New chemometric strategies in the spectrophotometric determination of pKa

Judith Amador-Hernández [a,*], Alberto Rojas-Hernández [b], Edith Madaí Colunga-Urbina [a],
Iliana Margarita De La Garza-Rodríguez [a], Miguel Velázquez-Manzanares [c],
and Luis Felipe Medina-Vallejo [a]

[a] *Facultad de Ciencias Químicas, Universidad Autónoma de Coahuila, Blvd. V. Carranza s/n, 25280 Saltillo, Coahuila, México*
[b] *Departamento de Química, Universidad Autónoma Metropolitana, Apdo. Postal 55-534, 09340 México, Distrito Federal, México*
[c] *Instituto de Biotecnología, Universidad del Papaloapan, Circuito Central 200, 68301 San Juan Bautista Tuxtepec, Oaxaca, México*

*Corresponding author at: Facultad de Ciencias Químicas, Universidad Autónoma de Coahuila, Blvd. V. Carranza s/n, 25280 Saltillo, Coahuila, México.
Tel.: +52.844.4159534 (Ext. 112). Fax: +52.844.4159534. E-mail address: amadorjudith@live.com.mx (J. Amador).

ARTICLE INFORMATION

ABSTRACT

In this work, principal component regression and partial least squares regression were used for the estimation of acid dissociation constants through UV-Vis spectrophotometric measurements, considering five well-known acid-base indicators as well as two herbicides as analytes. In each case, an acid-base titration was carried out. Then, the multivariate calibration model was constructed with a few absorption spectra of the series at extreme pH values, to which values of the dissociation fraction (α) of 1 or 0 were assigned, in the case of HA or A species. After that, the prediction step consisted in the estimation of α for the rest of the series. Then, distribution diagrams were built up with α *vs* pH, to find $\alpha = 0.5$ where pH = pK_a. The results were compared with those obtained through multivariate curve resolution-alternating least squares and program stability quotients from absorbance data (SQUAD), which showed an excellent correspondence.

1. Introduction

The study of acid-base equilibriums is of great importance, because the ionic and the neutral forms of a compound exhibit different physicochemical properties (solubility, partition coefficient, etc.) [1,2]. In other words, the predominance of one of the two forms will condition their distribution in the environment, their biological activity or chemical reactivity, among others.

Thus, the determination of acid dissociation constants is a topic of current interest. Several research groups are continuing to look for new methodologies to estimate them, based on techniques such as Nuclear Magnetic Resonance [3], Electric Impedance Spectroscopy [4], Capillary Electrophoresis [5], or Gas Chromatography [6], to name a few. Furthermore, the use of UV-Visible spectrophotometry is still common, through new approaches to data processing [7-10].

Particularly, the technique of Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is increasingly used for the analysis of component mixtures, both for quantitative measures and for the study of chemical equilibriums [7,11,12]. However, Partial Least Squares Regression (PLS) or Principal Component Regression (PCR)

have not been reported for pKa estimation as far as is known to the authors.

In this work, the techniques of MCR-ALS, PLS and PCR are used for the estimation of acid dissociation constants through UV-Vis spectrophotometry. The substances of interest were five acid-base indicators, whose pK_a values are widely reported in the literature, in order to show the applicability of the techniques. Later, the pK_a values for isomethiozin and methoprotryne, two triazine herbicides, are reported. In all cases, the pK_a values were compared with those estimated through SQUAD [13,14], an algorithm that is widely used for this purpose [15,16].

1.2. Fundamental considerations

Consider the general reaction for an acid dissociation:

$$HA + H_2O \rightleftarrows H_3O^+ + A^- \tag{1}$$

where HA and A⁻ represent the acid and its conjugated base, respectively. From the expression of the acid-base equilibrium, it follows the Henderson-Hasselbalch equation:

$$pH = pK_a + log \frac{[A^-]}{[HA]} \qquad (2)$$

According to this, at the equilibrium the concentration of the acid and its conjugated base is the same when the pH equals to pK_a. Similarly, through expression 2 it can be deduced that:

If $pH \leq pK_a - 2$ then $[HA] \geq 0.99 C_T$, $[A^-] \leq 0.01 C_T$ $\qquad (3)$

If $pH \geq pK_a + 2$ then $[HA] \leq 0.01 C_T$, $[A^-] \geq 0.99 C_T$ $\qquad (4)$

where C_T is the total concentration of the analyte, i.e. $C_T = [A^-] + [HA]$. Finally, the fractions (α) of the chemical species involved are:

$$\propto_{HA} = \frac{[HA]}{C_T} \qquad (5)$$

$$\propto_{A^-} = \frac{[A^-]}{C_T} \qquad (6)$$

1.2.1. MCR-ALS

Resolution techniques decompose the instrumental responses of mixtures (original data matrix, **X**) into the contributions linked to each of the pure components in the system (concentration profiles matrix, **C**) and the pure response profiles matrix (**S**^{T}), according to:

$$X = CS^T + E \qquad (7)$$

where **E** corresponds to the residuals of the modeling.

The mandatory prerequisite is an inner linear structure of the data set. Also, three constraints are commonly used to reduce the ambiguity in data decomposition: a) nonnegativity, which means that the instrumental values will always be positive; b) unimodality, which considers only one maximum per profile; and c) closure, i.e. that the sum of the concentrations for the whole species involved is constant through the experiment. Particularly, MCR-ALS uses an approach to iteratively find the matrices of concentration profiles and instrumental responses, so that neither the **C** nor the **S**^{T} matrices have priority over the other [17].

For an acid-base reaction monitored through UV-Vis spectrophotometry, the prerequisite for a linear structure of the data set is fulfilled according to the Beer Law, as well as the three constraints mentioned above. In this case, the MCR-ALS decomposition of **X** is represented in Figure 1. Once the concentration profiles of pure components are obtained, they are considered an expression of the fractions of dissociation (α), so $pK_a = pH$ if $\alpha = 0.5$ for both the acid and the base species, according to equations 2, 5 and 6.
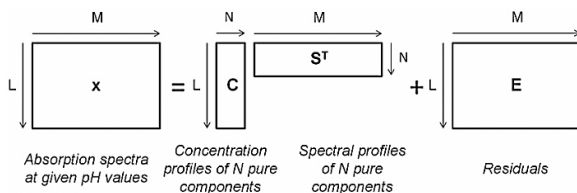

Figure 1. Estimation of pK_a through MCR-ALS. The dimensions of matrices are: (*L*) number of samples at given pH values, (*M*) number of wavelengths under study, (*N*) number of pure components in the mixture (two in this case).

1.2.2. PCR and PLS

Both techniques correspond to multivariate calibration, whose main purpose is to involve multiple measurements to predict the value of an underlying parameter or property. They are especially valuable in the analysis of multicomponent systems characterized by non-selective measurements, noisy data, etc.

PCR and PLS are based on Principal Component Analysis (PCA), whereby the dimensionality of the data set consisting of a large number of interrelated variables (**X**) is reduced to a new set of variables, named Principal Components (PCs), which retain as much as possible of the main information [18]. In general, the modeling of the underlying parameter or property (**Y**) can be represented according to:

$$T = XV \qquad (8)$$

$$X = TP' + E \qquad (9)$$

$$Y = TQ' + F \qquad (10)$$

where **T** is the scores' matrix; the loading matrix **P'** represents the regression coefficients of **X** on **T** in the same way as **Q'** represents the regression coefficients of **Y** on **T**; **E** and **F** are residual matrices. **V** contains coefficients estimated through many different methods. For PLS, **X** and **Y** are used simultaneously to estimate **V**. In PCR, **V** is estimated from **X**, so **Y** is only used to select how many factors will be integrated in the model [19].

Experimentally, two stages are carried out: calibration and prediction. In the first one, a series of objects described by **X** and **Y** are used to build the calibration model, after the proper selection of PCs. Secondly, equation 10 is used to estimate **Y** for objects only described by **X**.

In this case, at least four absorption spectra (two at each extreme of the samples series at different pH values) were considered to construct the calibration model, while assigning the corresponding α values of 0 or 1 as **Y**, due to the fulfillments of equations 3 and 4. Calibration models were built after the selection of the optimal number of PCs. Finally, α values were predicted for the rest of the samples and made up the distribution diagrams, in which the pH = pK_a if $\alpha = 0.5$ (Figure 2).

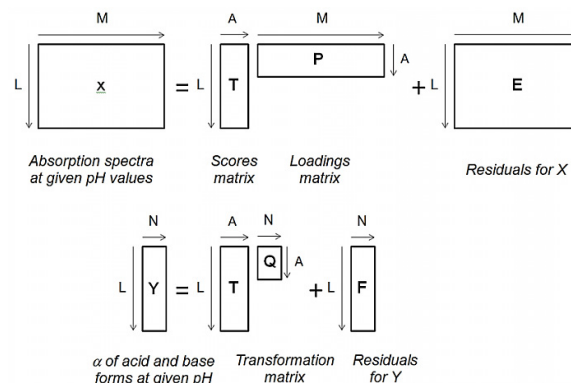**Figure 2.** Estimation of pK_a through PCR and PLS. The dimensions of matrices are: (*L*) number of samples at given pH values, (*M*) number of wavelengths under study, (*N*) number of components in the mixture (two in this case), (*A*) number of optimal PCs.

1.2.3. SQUAD

SQUAD is a program, created by David J. Leggett [13-16] that fits a set of absorption spectra corresponding to systems with variable chemical composition using a model (considering the formation constants of several chemical equilibria, the absorptivity coefficients for species and Beer's and additivity spectroscopic laws), by means of a Non Linear Least Squares algorithm. This program has been extensively used to determine equilibrium constants by UV-Visible measurements.
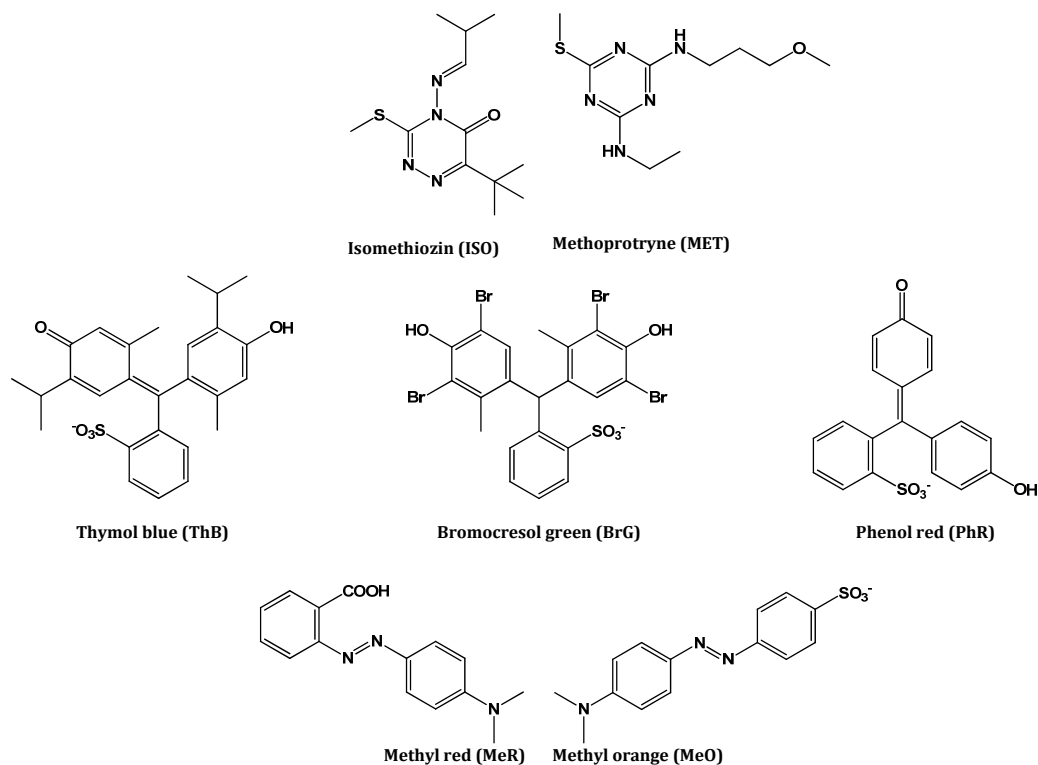
Figure 3. Developed formulae of the substances studied in this contribution.

Table 1. pK$_a$ values determined through UV-Vis spectrophotometry, using four algorithms for data processing.

Compound	Spectral range (nm)	pH range studied	C_T (M)	pK$_a$ expected [21,22]	pKa SQUAD	pKa MCR-ALS	pKa PLS	pKa PCR	Mean value ± standard deviation
PhR	370-600	4.7-10.8	1.02×10^{-5}	7.9	7.39	7.40	7.41	7.40	7.40±0.01
ThB	360-600	1.0-7.0	9.35×10^{-6}	1.6	1.49	1.55	1.54	1.53	1.53±0.03
MeO	350-580	2.0-7.2	1.50×10^{-5}	3.8, 3.4	3.33	3.33	3.35	3.36	3.34±0.02
MeR	350-600	2.3-8.0	4.46×10^{-5}	5.1, 4.9	4.70	4.69	4.68	4.68	4.69±0.01
BrG	400-700	1.0-8.0	1.07×10^{-5}	4.5, 4.7	4.37	4.33	4.33	4.34	4.34±0.02
ISO	205-400	0.2-5.5	4.80×10^{-5}	-	0.97	1.02	0.96	0.95	0.98±0.03
MET	205-280	2.0-7.0	1.55×10^{-5}	-	4.23	4.29	4.31	4.31	4.28±0.04

2. Experimental

2.1. Instrumentation

A UV-Visible spectrophotometer (Perkin-Elmer, model lambda EZ 210) was used, controlled by a PC while applying the program PESSW v1.2, as well as a pH meter attached to a pH combination electrode (Conductronic). Data treatment was carried out with the software packages by Pirouette v3.11 by Infometrix Inc., OriginPro 8 SR0 v8.0724 by OriginLab Corporate and SQUAD (compiled by one of the authors of the present work (AR-H) and co-workers).

2.2. Reagents

All reagents were at least analytical grade. The herbicides isomethiozin (ISO) and methoprotryne (MET) were pesticides for analysis, from Riedel de Haen. The acid base indicators of thymol blue (ThB), bromocresol green (BrG), methyl orange (MeO), methyl red (MeR), and phenol red (PhR) came from Sigma Aldrich. The developed formulae of these substances are presented in Figure 3. In all tests, ultrapure water was used (Barnstead).

Standard solutions were prepared from the substances of interest in methanol (ISO and MET at 1×10^{-4} M, the acid-base indicators at 1×10^{-3} M), which were stored at 4 °C and protected from light. Working solutions were prepared daily by appropriate dilution; their concentrations are mentioned in Table 1.

2.3. Procedure

For each of the substances of interest, 100 mL of aqueous solution in sodium chloride 0.5 M were prepared, with the proper concentration to obtain absorbance values lower than 1. NaCl was used for setting the ionic strength of the medium An acid-base titration was conducted in a thermostated cell at 25 °C with constant stirring, provided with a thermometer and a pH combination electrode. Aqueous solutions of hydrochloric acid or sodium hydroxide at different concentrations were added, in order to obtain increases in pH of 0.2 units, approximately. Absorption spectra were recorded at each pH condition against a reagent blank (NaCl 0.5 M), with a resolution of 0.5 nm.

2.4. Data processing

In all cases, the absorption spectra were restricted to one band per chemical form (i.e., a spectral profile with one absorption maximum for the A- and HA species, respectively).

For PLS and PCR analyses, mean centering was used as pretreatment of data. Calibration models were built with three PCs, selecting the optimal number by the F test (two-tailed, p = 0.05). Then, α values were predicted for the rest of the series. The estimated pK_a values through MCR-ALS, PLS and PCR were compared with those obtained through SQUAD.

3. Results and discussion

Next, the case of PhR is discussed, although it should be noted that the same procedure was followed for the rest of the compounds. The absorption spectra of PhR at different pH conditions are presented in Figure 4, where an isosbestic point is observed at 481 nm as a consequence of an acid-base equilibrium. Absorption maxima can be observed at 431 and 560 nm, which correspond to the predominant presence of the acidic and basic forms, respectively. Also, it was noticed that at pH < 6, practically the whole indicator is in the form of *HA*, while at pH > 10 it can be considered that the predominant form is *A*. Therefore, the spectra profiles remained virtually constant.
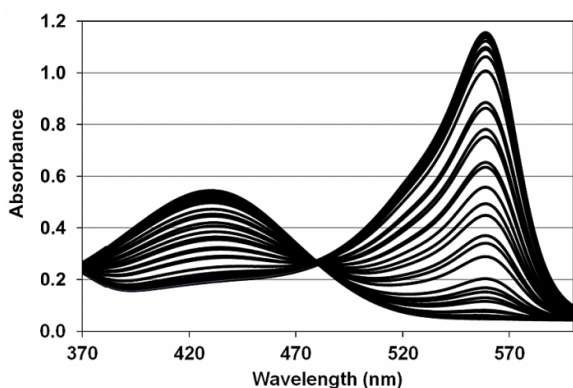

Figure 4. Absorption spectra of PhR as function of pH, between 4.7 and 10.8.

3.1. Data processing with SQUAD

Twenty four spectra with wavelength increments of 8 nm – taken from those shown in Figure 4 corresponding to aqueous solutions of phenol red with C_T = 1.02 × 10^{-5} M in the 5.72 to 9.28 pH range, were introduced to program SQUAD with a chemical model constituted by one acid-base equilibrium. The pKa = 7.3923 ± 0.0015 was refined by the program with a standard deviation for the absorbance data (s_A) such as s_A = 0.0035. The program SQUAD also allows determining the molar absorptivity coefficients for each species (Figure 5, up), with an uncertainty lower than 5% in this case.

3.2. Data processing with MCR-ALS

As was expected, two pure components were identified through MCR, modeling with one PC. In Figure 5 (down), the spectral profiles of the two pure components can be observed, very similar to the profiles of absorptivity coefficients obtained through SQUAD (Figure 5, up).

It should be noted the great similitude in shape for the corresponding absorptivity coefficients and source profiles of MCR algorithm.

Later, MCR-ALS was developed considering two pure components as maximum sources; constraints were defined as follows: a) non-negativity in amounts and profiles, b) unimodality in amounts and profiles, and c) closure in amounts. Initial estimates were chosen from rows. In Figure 6, the concentration profiles of the two pure components can be observed, which in this case coincide with the α fraction of the chemical species involved in the acid-base equilibrium.

According to the distribution diagram, α = 0.5 at pH = 7.4, which corresponds to the pK_a.
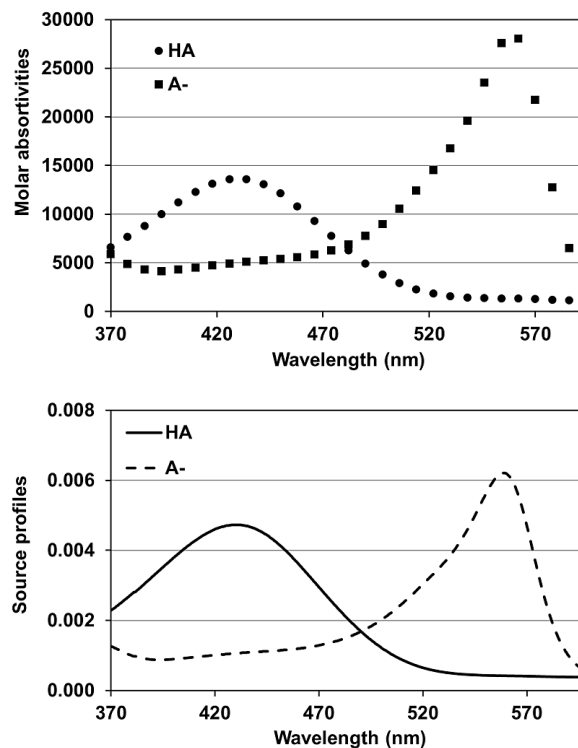

Figure 5. Fundamental spectroscopic data obtained for PhR in this work: (up) absorptivity coefficients for the species obtained by SQUAD, (down) source profiles obtained for the acidic and basic forms of PhR through MCR.
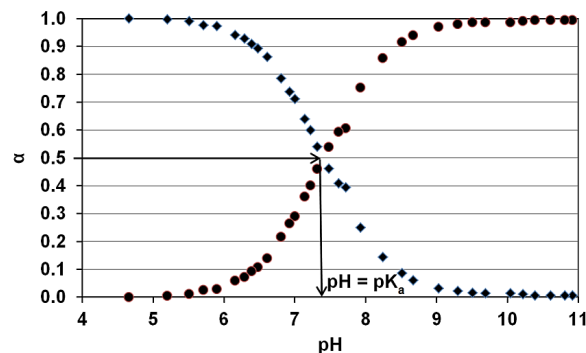

Figure 6. Distribution diagram for the fractions of HA (squares) and A- (circles), from which the pK_a can be deduced.

3.3. Data processing with PCR and PLS

After that, PCR and PLS were developed from the same data matrix as MCR-ALS (**X**). For the calibration step, three absorption spectra were considered in each case (for α_{HA} = 1, α_{A^-} = 0; and for α_{HA} = 0, α_{A^-} = 1) and the optimal number of factors was obtained by a cross validation leaving out one sample at a time. Through the F-test [18], it was observed that only one factor was needed for modeling (Figure 7), since the variance described by PCs 2 and 3 was negligible for both chemical forms. Then, for the rest of the series of samples α values were predicted. Outliers were not identified. Distribution diagrams obtained by PLS and PCR were very similar to those for MCR-ALS (see Figure 7).

Finally, the results obtained for all chemical compounds are shown in Table 1. As can be observed, there is clear agreement between the pK$_a$ reported in the literature and estimated values in all cases. Finally, the pK$_a$ values for ISO and MET are reported, since no information was found in the sources consulted.
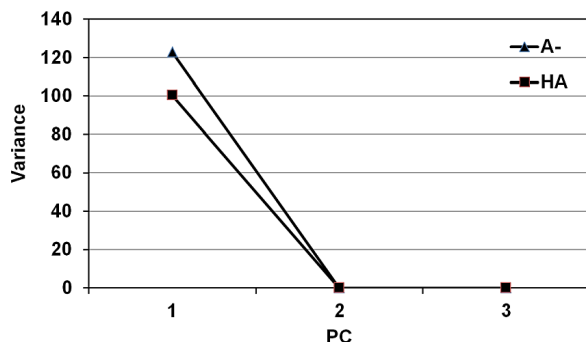

Figure 7. Selection of the optimum number of PC according to modeled variance. Factors 2 and 3 are negligible.

4. Conclusion

As shown by the results obtained through MCR-ALS and SQUAD, the PCR and PLS techniques were very useful for the estimation of the acid dissociation constants from spectrophotometric data for well-known acid-base indicators. Through these tools, these parameters were calculated for ISO (pK$_a$ = 0.98 ± 0.03) and MET (pK$_a$ = 4.28 ± 0.04).

Acknowledgements

References

[1]. Narasimham, L.; Barhate, V. D. *Eur. J. Chem.* **2011**, *2*, 36-46.
[2]. Nuryanti S.; Matsjeh S.; Anwar, C.; Raharjo, T. J.; Hamzah, B. *Eur. J. Chem.* **2013**, *4*, 20-24.
[3]. Kręzel, W.; Szczepanik, W.; Swiątek, M.; Jezowska-Bojczuk, M. *Bioorg. Med. Chem.* **2004**, *12*, 4075-4080.
[4]. De Oliveira, H. P. *Microchem. J.* **2008**, *88*, 32-37.
[5]. Ehala, S.; Grishina, A. A.; Sheshenev, A. E.; Lyapkalo, I. M.; Kasicka, V. *J. Chromatogr. A* **2010**, *1217*, 8048-8053.
[6]. Shi, B.; Qi, D. *J. Chromatogr. A* **2012**, *1231*, 73-76.
[7]. Argemi, A.; Saurina, J. *J. Pharmaceut. Biomed.* **2007**, *44*, 859-866.
[8]. Li, W.; Xian-Fa, S.; Zhong-Liang, Z. *Spectrochim. Acta A* **2007**, *67*, 789-792.
[9]. Alizadeh, K., Ghiasvand, A. R.; Borzoei, M.; Zohrevand, S.; Rezaei, B.; Hashemi, P.; Shamsipur, M.; Maddah, B.; Morsali, A.; Akhbari, K.; Yavari, I. *J. Mol. Liq.* **2009**, *149*, 60-65.
[10]. Batistela, V. G.; Silva Pellosi, D.; Dutra de Souza, F.; Ferreira da Costa, W.; De Oliveira Santin, S. M.; De Souza, V. R.; Caetano, W.; De Oliveira, H. P. M.; Spacino Scarminio, I.; Hioka, N. *Spectrochim. Acta A* **2011**, *79*, 889-897.
[11]. Garrido, M.; Rius, F. X.; Larrechi, M. S. *Anal. Bioanal. Chem.* **2008**, *390*, 2059-2066.
[12]. Gargallo, R.; Eritja, R.; Kudrev, A. G. *Chemometr. Intell. Lab.* **2013**, *122*, 78-83.
[13]. Leggett, D. J.; McBryde, W. E. A. *Anal. Chem.* **1975**, *47*, 1065-1070.
[14]. Leggett, D. J. (ed.). Computational Methods for the Determination of Formation Constants. Chapter 6. Plenum Press, New York, 1985.
[15]. Meloun, M.; Ferenikova, Z.; Malkova, H.; Pekarek, T. *Cent. Eur. J. Chem.* **2012**, *10*, 338-353.
[16]. Gomez-Zaleta, B.; Ramirez-Silva, M. T.; Gutierrez, A.; Gonzalez-Vergara, E.; Güizado-Rodríguez, M.; Rojas-Hernandez, A. *Spectrochim. Acta A* **2006**, *64*, 1002-1009.
[17]. Gemperline, P. (ed.). Practical Guide to Chemometrics. CRC Press, 2006.
[18]. Jollife, I. T. Principal Component Analysis, second edition, Springer-Verlag, 2002.
[19]. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, 2002.
[20]. Haaland, D. M.; Thomas, E. V. *Anal. Chem.* **1988**, *60*, 1193-1202.
[21]. Sabnis, R. W. Handbook of Acid-Base Indicators, CRC Press, 2008.
[22]. Jeffery, G. H.; Bassett, J.; Mendham, J.; Denney, R. C. Vogel's Textbook of Quantitative Chemical Analysis, fifth edition, Longman Scientific & Technical, 1989.