Resolution of Binary Mixtures of Rifamycin SV and Rifampicin by UV/VIS Spectroscopy and Partial Least-Squares Method (PLS)

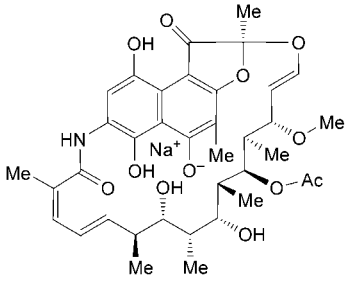
by **M. Asunción Alonso Lomillo**, **Olga Domínguez Renedo**, and **M. Julia Arcos Martínez***

Dpto. de Química, Área de Química Analítica, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos, s/n, E-09001 Burgos
(Phone: +34947258818; fax: +34947258831; e-mail: asunal@beca.ubu.es, olgado@ubu.es, jarcos@ubu.es)

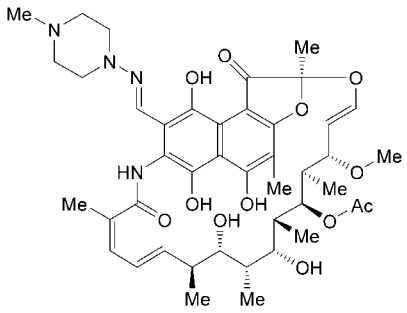
A procedure is proposed for the joint determination of rifamycin SV and rifampicin by UV/VIS spectroscopy. A partial least-squares regression (PLS) was used for the resolution of the overlapping spectrophotometric signals from mixtures of the two drugs. The application of a genetic algorithm to select some of the predictor variables allows one to considerably reduce the number of experimental variables, as well as to improve the prediction capacity of the PLS model constructed with these selected potentials. Finally, the methods were applied to the determination of these drugs in biological samples.

1. Introduction. – Rifamycins are antibiotics belonging to the group of naphthalenic ansamycins, which exert their activity against a large variety of organisms, such as bacteria, eukaryotes, and viruses [1], by specific inhibition of bacterial DNA-dependent RNA polymerase.

Rifamycins A, B, C, D, and E were isolated from the fermentation broth of the bacteria *Streptomyces mediterranei* [2]. Rifamycin B is the most active of these antibiotics. Rifamycin SV is obtained by oxidation, hydrolysis, and reduction of rifamycin B. Therefore, it has a semisynthetic origin, and, at the same time, it is more powerful than rifamycin B. When isolated from rifamycin SV, rifampicin differs only in a group attached to the naphthohydroquinone ring. The latter has many characteristics that make it the ideal product of the group.



Rifamycin SV



Rifampicin

Traditionally, rifamycin SV has been used as a chiral selector in capillary electrophoresis to enantioselectively resolve a number of chiral compounds [3]. It has been determined by adsorptive stripping voltammetry [4]. Several methods for the determination of rifampicin including HPLC [5–7], thin-layer chromatography (TLC) [8], spectrophotometry [9][10], and voltammetry [11] were reported.

To our knowledge, there have been no reports on the study of these molecules in mixtures. This sort of analysis would be useful to determine, for example, the purity grade in a rifampicin sample obtained from rifamycin SV. In this way, we propose a procedure for simultaneous determination of rifamycin SV and rifampicin in binary mixtures by UV/VIS spectroscopy. Such compounds absorb strongly in both UV and VIS spectral regions (*Fig. 1*).

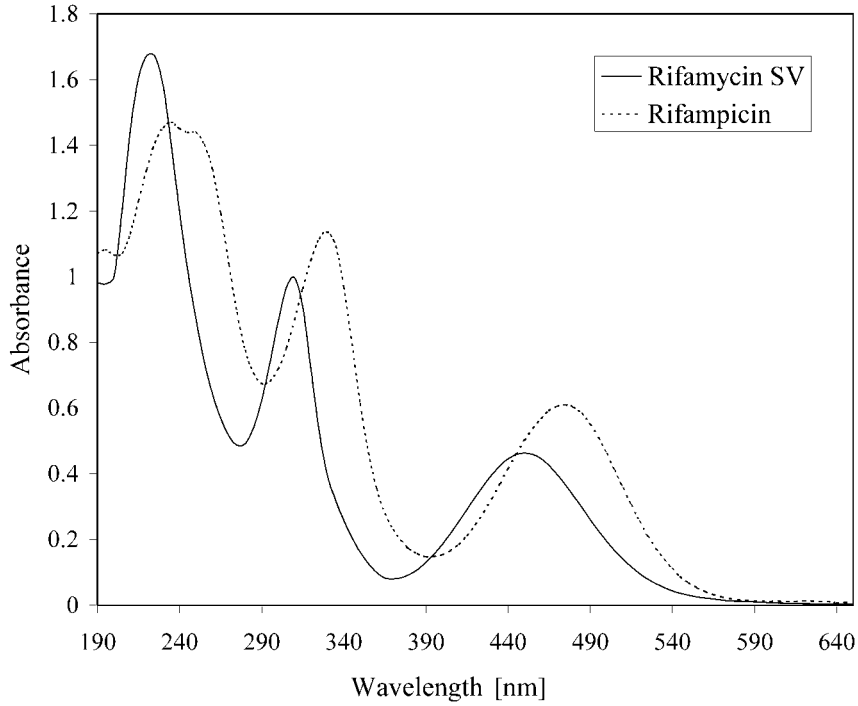


Fig. 1. *UV/VIS Spectra of* 5×10^{-5} M *solutions of rifamycin SV and rifampicin in* Britton Robinson, *pH 7*

Methods that involve a soft calibration, such as partial least-squares (PLS), have been demonstrated as being very useful in the resolution of overlapping signals [12][13]. These methods imply the consideration of a large number of original variables, and, in fact, most of them will only incorporate noise into the PLS model constructed. Therefore, their elimination may improve the prediction potential of the model. A modified steady-state genetic algorithm (GA) without duplicates [14] has been used to select the predictor variables.

2. Results and Discussion. – To perform the determination of rifamycin SV and rifampicin in the simplest way, the viability of the methodology of univariate regression was studied. Therefore, regressions based on the least-squares (OLS) criterion were carried out in order to obtain optimal precision and accuracy of both slope and intercept. This methodology requires an independent regression analysis for each drug.

The absorbance corresponding to the maximum points has been taken to perform the rifampicin calibrations in presence of different amounts of rifamycin SV (*Table 1*) as well as the rifamycin SV calibrations in presence of different amounts of rifampicin (*Table 2*). Not all of these regressions displayed a high coefficient of determination (R^2), reduced confidence intervals for the coefficients in the model, and low standard

Table 1. *Rifamycin SV Calibrations by OLS Regression in the Presence of Different Amounts of Rifampicin*

λ	$C_{\text{rifampicin}}$				
	10^{-5} mol dm^{-3}	3×10^{-5} mol dm^{-3}	5×10^{-5} mol dm^{-3}	7×10^{-5} mol dm^{-3}	9×10^{-5} mol dm^{-3}
230 nm	Intercept =	Intercept =	Intercept =	Intercept =	Intercept =
	0.44 ± 0.39	1.19 ± 0.51	1.74 ± 0.57	2.26 ± 0.48	2.61 ± 0.21
	Slope = 0.25 ± 0.07	Slope = 0.19 ± 0.09	Slope = 0.14 ± 0.10	Slope = 0.08 ± 0.08	Slope = 0.03 ± 0.04
	$R^2 = 0.9793$	$R^2 = 0.9402$	$R^2 = 0.8636$	$R^2 = 0.7371$	$R^2 = 0.6840$
	$S_{yx} = 0.1355$	$S_{yx} = 0.1751$	$S_{yx} = 0.1985$	$S_{yx} = 0.1667$	$S_{yx} = 0.0739$
330 nm	Intercept =	Intercept =	Intercept =	Intercept =	Intercept =
	0.20 ± 0.09	0.68 ± 0.05	1.08 ± 0.10	1.56 ± 0.23	1.90 ± 0.08
	Slope = 0.18 ± 0.02	Slope = 0.16 ± 0.01	Slope = 0.13 ± 0.02	Slope = 0.10 ± 0.04	Slope = 0.08 ± 0.01
	$R^2 = 0.9976$	$R^2 = 0.9979$	$R^2 = 0.9956$	$R^2 = 0.9564$	$R^2 = 0.9911$
	$S_{yx} = 0.0320$	$S_{yx} = 0.0260$	$S_{yx} = 0.0337$	$S_{yx} = 0.0812$	$S_{yx} = 0.0263$
470 nm	Intercept =	Intercept =	Intercept =	Intercept =	Intercept =
	0.11 ± 0.01	0.38 ± 0.02	0.59 ± 0.03	0.85 ± 0.14	1.07 ± 0.06
	Slope = 0.09 ± 0.01	Slope = 0.08 ± 0.01	Slope = 0.09 ± 0.01	Slope = 0.08 ± 0.24	Slope = 0.08 ± 0.02
	$R^2 = 0.9998$	$R^2 = 0.9978$	$R^2 = 0.9987$	$R^2 = 0.9733$	$R^2 = 0.9950$
	$S_{yx} = 0.0047$	$S_{yx} = 0.0142$	$S_{yx} = 0.0112$	$S_{yx} = 0.0478$	$S_{yx} = 0.0199$

Table 2. *Rifampicin Calibrations by OLS Regression in the Presence of Different Amounts of Rifamycin SV*

λ	$C_{\text{rifamycin SV}}$				
	10^{-5} mol dm^{-3}	3×10^{-5} mol dm^{-3}	5×10^{-5} mol dm^{-3}	7×10^{-5} mol dm^{-3}	9×10^{-5} mol dm^{-3}
230 nm	Intercept =	Intercept =	Intercept =	Intercept =	Intercept =
	0.42 ± 0.22	1.21 ± 0.44	1.87 ± 0.45	2.36 ± 0.33	2.64 ± 0.15
	Slope = 0.24 ± 0.04	Slope = 0.19 ± 0.08	Slope = 0.12 ± 0.08	Slope = 0.06 ± 0.06	Slope = 0.02 ± 0.03
	$R^2 = 0.9925$	$R^2 = 0.9526$	$R^2 = 0.8859$	$R^2 = 0.7879$	$R^2 = 0.7562$
	$S_{yx} = 0.0769$	$S_{yx} = 0.1541$	$S_{yx} = 0.1574$	$S_{yx} = 0.1160$	$S_{yx} = 0.0526$
330 nm	Intercept =	Intercept =	Intercept =	Intercept =	Intercept =
	0.19 ± 0.10	0.62 ± 0.10	1.02 ± 0.20	1.40 ± 0.20	1.77 ± 0.15
	Slope = 0.20 ± 0.02	Slope = 0.17 ± 0.02	Slope = 0.15 ± 0.03	Slope = 0.12 ± 0.03	Slope = 0.09 ± 0.05
	$R^2 = 0.9978$	$R^2 = 0.9967$	$R^2 = 0.9839$	$R^2 = 0.9769$	$R^2 = 0.9770$
	$S_{yx} = 0.034$	$S_{yx} = 0.0365$	$S_{yx} = 0.0704$	$S_{yx} = 0.0702$	$S_{yx} = 0.0521$
470 nm	Intercept =	Intercept =	Intercept =	Intercept =	Intercept =
	0.10 ± 0.04	0.28 ± 0.05	0.45 ± 0.09	0.64 ± 0.04	0.83 ± 0.05
	Slope = 0.11 ± 0.01	Slope = 0.46 ± 0.01	Slope = 0.11 ± 0.03	Slope = 0.11 ± 0.07	Slope = 0.10 ± 0.01
	$R^2 = 0.9988$	$R^2 = 0.9984$	$R^2 = 0.9946$	$R^2 = 0.9988$	$R^2 = 0.9979$
	$S_{yx} = 0.0140$	$S_{yx} = 0.0169$	$S_{yx} = 0.0307$	$S_{yx} = 0.0141$	$S_{yx} = 0.0169$

deviation of the residuals (S_{xy}). Therefore, the direct detection of these analytes at wavelengths of maximum absorbance are not suitable for resolving mixtures of these drugs without a separation step due to the interferences between them.

Bearing this result in mind, PLS models were constructed for the resolution of binary mixtures of the drugs mentioned above.

PLS is a widely used regression method. It is known that information from the concentration values is introduced into the calculation of the so-called latent variables, which are linear combinations of the original variables. To maintain the maximum prediction ability of the model, it is appropriate to optimize the sum of squares in prediction (PRESS) of the PLS models constructed with the calibration data [12][13].

$$\mathrm{PRESS}(k) = \sum_{i=1}^{m} \left(c_i - \hat{c}_{k/i}\right)^2$$

in which c_i is the concentration corresponding to the ith calibration sample (ith element of the vector $\boldsymbol{c}$) and $\hat{c}_{k/i}$ is the concentration estimated by the PLS model with k latent variables computed when the ith sample is removed. In practice, a more-stable estimate is obtained when, instead of eliminating only one sample to calculate the concentration of k latent variables, the highest possible fraction of the samples is cancelled. The importance of full cross-validation [15], compared with partial cross-validation [16][17], has been shown. In other words, it is essential that, in the calculation process for the PLS model, neither the cancellation group nor an initial autoscaling that affects all the samples intervenes in any way. If the data were autoscaled, the mean and variance of all the samples would intervene. In this work, the full cross-validation procedure, PLSC, is used instead of partial cross-validation.

The calculation of PRESS was performed with three cancellation groups, which is to say that a PLSC model was constructed three times for a number of latent variables, eliminating 9, 8, and 8, respectively, of the 25 absorption spectra [18][19].

Table 3 shows the results in percentages of explained variance and cross-validation variance as a function of the number of latent variables. It is obvious that, upon including new latent variables, the explained variance rises; however, when the model includes a ith latent variable not related to the response, the cross-validate variance will not continue increasing but will rather decrease. The minimum PRESS is reached for the number of latent variables that give the maximum cross-validate variance.

According to this criterion, one must take 7 and 5 latent variables for rifamycin SV and rifampicin, respectively. In all of the cases, the cross-validate variance exceeded 99.877%, and there is always more than 99.881% explained variance associated with it.

The concentration found with this model for each drug was compared to the real value. The average relative absolute error obtained was 1.641% in the calibration of rifamycin SV and 3.972% in the case of rifampicin.

To check the performance of the PLSC-calibration models, they were applied to a test set of five additional solutions (*Fig. 2*) different from those that built the model. The figure of merit considered at this stage was accuracy, which includes precision and trueness (*Table 4*).

Table 3. *Variance Explained in the Blocks of Predictors and Response, and Cross-Validation Variance* (C. V.) *for the Concentration of Rifamycin SV and Rifampicin by Using the PLS Model Constructed with the Original Signal*

Latent variables index	Rifamycin SV			Rifampicin		
	Explained variance of *Y* block [%]	C. V. Explained variance of *Y* block [%]	Variance of *X* block [%]	Explained variance of *Y* block [%]	C. V. Explained variance of *Y* block [%]	Variance of *X* block [%]
1	83.050	83.523	89.467	89.607	89.928	89.466
2	94.939	95.163	93.236	98.510	98.633	94.227
3	99.658	99.678	97.815	99.726	99.739	97.815
4	99.829	99.839	98.562	99.836	99.832	98.375
5	99.932	99.914	99.328	99.881	99.877[a])	99.326
6	99.966	99.940	99.469	99.909	99.803	99.470
7	99.979	99.945[a])	99.550			
8	99.984	99.940	99.567			

[a]) Maximum cross-validation variance reached.

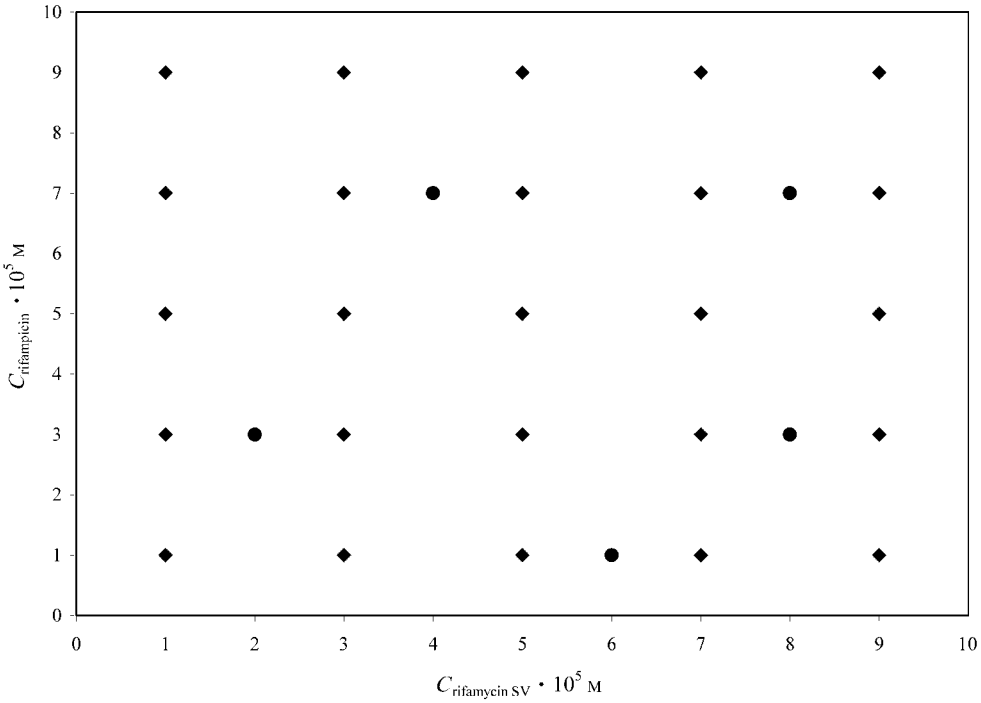


Fig. 2. *Set of calibration samples* (◆) *and test samples* (●) *of rifamycin SV and rifampicin*

The accuracy of the predictions, or total error, was calculated as the mean-square error of prediction (MSEP) and as a percentage, through the relative root mean-square error of prediction (RRMSEP).

Table 4. *Predictions for the Models Constructed with the Original Signal and the Selected Potentials*

	PLS Models constructed with the entire set of wavelengths		PLS Models constructed with a previous selection of wavelengths	
	Rifamycin SV	Rifampicin	Rifamycin SV	Rifampicin
MSEP	2.950×10^{-12}	8.149×10^{-13}	1.540×10^{-12}	2.760×10^{-12}
RRMSEP	3.067	2.149	2.216	3.959
BCMSEP	3.687×10^{-12}	1.019×10^{-12}	1.925×10^{-12}	3.450×10^{-12}
$F_{calc.}$ (JCIT)	0.096	0.726	0.320	1.060

$$\mathrm{MSEP}(k) = \frac{\sum_{i=1}^{e} (\hat{c}_i(k) - c_i)^2}{e} \quad \text{and} \quad \mathrm{RRMSEP}(k) = \frac{100}{\bar{c}} \sqrt{\mathrm{MSEP}}$$

where c_i is the concentration corresponding to the ith evaluation sample, $\hat{c}_i(k)$ is the concentration estimated by the PLS model with only k latent variables for the same sample, e the number of samples in the test set, and c is the mean of the real concentrations.

Precision or variance in the prediction can be estimated by calculating the bias-corrected mean-square error of prediction (BCMSEP),

$$\mathrm{BCMSEP}(k) = \frac{\sum_{i=1}^{e} (\hat{c}_i(k) - c_i)^2 - \frac{\left[\sum_{i=1}^{e} (\hat{c}_i(k) - c_i)\right]^2}{e}}{e - 1}$$

and can be statistically compared to the precision of another method, or with the same method under different conditions, by an F-test of comparison of variances.

Trueness is verified by the absence of bias, which can be evaluated with the joint confidence interval test of the slope and the intercept, taking into account errors on both axes (JCIT) for the real concentrations and the concentrations predicted by the model.

Since the predictions were unbiased at the usual 95% significance level ($F_{0.05,2,3} = 9.550$), while the prediction errors in RRMSEP terms were 0.096 for rifamycin SV and 0.726 for rifampicin, we can affirm that the procedure proposed is suitable for the joint calibration of these drugs.

Wavelength Selection by the Genetic Algorithm. With the aim of increasing the quality of the calibration, a selection of wavelengths was performed, in such a way that those providing little or nothing to the model were eliminated. This selection of variables was carried out using the genetic algorithm GA as described in [20]. The steady-state GA without duplicates, which was used in this case, has already been used with success [12][13][18]. With the GA applied under the conditions indicated in [12], the predictor variables (wavelengths) selected were 305, 340, 350, 365, 380, 395, 400, 405, 425, 485, 515, 580, and 650 nm for rifamycin SV, and 240, 285, 400, 485, 530, 560, 570, 575, and 580 nm for rifampicin.

The number of variables necessary to make a calibration was reduced from the original 114 to 13 and 9 in the case of rifamycin SV and rifampicin, respectively.

The PLSC model constructed with these selected variables is shown in *Table 5*. Comparing these results with those in *Table 3*, it can be seen that the reduction in the original variables reduced the number of latent variables needed to reach the minimum PRESS. This is probably because the wavelengths that do not provide fundamental information have been eliminated. Those wavelengths do not only fail to provide important information, but also lead to variance in the spectra, resulting in the introduction of a new latent variable.

Table 5. *Variance Explained in the Blocks of Predictors and Response, and Cross-Validation Variance* (C. V.) *for the Concentration of Rifamycin SV and Rifampicin by Using the PLS Model Constructed with Wavelengths Selection by Genetic Algorithm*

Latent variables index	Rifamycin SV			Rifampicin		
	Explained variance of *Y* block [%]	C. V. Explained variance of *Y* block [%]	Variance of *X* block [%]	Explained variance of *Y* block [%]	C. V. Explained variance of *Y* block [%]	Variance of *X* block [%]
1	82.028	82.544	87.938	89.287	89.615	92.116
2	99.239	99.275	97.831	94.359	94.673	97.371
3	99.923	99.922	99.575	99.899	99.596	99.794
4	99.980	99.977[a])	99.810	99.918	99.908[a])	99.844
5	99.980	99.975	99.892	99.922	99.859	99.944

[a]) Maximum cross-validation variance reached.

The average relative absolute error obtained was 0.985% in the calibration of rifamycin SV and 4.434% in the case of rifampicin.

These PLSC-calibration models were also applied to the test set (*Fig. 2*). The figures of merit considered in this stage are shown in *Table 4*.

From a statistical point of view, a test of the *t* of paired measurements conducted on the concentration values calculated by PLSC minus the concentration values calculated with the GA show that, at the level of $\alpha = 0.050$, both procedures must be considered equal, for rifamycin SV ($\alpha_{\text{actual}} = 0.618$), and rifampicin ($\alpha_{\text{actual}} = 0.327$).

Therefore, we can conclude that both procedures are viable for the resolution of mixtures of rifamycin SV and rifampicin. Furthermore, the selection of the predictor variables (wavelengths) by means of GA improves the prediction capacity and reduces the number of latent variables to construct the PLS model.

These methods have been successfully applied in the resolution of not only synthetic samples but also in complex matrices such as spiked urine samples. Here, the concentrations found for rifamycin SV ($C_{\text{theor.}} = 0.400$ μmol dm^{-3}) and rifampicin ($C_{\text{theor.}} = 0.400$ μmol dm^{-3}) were 0.395 ± 0.019 and 0.387 ± 0.015 μmol dm^{-3} ($n = 4$ and $\alpha = 0.050$), respectively.

This work was supported by the *Junta de Castilla y León* (BU08/01) and *Ministerio de Ciencia y Tecnología* (BQU2001-1126). *M. Asunción Alonso Lomillo* also thanks the *Junta de Castilla y León* and the *European Social Fund* for the *F. P. I.* grant.

Experimental Part

General. Anal.-grade chemicals were used without further purification. All solns. were prepared with *Milli-Q* water. N_2 (99.99%) was used to remove dissolved O_2. Solns. of rifamycin SV and rifampicin (*Fluka*, CH-Buchs) were prepared by dissolving the adequate amount of each compound in H_2O. *Britton Robinson* buffer, pH 7, was used. A *Perkin-Elmer* model *Lambda 3B* spectrophotometer (*Norwalk*, Connecticut, USA) with double beam was used. The spectra of solns. were recorded over the range 650 to 190 nm; scan speed, 60 nm/min, and wavelength interval, 5 nm. The pH of the solns. was measured with a *Crison* model *2002* (E-Barcelona) pH-meter. Absorbance spectra were acquired and processed by using software supplied by *Perkin-Elmer* [21]. Data analysis was performed with PARVUS [22] for the regression models and PROGRESS [23] for the robust regression. Test of the *t* of paired measurements was achieved with STATGRAPHICS [24].

Experimental Procedure. The calibration set in the resolution of samples of rifamycin SV and rifampicin were 25 synthetic samples containing different amounts of each drug (*Fig. 2*). The design of the concentrations enabled us to analyze five levels of concentration for each analyte in presence of a different amount of the other one. All spectra were digitalized accounting for the absorbance read at 93 wavelengths between 650 and 190 nm.

REFERENCES

[1] A. Bacchi, G. Pelizzi, M. Nebuloni, P. Ferrari, *J. Med. Chem.* **1998**, *41*, 2319.
[2] M. Litter, Farmacología Experimental y Clínica, El Ateneo, Argentina, 1986.
[3] T. J. Ward, C. Dann III, A. Blaylock, *J. Chromatogr., A,* **1995**, *715*, 337.
[4] M. A. Alonso, S. Sanllorente, L. A. Sarabia, M. J. Arcos, *Anal. Chim. Acta* **2000**, *405*, 123.
[5] B. Ratti, R. R. Parenti, A. Toselli, L. F. Zerilli, *J. Chromatogr., A* **1981**, *225*, 526.
[6] K. L. Swart, M. Papgis, *J. Chromatogr., A* **1992**, *593*, 21.
[7] Y. Y. Lau, G. D. Hanson, B. J. Carel, *J. Chromatogr., B* **1996**, *676*, 125.
[8] K. C. Jindal, R. S. Chaudhary, S. S. Gangwal, A. K. Singla, S. Khanna, *J. Chromatogr., A* **1994**, *685*, 195.
[9] C. R. Pasqualucci, A. Vigevani, P. Radaelli, G. G. Gallo, *J. Pharm. Sci.* **1970**, *59*, 685.
[10] M. I. Walash, F. Belal, M. E. Metwally, M. M. Hefnawy, *Anal. Lett.* **1993**, *26*, 1905.
[11] M. A. Alonso Lomillo, O. Domínguez Renedo, M. J. Arcos Martínez, *Electroanalysis* **2002**, *14*, 634.
[12] M. J. Arcos, M. C. Ortiz, B. Villahoz, L. A. Sarabia, *Anal. Chim. Acta* **1997**, *339*, 63.
[13] M. J. Arcos, C. Reguera, M. C. Ortiz, *Electrochim. Acta* **1998**, *43*, 479.
[14] L. Davis, ' Handbook of Genetic Algorithms', Van Nostrand Reinhold, New York, 1991.
[15] S. Lanteri, *Chemlab Intell. Lab. Syst.* **1992**, *15*,159.
[16] M. Stone, *J. Royal Stat. Soc.* **1974**, *36B*, 111.
[17] S. Wold, *Technometrics* **1978**, *20*, 397.
[18] O. Domínguez, M. J. Arcos, *Electroanalysis* **2000**, *12*, 449.
[19] C. Reguera, M. C. Ortiz, M. J. Arcos, *Quím. Anal.* **1999**, *18*, 105.
[20] R. Leardi, *J. Chemom.* **1994**, *8*, 65.
[21] PECSS, Perkin-Elmer Computerized Spectroscopy Software, The Perkin-Elmer Corporation, Norwalk, Connecticut, USA, 1988.
[22] M. Forina, S. Lanteri, C. Armanino, R. Leardi, G. Drava, 'PARVUS: An Extendable Package of Programs for Data Explorative Analysis, Classification and Regression Analysis', Version 1.3, available from the authors, Istituto di Analisi e Tecnologie Farmaceutiche ed Alimentari, Universitá di Genova, 1994.
[23] P. J. Rousseuw, A. M. Leroy, 'Robust Regression and Outlier Detection', John Wiley & Sons, New York, 1987.
[24] Statgraphics Plus for Windows, Version 4.0, Copyright 1994 – 1999 by Statistical Graphics Corp.

Received February 25, 2004