Available online at www.sciencedirect.com
SCIENCE DIRECT®
Thermochimica Acta 428 (2005) 119–123
thermochimica acta
www.elsevier.com/locate/tca

Modeling acid dissociation constant of analytes in binary solvents at various temperatures using Jouyban–Acree model

Abolghasem Jouyban[a,*], Somaieh Soltani[b], Hak-Kim Chan[c], William E. Acree Jr.[d]

[a] *School of Pharmacy and Drug Applied Research Center, Tabriz University of Medical Sciences, 29th Bahman St., Tabriz 51664, Iran*
[b] *Kimia Research Institute, P.O. Box 51665-171, Tabriz, Iran*
[c] *Faculty of Pharmacy, The University of Sydney, Sydney, NSW 2006, Australia*
[d] *Department of Chemistry, University of North Texas, Denton 76203-5070, USA*

Received 18 August 2004; received in revised form 19 October 2004; accepted 1 November 2004
Available online 13 December 2004

Abstract

A mathematical equation, namely Jouyban–Acree model, for calculating apparent acid dissociation constants (pK_a) in hydro-organic mixtures with respect to the concentration of organic solvent and temperature is proposed. The correlation ability of the model is evaluated by employing pK_a values of 17 different acids in water–cosolvent systems. The results show that the model is able to correlate the pK_a values with an overall average percentage differences (APD) of $1.71 \pm 1.86\%$. In order to test the prediction capability of the model, nine experimental pK_a values from each data set have been employed to train the model, then the pK_a values at other solvent compositions and temperatures were predicted and the overall APD obtained is $2.10 \pm 2.42\%$. The applicability of the extended form of Jouyban–Acree model on pK_a data of analytes in ternary solvent mixtures is also shown.

Keywords: Cosolvency; Acid dissociation constant; Hydro-organic mixtures; Different temperatures; Jouyban–Acree model

1. Introduction

During the lead optimization phase of drug discovery studies, where a large number of molecules are biologically evaluated in parallel, the determination of physico-chemical properties is essential to ensure adequate characterization and quality of developed candidates [1]. The knowledge of acid dissociation constants of these new chemical entities is of fundamental importance in order to provide information for scientists working on chromatographic separations (retention times and selectivity dependence of mobile phase pH) [2], capillary electrophoresis separations (migration times or mobilities of the ionic species over a range of pH values) [2], pharmaceutical drug discovery and development [3], chemical reactivity (selection of conditions for synthesis by considering the effects of pH on reaction products and properties of postulated intermediates), salt formation, purification process [1] and pharmaceutical formulation development [1].

Experimental measurements of pK_a values are expensive, difficult, time consuming and limited by purity of compounds (almost all experimental methods except capillary electrophoresis) [3], low analyte solubility (in potentiometry), the range of pH (in high performance liquid chromatography), spectral similarities (in spectrometric methods), and stability of analytes (e.g. chemical reactions intermediates). The most important limitation is that before synthesis of a compound, its pK_a value cannot be estimated experimentally.

Although water is the most common solvent in chemical/pharmaceutical applications, organic solvents are used as a cosolvent in order to adjust separation selectivity and modify solubility, stability, pK_a and other characteristics of the analytes. Temperature plays a significant role in most of the chromatographic and electrophoretic methods and it is recognized as the most relevant parameter in gas chromatography. Many separation scientists have traditionally

* Corresponding author. Tel.: +98 411 3392584; fax: +98 411 3344798.
E-mail address: ajouyban@hotmail.com (A. Jouyban).

0040-6031/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.tca.2004.11.002

disregarded temperature effects in liquid chromatography whereas elevated temperature is a usual controlling variable in reversed phase liquid chromatography. Today publications stated the reasons for fearing of such methods and now temperature in combination with pH and cosolvent addition are introduced as variables to adjust selectivity towards acidic and basic compounds e.g. see Refs. [4–6].

Since the pK_a value of many compounds is determined commonly at 25 °C, prediction of pK_a at different temperatures (e.g. body temperature 37 °C) would be very useful tool for biological and biomedical applications. Using mixed solvents in the analytical/pharmaceutical areas is a common method to optimize solubility and/or separation efficiency. However, the number of solvent compositions and temperature combinations is quite large and it is difficult to determine all possible combinations by experiments. Thus, a good alternative is to use computational methods. The aim of this work is to present a mathematical treatment of pK_a values as a function of solvent composition and temperature. The accuracy of the proposed model is assessed by using available pK_a values in mixed solvents at various temperatures collected from the literature.

2. Theoretical treatment

A dissociation reaction of a monoprotic acid (HA) in a solvent can be represented as:

$$HA \leftrightarrow H^+ + A^-, \quad K_a = \frac{a_{H^+} \cdot a_{A^-}}{a_{HA}}$$

where a is the activity of the chemical species. The logarithm of K_a expressed as:

$$2.303RT \log K_a = \mu_{H^+} + \mu_{A^-} - \mu_{HA} \tag{1}$$

where μ denotes chemical potential of the species.

The chemical potential in aqueous cosolvent mixtures can be expressed as the mole fractions of the cosolvents:

$$\mu^m_{H^+} = X^c\mu^c_{H^+} + X^w\mu^w_{H^+} + A_0X^cX^w + A_1X^cX^w(X^c - X^w) \tag{2}$$

$$\mu^m_{A^-} = X^c\mu^c_{A^-} + X^w\mu^w_{A^-} + B_0X^cX^w + B_1X^cX^w(X^c - X^w) \tag{3}$$

$$\mu^m_{HA} = X^c\mu^c_{HA} + X^w\mu^w_{HA} + C_0X^cX^w + C_1X^cX^w(X^c - X^w) \tag{4}$$

where superscripts m, c and w denote mixed solvent, pure cosolvent and pure water, respectively, X is the mole fraction of the solvents, and A, B and C the solute–solvent and solvent–solvent interaction terms. These terms represent the two body and three body interactions in the solution [7].

Summation of Eqs. (2)–(4) yields:

$$\mu^m_{H^+} + \mu^m_{A^-} - \mu^m_{HA} = X^c(\mu^c_{H^+} + \mu^c_{A^-} - \mu^c_{HA}) + X^w(\mu^w_{H^+} + \mu^w_{A^-} - \mu^w_{HA}) + (A_0 + B_0 - C_0)X^cX^w + (A_1 + B_1 - C_1)X^cX^w(X^c - X^w) \tag{5}$$

Replacing the corresponding equals from Eq. (1) into Eq. (5) with appropriate rearrangements give:

$$2.303R \log K_a^{m,T} = X^c(2.303R \log K_a^{c,T}) + X^w(2.303R \log K_a^{w,T}) + (A_0 + B_0 - C_0)\frac{X^cX^w}{T} + (A_1 + B_1 - C_1)\frac{X^cX^w(X^c - X^w)}{T} \tag{6}$$

Since $(A_0 + B_0 - C_0)$, $(A_1 + B_1 - C_1)$ and $2.303R$ are constant values and $pK_a = -\log K_a$, it is possible to simplify Eq. (6) as:

$$pK_a^{m,T} = X^c\,pK_a^{c,T} + X^w pK_a^{w,T} + W_0\frac{X^cX^w}{T} + W_1\frac{X^cX^w(X^c - X^w)}{T} \tag{7}$$

where $W_0 = (A_0 + B_0 - C_0)/2.303R$ and $W_1 = (A_1 + B_1 - C_1)/2.303R$. It is obvious that one can use the volume /weight fractions of the solvents instead of the mole fractions [7] and rewrite Eq. (7) as:

$$pK_a^{m,T} = f^c\,pK_a^{c,T} + f^w\,pK_a^{w,T} + K_0\frac{f^cf^w}{T} + K_1\frac{f^cf^w(f^c - f^w)}{T} \tag{8}$$

where K_0 and K_1 are the curve-fitting parameters. The numerical values of K_0 and K_1 can be computed by fitting the experimental values of $(pK_a^{m,T} - f^c\,pK_a^{c,T} - f^w\,pK_a^{w,T})$ against $\frac{f^cf^w}{T}$ and $\frac{f^cf^w(f^c-f^w)}{T}$ by using a no intercept least square analysis.

The general form of the proposed equation can be expressed as:

$$pK_a^{m,T} = f^c\,pK_a^{c,T} + f^w\,pK_a^{w,T} + f^cf^w\sum_{q=0}^{n}\frac{K_q(f^c - f^w)^q}{T} \tag{9}$$

The model could possess as many curve-fitting parameters as needed for accurate representation of experimental data. However, it is preferred to employ the lowest number of curve-fitting parameters, since it requires minimum number of experimental data in model training process. In some cases, the numerical values of pK_a^c are not available. If so, it is pos-

sible to rewrite Eq. (8) as:

$$pK_a^{m,T} = f^w pK_a^{w,T} + Jf^c + K_0\frac{f^c f^w}{T} + K_1\frac{f^c f^w(f^c - f^w)}{T} \quad (10)$$

where J, K_0 and K_1 are the model constant. These are computing via fitting ($pK_a^{m,T} - f^w pK_a^{w,T}$) against f^c, $f^c f^w$ and $f^c f^w(f^c - f^w)$. Eq. (10) can be written as Eq. (11) for calculating pK_a values in mixed solvents at a fixed temperature [7]:

$$pK_a^{m,T} = f^w pK_a^w + Mf^c + M_0 f^c f^w + M_1 f^c f^w(f^c - f^w) \quad (11)$$

where M, M_0 and M_1 are the model constants. Eq. (11) produced reasonably accurate results for both calculation of pK_a of a single analyte and a number of related analytes in binary solvents [7]. Eq. (11) can also be extended to calculate pK_a of analytes in ternary solvent mixtures as:

$$\begin{aligned} pK_a^m = {} & f^w pK_a^w + f^c pK_a^c + f^s pK_a^s + Q_0 f^w f^c \\ & + Q_1 f^w f^c(f^w - f^c) + Q_0' f^w f^s \\ & + Q_1' f^w f^s(f^w - f^s) + Q_0'' f^c f^s \\ & + Q_1'' f^c f^s(f^c - f^s) + Q_0''' f^w f^c f^s \\ & + Q_1''' f^w f^c f^s(f^w - f^c - f^s) \end{aligned} \quad (12)$$

where superscript s denote the second cosolvent, and Q terms are the model constants. If the numerical values of pK_a of analytes in pure solvents (e.g. pK_a^c or pK_a^s) are not available, it is possible to consider them as constant terms. The other versions of Eq. (9) have been successfully applied for calculation of solute solubilities in mixed solvents at various temperatures [8], dielectric constants [9] and also surface tensions [10] of liquid mixtures. The model could be considered empirical in nature, however, it produces accurate calculations and has its position in data modeling of physico-chemical properties in mixed solvents.

To assess the accuracy of the proposed equations, the average percentage differences (APD) between experimental and calculated pK_a values are considered as an accuracy criterion:

$$APD = \left(\frac{100}{N}\right)\sum\left|\frac{pK_a^{calculated} - pK_a^{observed}}{pK_a^{observed}}\right|$$

where N denotes the number of experimental data point in each set. The individual percentage difference (IPD) is calculated by:

$$IPD = 100\left|\frac{pK_a^{calculated} - pK_a^{observed}}{pK_a^{observed}}\right|$$

3. Results and discussion

In order to evaluate the accuracy of the proposed model, available pK_a values in different concentrations of the organic solvents at various temperatures including more than 15 data points were collected from the literature. The details of the collected data including the solute, the cosolvent, the number of experimental data points in each set, the temperature range and the references are listed in Table 1.

To evaluate the correlation ability of the Jouyban–Acree model, all data points in each set have been fitted to Eq. (9) with $q = 0$, 1 and 2, and the back calculated pK_a values have been used to compute APDs. This numerical method has been called correlative analysis and the results are shown in Table 1. The overall APD and the standard deviations for $q = 0$, 1, and 2 are 1.71 ± 1.86, 1.26 ± 1.30 and 0.99 ± 0.85, respectively. There is no significant differences in the overall APD between $q = 0$ and $q = 1$ or between $q = 1$ and 2 (paired t-test, $P > 0.07$). Therefore, Eq. (9) with $q = 0$ is selected as appropriate version of the model. For Eq. (9) possessing three constant terms (i.e. $pK_a^{w,T}$, J and K_0), the least APD value (0.27%) has been observed for eriochrome black T in water-ethanol and the highest APD value (4.98%) has been observed for maleic acid in water–ethanol mixtures. The IPD values produced by correlative analysis of Eq. (11) with $q = 0$ sorted in three subgroups, i.e. IPD ≤ 2, 2–4 and >4% are illustrated in Fig. 1. As seen in more than 75% of the cases the IPD is less than 2% whereas the experimental relative standard deviation for repeated experiments is up to 20% [3].

In order to evaluate the prediction capability of the selected version of the Jouyban–Acree model, nine experimental pK_a values (high, low and medium T, f^c and f^w) have been used to train the model and then the trained models are then employed to predict the pK_a values of remainder points in each data set. The minimum prediction APD (0.19%) is observed for adipic acid in water–methanol and the maximum value (9.12%) for pK_{a2} of tartaric acid in water–ethanol mixtures. The overall APD obtained is 2.10 ± 2.42%. Fig. 1 also shows the relative frequency of IPD values for predictive analysis. As expected, error percentage less than 2% shows the highest relative frequency.

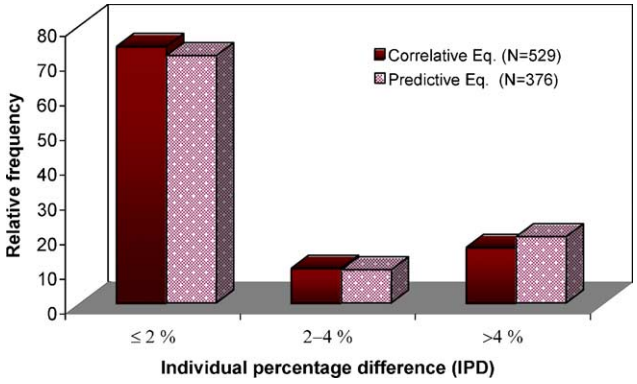

Fig. 1. Relative frequency of IPD values for correlative and predictive analyses.

Table 1
Details of the systems studied and the average percentage deviation (APD) for correlative and predictive analyses

No.	Solute	Cosolvent	Temperature (°C)	Reference	N[a]	Correlative			Predictive, $q=0$[b]
						$q=0$	$q=1$	$q=2$	
1	Adipic acid, pK_{a1}	Methanol	30–60	[14]	16	0.35	0.35	0.35	0.53
2	Adipic acid, pK_{a2}	Methanol	30–60	[14]	16	2.07	2.07	2.07	2.53
3	EBB[c], pK_{a1}	Dimethylformamide	20–40	[15]	26	0.55	0.49	0.42	0.64
4	EBB[c], pK_{a2}	Dimethylformamide	20–40	[15]	26	0.48	0.44	0.41	0.57
5	EBB[c], pK_{a1}	Ethanol	20–45	[15]	30	0.53	0.53	0.52	0.63
6	EBB[c], pK_{a2}	Ethanol	20–45	[15]	30	0.44	0.43	0.40	0.49
7	EBT[d], pK_{a1}	Ethanol	20–40	[15]	25	0.27	0.22	0.22	0.3
8	EBT[d], pK_{a2}	Ethanol	20–40	[15]	25	0.47	0.42	0.39	0.49
9	Maleic acid, pK_{a1}	Ethanol	30–55	[16]	36	4.98	4.77	3.05	5.63
10	Maleic acid, pK_{a2}	Ethanol	30–55	[16]	36	1.91	1.70	1.62	2.22
11	Phthalic acid, pK_{a1}	Ethanol	30–55	[16]	36	1.55	1.40	1.30	2.56
12	Phthalic acid, pK_{a2}	Ethanol	30–55	[16]	36	1.24	1.15	1.08	1.37
13	PAN[e]	Ethanol	20–30	[17]	33	4.94	1.37	1.15	5.12
14	Tartric acid, pK_{a1}	Ethanol	30–55	[16]	36	1.68	1.06	0.55	1.99
15	Tartric acid, pK_{a2}	Ethanol	30–55	[16]	36	6.21	4.03	2.46	9.12
16	Terazodone	Ethanol	15–45	[18]	36	0.71	0.7	0.70	0.66
17	Tris[f]	2-Methoxyethanol	20–45	[19]	50	0.68	0.32	0.17	1.08
Overall APD S.D.						1.71 ± 1.86	1.26 ± 1.30	0.99 ± 0.85	2.10 ± 2.42

[a] The number of data points in each set.
[b] APD of predictive analysis using models trained by nine experimental data points. The number of predicted points is $N-9$.
[c] Eriochrome blue black RC.
[d] Eriochrome black T.
[e] 1-(2-Pyridylazo)-2-naphthol.
[f] Tris-(hydroxymethyl) aminomethane.

Four available pK_a data sets in ternary solvents at 25 °C taken from the literature [11] are used to check the applicability of Eq. (12) to reproduce such data. The APDs of pK_a of heptanoic, hexanoic, pentanoic and butanoic acids in water–methanol–dioxane are 0.41, 0.38, 0.48 and 0.39%, respectively and the overall APD is 0.42%. All data points of alkanoic acids in water–methanol–dioxane are fitted to Eq. (12) and the obtained APD is 0.56%. To check the prediction capability of the model to predict the pK_a data, experimental data of heptanoic and butanoic acids are used to train the model, and then pK_a of hexanoic and pentanoic acids are predicted. The resulted prediction of APD is 0.51%, which shows a good predictability of the proposed model. It is obvious that such data is required in analytical [12] and pharmaceutical [13] areas where ternary solvents are used for mobile phases and/or drug formulations. To the best of our knowledge, there is no published pK_a data in ternary solvents at various temperatures to evaluate the accuracy of the proposed model. However, it is anticipated that the model is able to predict such data and it could be employed in generating pK_a data in ternary and even higher order multicomponent systems at various temperatures after training by a minimum number of experimental data.

In conclusion, the APD value for correlative studies using 17 data sets and employing a three constant term model is $1.71 \pm 1.86\%$ ($N = 529$) and for predicted data points using the trained models the APD value is $2.10 \pm 2.42\%$ ($N = 376$). These low correlation and prediction errors mean that the proposed model is able to calculate pK_a values in binary solvents at various temperatures within an acceptable error range. The corresponding values for pK_a of analytes in ternary solvents at a fixed temperature are 0.56 and 0.51%, respectively. It is suggested that employing the proposed models could help the researchers to speed up the optimization procedure.

References

[1] G.A. Caliaro, C.A. Herbotes, J. Pharm. Biomed. Anal. 26 (2001) 427.
[2] G. Lunn, HPLC Methods for Pharmaceutical Analysis, Wiley, Germany, 2000.
[3] A. Avdeef, K.J. Box, J.E.A. Comer, M. Gilges, M. Hadley, C. Hibbert, W. Patterson, K.Y. Tam, J. Pharm. Biomed. Anal. 20 (1999) 631.
[4] Y. Guillaume, C. Guinchard, J. Liq. Chromatogr. 17 (1994) 2809.
[5] C.B. Castells, C. Rafols, M. Roses, E. Bosch, J. Chromatogr. A 1002 (2003) 41.
[6] S. Heinisch, J.L. Rocca, J. Chromatogr. A 1048 (2004) 183.
[7] A. Jouyban, H.K. Chan, B.J. Clark, W.E. Acree Jr., Int. J. Pharm. 246 (2002) 135.
[8] A. Jouyban-Gh., W.E. Acree Jr., Int. J. Pharm. 167 (1998) 177.
[9] A. Jouyban, Sh. Soltanpour, H.-K. Chan, Int. J. Pharm. 269 (2004) 353.
[10] A. Jouyban, A. Fathi-Azarbayjani, W.E. Acree Jr., Chem. Pharm. Bull. 52 (2004) 1219.
[11] G. Papanastasiou, I. Ziogas, I. Moumtzis, Anal. Chim. Acta 186 (1986) 213.

[12] A. Jouyban, S.C. Grosse, H.-K. Chan, M.W. Coleman, B.J. Clark, J. Chromatogr. A 994 (2003) 191.
[13] A. Martin, P.L. Wu, T. Velasquez, J. Pharm. Sci. 74 (1985) 277.
[14] G.A. El-Naggar, Talanta 47 (1998) 1013.
[15] M.S. Masoud, H.H. Hammud, H. Beidas, Thermochim. Acta 381 (2002) 119.
[16] A.A. Zaghloul, G.A. El-Naggar, Sh.A. El-Shazly, M.F. Amira, Talanta 45 (1997) 189.
[17] L. Coo, J.T. Cardwell, R.W. Cattrall, S.D. Kolev, Anal. Chim. Acta 360 (1998) 153.
[18] I. Shehatta, Helv. Chim. Acta 85 (2002) 2125.
[19] A.A. El-Harakany, F.M. Abdel Halim, A.O. Barakat, J. Electroanal. Chem. 162 (1984) 285.