experimental difference	
Paracetamol	13	1.00	0.85	0.15
Oxolinic acid	11	0.70	0.80	0.1
Methylhydantoic acid	6	0.50	0.60	0.1
Strychnine	7	0.80	0.80	0.0
Phenobarbital	12	1.00	0.90	0.1
Methobarbital	11	0.80	0.80	0.0
Indomethasine	10	1.00	1.00	0.0
Barbital	11	0.80	0.90	0.1
Benzoic acid	11	1.00	1.00	0.0
Anthracene	11	1.00	1.00	0.0
Biphenyl	11	1.00	1.00	0.0
Hydantoic acid	6	0.00	0.00	0.0
5-Ethylhydantoin	7	0.60	0.60	0.0
Hydantoin	7	0.00	0.00	0.0
Alprazolam	9	1.00	1.00	0.0
Diazepam	11	1.00	0.90	0.1
Didanosine	11	0.50	0.40	0.1
Furosemide	13	1.00	1.00	0.0
Zidovudine	11	0.70	0.70	0.0
Zalcitabine	11	0.50	0.30	0.2
Aspartic acid	9	0.00	0.00	0.0
Norleucine	10	0.00	0.00	0.0
DL-Valine	7	0.00	0.00	0.0
Glycyglycine	7	0.00	0.00	0.0
Histidne	8	0.00	0.00	0.0
Tryptophan	8	0.00	0.00	0.0
Alanine	10	0.00	0.00	0.0
Aminocaproic acid	10	0.10	0.00	0.1
Phenylalanine	8	0.00	0.00	0.0
Tartaric acid	12	0.00	0.00	0.0
Leucine	5	0.00	0.00	0.0
β-estradiol	6	1.00	1.00	0.0
Caffeine	6	0.60	0.60	0.0
Phenytoin	11	0.90	0.90	0.0
Ibuprofen	8	0.80	1.00	0.2
Benzocaine	11	1.00	0.90	0.1
p-Aminobenzoic acid	6	1.00	0.80	0.2
Salicylic acid	6	1.00	1.00	0.0
Camphoric acid	12	0.80	0.90	0.1
Glycine	10	0.00	0.00	0.0
Formylglycine	9	0.00	0.00	0.0
Formylleucine	8	0.80	0.90	0.1
Amino-*n*-butyric acid	6	0.00	0.00	0.0
Amino-isobutyric acid	5	0.00	0.00	0.0
Glutamic acid	6	0.00	0.00	0.0
Asparagine	5	0.00	0.00	0.0
Glutamine	5	0.00	0.00	0.0
Formyl-aminobutyric acid	7	0.90	0.80	0.1
Benzamide	14	0.80	0.83	0.03
Acetanilide	13	0.90	0.90	0.0
Triglycine	7	0.00	0.00	0.0

[a] n is the number of experimental points in each data set.

Table 4
AAE and RMSE calculated from the different models [Eq. (5), Ruckenstein and Eq. (4)] using oxolinic acid as the model compound

	AAE (log *S*)	RMSE (log *S*)
Ruckenstein's model	0.044	0.064
Eq. (4)	0.041	0.049
Eq. (5)	0.057	0.063

X_2^{b1} are the solute solubility in pure water and co-solvent.

The solute molar volume is determined by:

$$V = X_1V_1 + X_3V_3 + eX_1X_3 \quad (8b)$$

and is not necessarily equal to the experimental molar volume of the solute, where e is an empirical parameter and X_3 and X_1 are the molar volumes of co-solvent and water.

Note that although Eq. (8a) contains one less coefficient than Eq. (4), it requires an additional fitted value, i.e. the solubility in pure co-solvent.

4. Conclusion

It is apparent from Figs.1 and 2 and the average errors, in Table 4, that the suggested model is statistically a better predictor of the ethanol/water solubility profile and the f_{max} for oxolinic acid than the other models. The proposed model accounts for the initial log-linear relationship as well as the parabolic behavior of the solubility profile observed at higher fractions of ethanol. Furthermore this model has been proven to be statistically a better predictor of the ethanol solubility profile as well as the fraction of ethanol, which gives the maximum solute solubility (f_{max}).

Acknowledgements

Pilar Bustamante acknowledges the financial support provided by project PM 99-0127, Ministerio de Ciencia y Tecnología, Spain and a grant from the NATO Scientific Committee.

References

Jouyban, A., Romero, S., Chan, H.K., Clark, B.J., Bustamante, P., 2002. A cosolvency model to predict solubility of drugs at several different temperatures from a limited number of solubility measurements. Chem. Pharm. Bull. 50, 594–599.

Li An, Yalkowsky, S.H., 1994. Solubility of organic solutes in ethanol/water mixtures. J. Pharm. Sci. 83, 1735–1740.

Martin, A., Newburger, J., Adjei, A., 1979. New solubility equation. J. Pharm. Sci. 68, IV–V.

Martin, A., Paruta, A.N., Adjei, A., 1981. Extended Hildebrand solubility approach: methylxanthines in mixed solvents. J. Pharm. Sci. 70, 1115–1119.

Millard, J.W., Alvarez-Nunez, F.A., Yalkowsky, S.H., 2002. Solubilization by cosolvents establishing useful constants for the log-linear model. Int. J. Pharm. 245, 153–166.

Paruta, A.N., Sciarrone, B.J., Lordi, N.G., 1964. Solubility of salicylic acid as a function of dielectric constant. J. Pharm. Sci. 53, 1349–1353.

Ruckenstein, E., Shulgin, I., 2003. Solubility of drugs in aqueous solutions. Part 1. Ideal mixed solvent approximation. Int. J. Pharm. 258, 193–201.

Yalkowsky, S.H., Roseman, T.J., 1981. In Techniques of Solubilization of Drugs. Dekker, New York,, Chapter 3.