Simultaneous Determination of Two Antibiotics in Tablets by Spectrophotometry and Principal Component Regression (PCR) Analysis. An Advanced Undergraduate Experiment Involving Chemometrics

María É. Ribone,[†] Ariana P. Pagani,[†] Héctor C. Goicoechea,[†,‡] and Alejandro C. Olivieri[*,†]

Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario (2000), Argentina, and Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Paraje El Pozo, Santa Fe(3000), Argentina, aolivier@fbioyf.unr.edu.ar

Abstract: An advanced analytical chemistry laboratory experiment involving the simultaneous determination of the antibiotics sulfamethoxazole and trimethoprime in commercial tablets is described. It involves the following steps: (1) preparation of nine calibration mixtures and the recording of their absorption spectra in the region 200–320 nm, (2) dissolution of the tablet contents and the recording of spectrophotometrics data, and (3) processing the data with the multivariate calibration technique of principal component regression (PCR). The theory of PCR is discussed, and a Visual Basic program is made available for data processing. The latter program allows students to obtain and save relevant statistical information (root-mean-square deviation, correlation coefficient, and relative error of prediction), as well as calibrate spectral factors and spectral residuals for each test sample. This program helps to illustrate the PCR technique in detail. The reagents used are of low cost and nontoxic; the experiment is simple and provides students with insight into a the real practice of integrating chemistry, instrumentation and computer techniques.

Introduction

Chemometrics involves the use of multivariate calibration methods applied to spectroscopic data [1–4]. These methods are classified according to the type of spectral information used and as to whether the calibration process is direct or inverse (Table 1). The common task of all multivariate methods is to efficiently extract information concerning certain analytes of interest from spectra of multicomponent mixtures. Perhaps the simplest of these methods is classical least squares (CLS). CLS is an extension of the well-known method for resolving mixtures of l components by measuring the absorbance at l wavelengths, which students learn in basic analytical chemistry courses [5–7]. In CLS, measurements are performed at n wavelengths (in practice $n >> l$), and, thus, it is considered to be a full-spectrum method. This generally leads to higher precision as compared to using only a small number of wavelengths [3]; however, it also has some drawbacks. First, it uses a *direct* calibration step, which requires the knowledge of all sample components. The term *direct* refers to the usual definition of Beer's law ($A = kc$) extended to multicomponent mixtures. Second, it is very sensitive to spectral noise, baseline drift, and spectral overlap of the sample components [3]. Inverse least squares (ILS) circumvents these problems. Because the inverse Beer's law, $c = k'A$, (extended to multicomponent mixtures) is used, it only requires the knowledge of the concentration of the analyte of interest for calibration [3]; however, it is restricted to a small number of wavelengths. Better performance is obtained with methods which use spectral factors, such as principal component regression (PCR), partial least squares (PLS), and hybrid linear analysis (HLA) [4]. These methods use inverse calibration steps, combined with a prior optimization of the information contained in the calibration spectra. They display the following advantages: (1) they use full spectra, (2) they require knowledge of only the concentrations of the analytes of interest, and (3) the spectra can be decomposed into factors, avoiding the problems associated with overlapping, collinearities, noise, drifts, and other spectral artifacts [3]. They are ideally suited for the study of complex biological samples, such as drug or metabolite monitoring in blood [8] or pharmaceutical analysis of multicomponent preparations where the excipients may not be known [9]. A common requirement to all these multivariate methods is that the matrix should be modeled during calibration; that is, all compounds in which the chemist is not interested should be present in the calibration samples (although one need not know their concentrations). A summary of the advantages and drawbacks of these methods is shown in Table 1.

The simplest factor-based multivariate calibration method is PCR. We have, therefore, set out a simple system in which PCR can be successfully applied: the resolution of a mixture of antibiotics in pharmaceutical tablets. Specifically, the studied antibiotics are sulfamethoxazole, *N*′-(5-methyl-3-isoxazolyl)-sulfanilamide (SMZ) and trimethoprim, 2,4-diamino-5-(3,4,5-trimethoxybenzyl)-pyrimidine (TMP) [10]. Sulfonamides such as SMZ show a wide range of antimicrobial activity against both gram-positive and gram-negative bacteria, and they have found increasing usage owing to the introduction, in the mid-1970s, of their combination with TMP. The latter shows an antibacterial spectrum similar to SMZ, but the mixture of SMZ and TMP (known as cotrimoxazole) combines both drugs in a synergistic action, inhibiting the production of tetrahydrofolic

* Address correspondence to this author.
† Universidad Nacional de Rosario
‡ Universidad Nacional del Litoral

Table 1. Summary of Some Popular Multivariate Calibration Methods

Method	Advantages	Drawbacks
CLS Classical Least Squares	Uses full spectra.	All components should be known. Sensitive to collinearities, baseline drifts, noise, etc.
ILS Inverse *L*east *S*quares	Only the components of interest need to be known.	Uses a small number of wavelengths Sensitive to collinearities.
PCR *P*rincipal *C*omponents *R*egression	Only the components of interest need to be known. Uses spectral factors (less sensitive to collinearities). Uses full spectra.	The matrix should be modeled in the calibration.
PLS Partial Least Squares	Uses spectral and concentration factors (less sensitive to collinearities). Uses full spectra.	The matrix should be modeled in the calibration
HLA Hybrid Linear Analysis	Uses spectral factors (less sensitive to collinearities). Uses full spectra. Detects interferences.	The matrix should be modeled in the calibration

Table 2. PCR Calibration Data for Binary SMZ/TMP Mixtures Using the First Two Factors

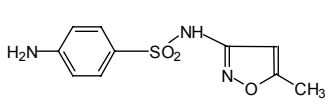
Parameter[a]	SMZ		TMP	
Spectral region (nm)	230–300		230–300	
Concentrations (mg L^{-1})	Actual	Predicted	Actual	Predicted
cal1	10.00	10.08	2.0	2.08
cal2	10.00	10.03	5.0	4.99
cal3	10.00	10.02	8.0	8.11
cal4	17.00	16.83	2.0	1.90
cal5	17.00	16.95	5.0	4.93
cal6	17.00	17.03	8.0	7.93
cal7	24.00	24.14	2.0	2.11
cal8	24.00	24.05	5.0	5.01
cal9	24.00	23.93	8.0	8.04
RMSD (mg L^{-1})	0.087		0.076	
r^2	0.9998		0.9990	
REP(%)	0.51		1.53	

[a] $RMSD = \left[\frac{1}{m}\sum_{1}^{m}(c_{act} - c_{pred})^2\right]^{1/2}$, $r^2 = 1 - \frac{\sum_{1}^{m}(c_{act} - c_{pred})^2}{\sum_{1}^{m}(c_{act} - \bar{c})^2}$,

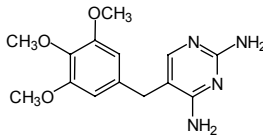
$REP(\%) = \frac{100}{\bar{c}}\left[\frac{1}{m}\sum_{1}^{m}(c_{act} - c_{pred})^2\right]^{1/2}$, $\bar{c}$ is the average component concentration in the m calibration mixtures.

acid by acting on two sequential steps of the bacterial biosynthesis [11]. Cotrimoxazole is indicated for several infections, including urinary, respiratory, gastrointestinal, and genital infections.

The experiment described here can be conveniently performed by students in their last year of Chemistry or Pharmacy in a course devoted to advanced aspects of Analytical Chemistry. The interest in this type of laboratory experience arises because it offers the opportunity of



SMZ



TMP

integrating a laboratory practice (involving weighing of reagents, preparation of solutions of different concentrations, processing real samples and reading their absorption spectra) with computer techniques (transforming spectra into ASCII format, preparing calibration data files, and using PCR software to analyze samples). Mathematically speaking, PCR is based on the technique of spectral decomposition, which involves obtaining eigenvectors and eigenvalues of a square matrix. Students have acquired this skill in previous mathematics courses; thus, we felt that our Advanced Analytical Chemistry course was an excellent opportunity to illustrate the application of eigenvectors to a real chemometrics experiment. Related multivariate experiments involving CLS [12] and target factor analysis [13] have been previously described in the literature. We further stress that both the reagents and pharmaceutical samples involved are easily accessible, nontoxic, and of low cost. Finally, we have developed simple Visual Basic 5.0 software that can be used by students with the Windows 95 or 98 operating systems. It is available to readers as supporting material.

Experimental Procedure

Calibration Samples. A training set of 9 samples (cal1–cal9) were prepared for calibration, using a three-level full-factorial design (Table 2). The meaning of full factorial is evident on examining the concentrations of both components in Table 2.

Synthetic and Commercial Samples. Two synthetic unknowns and two commercial samples were studied. Their concentrations are shown in Table 3.

Software. Although software for PCR analysis (and other multivariate techniques) is commercially available, we chose to write an in-house program that may be easily implemented by students. Thus, the PCR.exe program was written in Visual Basic 5.0 [14].

Details regarding the experiment and software are located in the supporting material.

Theory

PCR Analysis. PCR is best understood using matrix notation. With the spectra of m calibration samples measured at n digitized wavelengths, an $m \times n$ data matrix **A** is constructed. If this matrix is used without further manipulation, collinearity problems may arise. Thus, in PCR the first step is to decompose **A** into the product of two smaller matrixes

$$\mathbf{A} = \mathbf{T} \times \mathbf{B} \quad (1)$$

Table 3. Results Obtained by Applying PCR Analysis to Some Binary Mixtures

Sample	Component	Synthetic samples		Recovery (%)
		Actual	Found	
		mg L^{-1}		
Synthetic sample 1	SMZ	18.0	17.8	98.9
	TMP	3.6	3.6	100.0
Synthetic sample 2	SMZ	28.0	27.8	99.3
	TMP	5.6	5.7	101.8

			Commercial samples			
Tablet	Average tablet weight/mg	Sample weight/mg	Component	Actual mg	Found mg	Recovery (%)
Bactrim[a]	504.8	118.4	SMZ	400	399.1	99.8
			TMP	80	75.0	93.8
Cotrimoxazol[a]	658.3	146.4	SMZ	400	406.5	101.6
			TMP	80	74.8	93.5

[a] Actual contents were taken as those reported by the manufacturing laboratories. The content w of SMZ or TMP were calculated from $w = 4 \times c \times$ (tablet weight)/(sample weight), where c is the calculated concentration of each component (see the Experimental section for the preparation of commercial samples).

where **B** is an $h \times n$ matrix called "loading" and **T** is an $m \times h$ matrix named "scores." The rows of matrix **B** are the first significant h eigenvectors of the square matrix ($\mathbf{A}^T\mathbf{A}$). The latter are called factors. (See below for a discussion on how to select the appropriate number of factors.) The relevant spectral factors are extracted from **A** by efficient computer techniques in the order of their contribution to the variance; that is, they are selected so that their information content is maximized. In other words, the eigenvectors of ($\mathbf{A}^T\mathbf{A}$) are ranked according to the associated eigenvalues, selecting those corresponding to the higher eigenvalues. There are computer algorithms, however, that do not need to obtain all the eigenvectors; they simply extract them one by one in the order of their eigenvalues.

Once the factors have been obtained, the matrix **T** is easily calculated from eq 1 as

$$\mathbf{T} = \mathbf{A} \times \mathbf{B}^T \tag{2}$$

since $\mathbf{B}^T \times \mathbf{B} = \mathbf{I}$ (**I** is the unit $n \times n$ matrix and $\mathbf{B}^T$ is the transpose of **B**). In subsequent mathematical manipulations, **A** is replaced by **T**, greatly alleviating collinearity problems, because the orthogonal eigenvectors of **B** lead, according to eq 2, to orthogonal columns in **T**.

During calibration, the m concentrations of each of the l components of interest (which form an $m \times l$ matrix **C**) are related to the scores matrix as

$$\mathbf{C} = \mathbf{T} \times \mathbf{V} \tag{3}$$

where **V** is an $h \times l$ matrix of regression coefficients. Equation 3 is similar to the regression step in inverse calibration methods where the concentration is related to absorbance (the inverse Beer´s law). In PCR, however, absorbances are replaced by scores; thus, the matrix **V** may be seen as the proportionality constant relating concentration to scores. From eq 3

$$\mathbf{V} = (\mathbf{T}^T \times \mathbf{T})^{-1} \times \mathbf{T}^T \times \mathbf{C} \tag{4}$$

where $(\mathbf{T}^T \times \mathbf{T})^{-1}$ is easily obtained because the columns of **T** are orthogonal to each other.

Once **V** is known, unknown samples can be studied (this is called the prediction step). During prediction, the components score is obtained from the unknown $n \times 1$ spectrum **a** as

$$\mathbf{t} = \mathbf{B} \times \mathbf{a} \tag{5}$$

where **t** is an $h \times 1$ vector. The unknown concentrations are then obtained as the $l \times 1$ vector $\mathbf{c}_{pred}$;

$$\mathbf{c}_{pred} = \mathbf{V}^T \times \mathbf{t} \tag{6}$$

therefore, each element of the column vector, $\mathbf{c}_{pred}$, is the concentration of a given sample component.

Finally, the $n \times 1$ vector **e** of spectral residues for a given sample are obtained by subtracting the calculated spectrum $\mathbf{B}_i^T \times \mathbf{t}_i = \mathbf{B}_i^T \times \mathbf{B}_i \times \mathbf{a}$ from that for the sample:

$$\mathbf{e} = \mathbf{a} - \mathbf{B}_i^T \times \mathbf{B} \times \mathbf{a} \tag{7}$$

Cross-validation Procedure. The cross-validation procedure consists of removing one sample at a time from the calibration set and using the remaining samples (eight in the present case) to calibrate and predict the concentration of the analyte in the sample removed. The process is repeated until all samples have been left out once. The actual and predicted concentrations of the analyte in the removed samples are used to define the PRESS (Predicted Error Sum of Squares). The later is defined by:

$$\text{PRESS} = \sum_{1}^{m} (c_{act} - c_{pred})^2 \tag{8}$$

The values of the PRESS are calculated as a function of the number of factors. Ideally, a certain value of h should lead to a minimum PRESS, and increasing h will only add uninformative data causing the PRESS to increase. In practice, h is selected by computing the following F ratio:

Table 4. Changes in PRESS as a Function of the Number of Factors for each Analyte and Statistical Analysis of the Optimum Number of Factors

SMZ				TMP			
Factor number	PRESS mg^2 L^{-2}	F[a]	P(%)[b]	Factor number	PRESS mg^2 L^{-2}	F[a]	P(%)[b]
0	372.09	5523	100	0	68.34	2269	100
1	27.39	406	100	1	84.09	2792	100
2	0.067	1	50	2	0.052	1.74	79
3	0.076			3	0.036	1.19	60
4	0.110			4	0.030	1	50

[a] F = PRESS(h)/PRESS(h*), see text.

[b] P = probability that $F > 1$

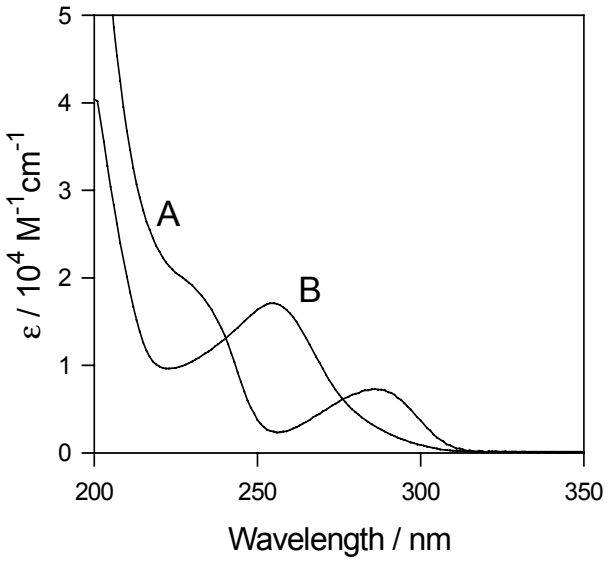


Figure 1. Electronic absorption spectra in aqueous solution (NaOH, 0.04 N) of: (A) SMZ and (B) TMP.

$$F(h) = \frac{\text{PRESS}(h)}{\text{PRESS}(h^*)} \tag{9}$$

where h* is the number of factors leading to the minimum PRESS and $h < h$*. The optimum number of factors is suggested to correspond to a probability of less than 80% that $F > 1$.

Students are asked to try other wavelength regions and to compare the values of PRESS that are obtained with those shown in Table 4. With the results saved in the PRESS.txt file, students are able to examine how the cross-validation predicted concentrations of SMZ and TMP in the calibration samples approach the actual values in going from h = 1 to h = 2.

The program not only saves the values of PRESS for different factors, but other statistical parameters as well. These are the RMSD (Root-Mean-Square Deviation); the correlation coefficient, r^2; and the REP (Relative Error of Prediction), defined by:

$$\text{RMSD} = \left[\frac{1}{m}\sum_1^m (c_{\text{act}} - c_{\text{pred}})^2\right]^{1/2} \tag{10}$$

$$r^2 = 1 - \frac{\sum_1^m (c_{\text{act}} - c_{\text{pred}})^2}{\sum_1^m (c_{\text{act}} - \bar{c})^2} \tag{11}$$

$$\text{REP}(\%) = \frac{100}{\bar{c}}\left[\frac{1}{m}\sum_1^m (c_{\text{act}} - c_{\text{pred}})^2\right]^{1/2} \tag{12}$$

where $\bar{c}$ is the average component concentration in the m calibration mixtures.

Spectral residues. For a particular sample, the program also saves the spectral residues in the file RESIDUE.txt (the filename can be changed, see Figure 4). A report box (Figure 4) shows the sample name, predicted concentration, number of factors used and spectral residue. The latter is defined by:

$$\text{Spectral Residue (\%)} = \frac{\sum_{i=1}^n |e_i|}{\sum_{i=1}^n |a_i|} \times 100 \tag{12}$$

where e_i and a_i are the components of the vectors **e** and **a** respectively (see above). The value of the spectral residue helps to characterize the performance of the model with the particular sample under study. The result is considered satisfactory if the value is less than about 2%. For a specific sample, the experimental and calculated spectra are shown in Figure 5, along with the spectral residues.

Results and Discussion

Figure 1 shows the absorption spectra of pure SMZ and TMP. As can be seen, both spectra are overlapped in the informative 200 to 320 nm spectral region. This is one of the reasons for applying techniques such as PCR for studying mixtures of SMZ and TMP. In the present case, a calibration set of samples was prepared using a full-factorial approach (Table 2). The first step in multivariate calibration analysis is the selection of the best wavelength working range. One usually selects regions where the mixture components show significant absorptions. High absorbance regions should be avoided since they may deviate from Beer's law. Examination of Figure 1 suggests that a convenient working range is the region 230 to 300 nm.

With the spectra shown in Figure 2, the program PCR.exe allows students to study the corresponding factors of the matrix **A**. The first four factors in the region 230 to 300 nm are

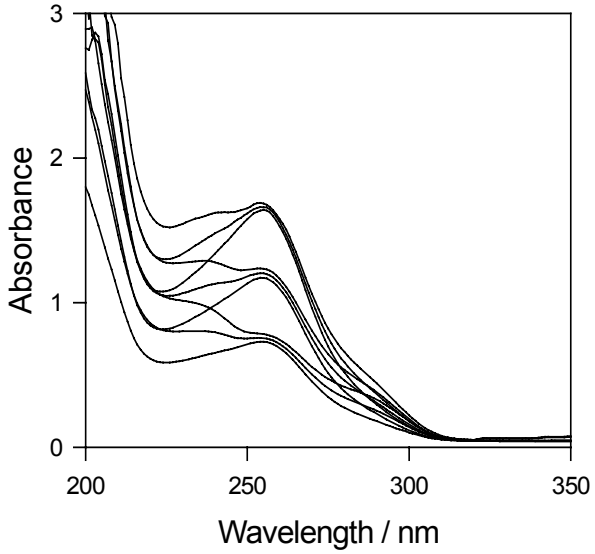


Figure 2. Electronic absorption spectra of the nine calibration samples.

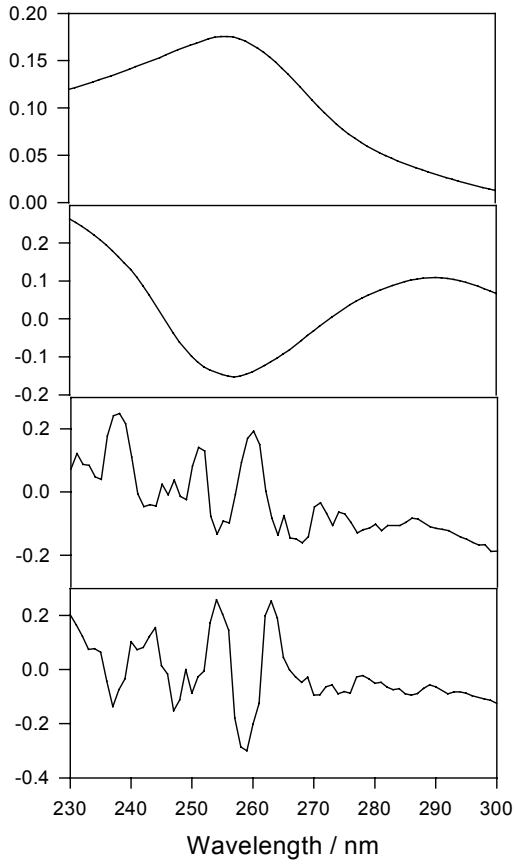


Figure 3. First four spectral factors (from top to bottom), as obtained by PCR.exe from the spectra shown in Figure 3.

shown in Figure 3. A significant difference is apparent between the first two factors on one hand and the last two on the other. These latter factors appear to contain only noise in contrast to the valuable information provided by the first two factors (notice that the first factor shown in Figure 3 resembles the spectrum of pure SMZ, the major component of these binary mixtures). This line of reasoning leads to the intuitivechoice of $h = 2$ as the appropriate number of factors. A less intuitive approach, the cross-validation procedure, has been proposed in the literature [3]. Application of the later criterion leads to $h = 2$ for both compounds, in good agreement with the intuitive approach discussed above. Also, notice that there is good agreement between the actual concentrations shown in Table 2, as well as the excellent statistical indicators.

Finally, unknown samples were studied, both synthetic, mixtures of SMZ and TMP prepared from the stock solutions of the pure compounds, and real, some commercial tablets from different laboratories. Using two factors, calibration and prediction of both concentrations in the unknown samples was performed with PCR.exe. The results are shown in Table 3. The recoveries for the synthetic samples are excellent. For the commercial samples, one should notice that the recoveries are calculated on the basis of the content reported by the manufacturing laboratories (Table 3); however, there exist recommended limits for quality control of pharmaceuticals in books such as the Pharmacopoeias. The US XXII Pharmacopeia recommends that SMZ/TMP tablets should contain not less than 90% and not more than 110% of the content declared by the laboratory [15]. The British Pharmacopoeia, on the other hand, recommends a range from 92.5% to 107.5% [16]. On this basis, all commercial samples analyzed by students yielded satisfactory results.

Conclusions

An experiment involving the interplay of pharmaceutical analytical chemistry and instrumentation and computer techniques has been described. It involves the simultaneous determination of two antibiotics in commercial tablets: sulfamethoxazole and trimethoprime. The analysis is performed by recording electronic absorption data and processing them with a multivariate calibration method: principal component regression (PCR), whose theory is discussed in detail. Further, a Visual Basic program is available as supporting material. It allows students to perform data processing, obtain and save statistical information, and produce several concentration and spectral plots which are relevant to PCR.

Acknowledgment. Financial support from CONICET (Consejo Nacional de Investigaciones Cientificas y Tecnicas), Universidad Nacional de Rosario and Agencia Nacional de Promocion Cientifica y Tecnologica is gratefully acknowledged. H.C.G. thanks FOMEC (Programa para el mejoramiento de la calidad de la enseñanza universitaria) for a fellowship.

References and Notes

1. Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: New York, 1989.
2. Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, *185*, 1-19.
3. Haaland, D. M.; Thomas, E. V. *Anal. Chem.* **1988**, *60*, 1193-1202.
4. Berger, A. J.; Koo, T.-W.; Itzkan, I.; Feld, M. S. *Anal. Chem.* **1998**, *70*, 623-627.

5. Sawyer, D. T.; Heineman, W. R.; Beebe, J. M. *Chemistry Experiments for Instrumental Methods*; Wiley: New York, 1984, p. 163.
6. Skoog, D. A. *Principles of Instrumental Analysis*, 3rd. Ed.; Saunders College Publishing: Philadelphia, 1985, p. 212.
7. Willard, H. H.; Merritt Jr., L. L.; Dean, J. A.; Settle Jr., F. A. *Instrumental Methods of Analysis*; Wadsworth: 1988, p. 170.
8. Goicoechea, H. C.; Olivieri, A. C. *Anal. Chim. Acta* **1999**, *384*, 95-103.
9. Goicoechea, H. C.; Olivieri, A. C. *Talanta* **1998**, *47*, 103-108.
10. Ribone, M. E.; Pagani, A. P.; Olivieri, A. C. *Anal. Lett.* **1999**, 32, 1389-1401
11. Goodman-Hillman, A.; Rall, T.; Nier, A.; Taylor, P. *The Farmacologycal Basis of Therapeutics*; McGraw-Hill: New York, 1996.
12. Blanco, M.; Iturriaga, H.; Maspoch, S.; Tarin, P. *J. Chem. Educ.* **1989**, *66*, 178-180.
13. Charles, M. J.; Martin, N. W.; Msimanga, H. Z. *J. Chem. Educ.* **1997**, *74*, 1114-1117.
14. *Visual Basic 5.0 Programmer's Guide*, Microsoft Press: Redmond, Washington, 1997.
15. *United States Pharmacopeia XXII*; United States Pharmacopeial Convention: Rockville, M.D., 1990, p. 1096.
16. *British Pharmacopoeia*, H. M. Stationery Office: London, 1998, Vol. II, p. 262.