Determination of Thiamine HCl and Pyridoxine HCl in Pharmaceutical Preparations Using UV–Visible Spectrophotometry and Genetic Algorithm Based Multivariate Calibration Methods

Durmus ÖZDEMİR*,[a] and Erdal DINC[b]

[a] Department of Chemistry, Faculty of Science, İzmir Institute of Technology; 35437 Urla-İzmir, Turkey: and [b] Department of Analytical Chemistry, Faculty of Pharmacy, University of Ankara; 06100 Tandoğan-Ankara, Turkey.
Received February 18, 2004; accepted April 22, 2004

Simultaneous determination of binary mixtures pyridoxine hydrochloride and thiamine hydrochloride in a vitamin combination using UV–visible spectrophotometry and classical least squares (CLS) and three newly developed genetic algorithm (GA) based multivariate calibration methods was demonstrated. The three genetic multivariate calibration methods are Genetic Classical Least Squares (GCLS), Genetic Inverse Least Squares (GILS) and Genetic Regression (GR). The sample data set contains the UV–visible spectra of 30 synthetic mixtures (8 to 40 μg/ml) of these vitamins and 10 tablets containing 250 mg from each vitamin. The spectra cover the range from 200 to 330 nm in 0.1 nm intervals. Several calibration models were built with the four methods for the two components. Overall, the standard error of calibration (SEC) and the standard error of prediction (SEP) for the synthetic data were in the range of $<$0.01 and 0.43 μg/ml for all the four methods. The SEP values for the tablets were in the range of 2.91 and 11.51 mg/tablets. A comparison of genetic algorithm selected wavelengths for each component using GR method was also included.

Key words UV–visisible spectrophotometry; multivariate calibration; genetic algorithms; genetic regression; classical least squares (CLS); Inverse Least Squares (ILS).

The mixture of pyridoxine hydrochloride and thiamine hydrochloride as an antinevritic is widely used in vitamin combinations. The resolution of the mixtures systems containing two or more compounds without any separation procedure in the presence of excipients existing in samples is one of the main issues of the simultaneous quantitative determination. The quantitative determination of the vitamins in combinations containing pyridoxine hydrochloride or thiamine hydrochloride with other active compounds using various methods including spectrophotometry,[1—3] HPLC,[4] fluorescence,[5] capillary electrophoresis,[6] and potentiometric,[7] have been described for several mixtures and vitamin combinations.

Modern spectroscopic instruments are so fast that they can produce hundreds of spectra in a few minutes for a given sample that contain multiple components. Unfortunately, univariate calibration methods are not suitable for this type of data, as they require an interference free system. Multivariate calibration deals with data containing instrument responses measured on multiple wavelengths for a sample that usually contains more than one component. In recent years, advances in chemometrics and computers have lead to the development of several multivariate calibration methods[8—11] for the analysis of complex chemical mixtures.

Genetic regression (GR) is a calibration technique that optimizes linear regression models using a genetic algorithm (GA) and has been applied to a number of multi-instrument calibration and wavelength selection problems.[12—16] GA's are non-local search and optimization methods that are based upon the principles of natural selection.[17—20] For a given full spectrum data, GR selects an optimum linear combination of wavelengths and simple mathematical operators to build a linear calibration model using simple least squares method.

Classical Least Squares (CLS) extends the classical Beer's Law model in which the absorbance at each wavelength is directly proportional to the component concentrations. Inverse Least Squares (ILS) is based on the inverse Beer's Law where concentrations of an analyte are modelled as a function of absorbance measurements. Genetic Classical Least Squares (GCLS) and Genetic Inverse Least Squares (GILS) are modified versions of original CLS and ILS methods in which a small set of wavelengths are selected from a full spectral data matrix and evolved to an optimum solution using a genetic algorithm.

In this work, CLS and three different genetic algorithm based calibration methods GCLS, GILS and GR were tested with the aim of establishing calibration models that have a high predictive capacity for the simultaneous determination of thiamine HCl and pyridoxine HCl in their binary mixtures and in a commercial vitamin preparation using the UV–visible spectrophotometry.

Genetic Regression Genetic algorithms (GA) are global search and optimization methods based upon the principles of natural evolution and selection as developed by Darwin. Computationally, the implementation of a typical GA is quite simple and consists of five basic steps including initialization of a gene population, evaluation of the population, selection of the parent genes for breeding and mating, crossover and mutation, and replacing parents with their offspring. These steps have taken their names from the biological foundation of the algorithm.

Genetic regression (GR) is an implementation of a GA for selecting wavelengths and mathematical operators to build linear calibration models. GR is a hybrid calibration between univariate and multivariate calibration techniques in which it optimizes simple linear regression models through an evolving selection of wavelengths and simple mathematical operators (+, −, *, /). GR follows the same basic initialize/breed/mutate/evaluate algorithm as other GA's but differs in the way it encodes genes. A gene is a potential solution to a given problem and the exact form may vary from application

* To whom correspondence should be addressed. e-mail: durmusozdemir@iyte.edu.tr © 2004 Pharmaceutical Society of Japan

to application. Here, the term gene is used to describe the collection of instrument response pairs combined with the above mentioned operators. These pairs, called 'base pairs', are then combined with an addition operator to produce a score, which relates the instrument response to component concentration. The term 'population' is used to describe the collection of individual genes in the current generation.

In the initialization step, first generation of genes is created randomly with a fixed population size. Although random initialization helps to minimize bias and maximize the number of possible recombinations, GR is designed to select initial genes in a somewhat biased random fashion in order to start with genes better suited to the problem than those that would be randomly selected. Biasing is done with a correlation coefficient by plotting the scores of initial genes against the component concentrations. The size of the gene pool is a user defined even number in order to allow breeding of each gene in the population. It is important to note that the larger the population size, the longer the computation time. The number of base pairs in a gene is determined randomly between a fixed low limit and high limit. The lower limit was set to 2 in order to allow single point crossover whereas the higher limit was set to eliminate overfitting problems and reduce the computation time. Once the initial gene population is created, the next step is to evaluate and rank the genes using a fitness function, which is the inverse of the standard error of calibration (SEC).

The third step is where the basic principle of natural evolution is put to work for GR. This step involves the selection of the parent genes from the current population for breeding using a roulette wheel selection method according to their fitness values. The goal is to give a higher chance to those genes with high fitness so that only the best performing members of the population will survive in the long run and will be able to pass their information to the next generations. Because of the random nature of the roulette wheel selection method, however, genes with low fitness values will also have some chance to be selected. Also, there will be genes that are selected multiple times and some genes will not be selected at all and will be thrown out of the gene pool. After the selection procedure is completed, the selected genes are allowed to mate top-down without ranking whereby the first gene mates with the second gene and the third one with the fourth one and so on as illustrated in the following example:

Parents

$$S_1=(A_{347}*A_{251})^{\#}+(A_{379}+A_{218}) \tag{1}$$

$$S_2=(A_{225}*A_{478})^{\#}+(A_{343}/A_{250})+(A_{451}-A_{358})+(A_{231}-A_{458}) \tag{2}$$

The points where the genes are cut for mating are indicated by $^{\#}$.

Offspring

$$S_3=(A_{347}*A_{251})+(A_{343}/A_{250})+(A_{451}-A_{358})+(A_{231}-A_{458}) \tag{3}$$

$$S_4=(A_{225}*A_{478})+(A_{379}+A_{218}) \tag{4}$$

Here the first part of the S_1 is combined with the second part of the S_2 to give the S_3, likewise the second part of the S_1 combined with the first part of the S_2 to give S_4. This process is called the singe point crossover and is the one used in GR. The single point crossover will not provide different offspring if both parent genes are identical, which may happen in the roulette wheel selection, and broken at the same point. Also note that mating can increase or decrease the number of base pairs in the offspring genes. After crossover, the parent genes are replaced by their offspring and the offspring are evaluated. The ranking process is based on their fitness values following the evaluation step. Then the selection for breeding/mating starts all over again. This is repeated until a predefined number of iterations are reached.

Mutation which introduces random deviations into the population was also introduced into the GR during the mating step at a rate of 1% as is typical in GA's. This is usually done by replacing one of the base pairs in an existing gene with a randomly generated new base pair. Mutation allows the GR to explore the search space and incorporate new material into the genetic population. It helps keep the search moving and can eject GR from a local minimum on the response surface. However, it is important not to set the mutation rate too high since it may keep the GA from being able to exploit the existing population.

In the end, the gene with the lowest SEC (highest fitness) is selected for the model building which is done by simple least squares. This model is used to predict the concentrations of component being analyzed in the validation (test) sets. The success of the model in the prediction of the validation sets are evaluated using standard error of prediction (SEP). Because the random processes are heavily involved in the GR as in all the GA's, the program has been set to run several times for each component in a given multi-component mixture during the course of this study. The best run, (*i.e.* the one generating the lowest SEC for the calibration set and at the same time producing SEP's for validation sets that are in the same range with the SEC) was subsequently selected for evaluation and further analysis. The termination of the algorithm can be done in many ways. The easiest way is to set a predefined iteration number for the number of breeding/mating cycles.

GR has some major advantages over classical univariate and multivariate calibration methods. It is a hybrid calibration method that uses the full spectral information and reduces it to a single score upon which simple calibration models are built. First of all, it is as simple as univariate calibration in terms of the mathematics involved in the model building and prediction steps, but at the same time it has the advantages of the multivariate calibration methods since it uses the full spectrum to extract genetic scores. It automatically corrects baseline fluctuations with the use of simple mathematical operators while forming the base pairs. Note that no data pretreatment is necessary before calibration, which saves the extra time in the data processing.

Genetic Classical Least Squares The classical least squares (CLS) method extends the classical Beer's Law model in which the absorbance at each wavelength is directly proportional to the component concentrations. Model errors are assumed to be in the measurement of the instrument responses as it was in the classical univariate method. In matrix notation, the CLS model for m calibration samples containing l chemical components whose spectra contain n wavelengths is described as:

$$\mathbf{A}=\mathbf{CK}+\mathbf{E_A} \tag{5}$$

Where $\mathbf{A}$ is the $m\times n$ matrix of the calibration spectra, $\mathbf{C}$ is

the $m \times l$ matrix of the component concentrations, **K** is the $l \times n$ matrix of absorptivity-pathlength constants and $\mathbf{E_A}$ is the $m \times n$ matrix of the spectral errors or residuals not fit by the model. Here the **K** matrix represents the first order estimates of the pure component spectra at unit concentration and unit pathlength. The method of least-squares can be used to estimate the **K** matrix. The least-squares estimate of the **K** is defined as:

$$\hat{\mathbf{K}} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{A} \quad (6)$$

Once the estimated $\hat{\mathbf{K}}$ matrix obtained, the concentrations of an unknown sample can be predicted from its spectrum by:

$$\hat{\mathbf{c}} = (\hat{\mathbf{K}}\hat{\mathbf{K}}')^{-1}\hat{\mathbf{K}}\,\mathbf{a} \quad (7)$$

Where **a** is the spectrum of the unknown sample and $\hat{\mathbf{c}}$ is the vector of the predicted component concentrations. Genetic Classical Least Squares (GCLS) is a modified version of the original CLS method in which a small set of wavelengths are selected from a full spectral data using a genetic algorithm. The algorithm used to select the optimum number of wavelengths in GCLS is quite similar to the GR algorithm, but differs in the way it encodes the gene. In GCLS, the term 'gene' describes a vector whose elements are randomly selected wavelengths. The size of the vector is also determined in a random fashion with an upper limit to reduce computation time.

In the initialization step, an even number of genes are formed from full a spectral data matrix and each gene is used to form a CLS model. These models are then evaluated and ranked using the fitness function described in GR. The roulette wheel method is then used to select the gene population for breeding. After the selection procedure is completed, the selected genes are allowed to mate top-down without ranking whereby the first gene mates with second gene and third one with fourth one and so on as described in above with one difference. Since the genes used in GCLS are only vector of wavelengths and contains no base pairs as described in GR, for each gene a random number is generated between 1 and the length of the gene and the single point crossover process is performed using this number. After crossover, the parent genes are replaced by their offspring and the offspring are evaluated. The ranking process is based on their fitness values and follows the evaluation step. Then the selection for breeding/mating starts all over again. This is repeated until a predefined number of iterations are reached. In each iteration, the best gene with the lowest SEC is stored in order to compare it with the best gene of the next generation. If the next generation produces a better gene then it is replaced with the older one; otherwise the old one kept for further iterations. At the end, the gene with the lowest SEC is selected for model building. This model is used to predict the concentrations of component being analyzed in the validation (test) sets as described in GR.

Genetic Inverse Least Squares The major drawback of the CLS is that all of the interfering species must be known and their concentrations included in the model. This need can be eliminated by using the inverse least squares (ILS) method which uses the inverse of Beer's Law. In the ILS method, concentrations of an analyte are modelled as a function of absorbance measurements. Because modern spectroscopic instruments are very stable and provide excellent signal-to-noise (S/N) ratios, it is believed that the majority of errors lie in the reference values of the calibration sample, not in the measurement of their spectra. The ILS model for m calibration samples with n wavelengths for each spectrum is described by:

$$\mathbf{C} = \mathbf{AP} + \mathbf{E_C} \quad (8)$$

Where **C** and **A** are the same as in CLS, **P** is the $n \times l$ matrix of the unknown calibration coefficients relating l component concentrations to the spectral intensities and $\mathbf{E_C}$ is the $m \times l$ matrix of errors in the concentrations not fit by the model. In the calibration step, ILS minimizes the squared sum of the residuals in the concentrations. The biggest advantage of ILS is that equation 8 can be reduced for the analysis of single component at a time since analysis is based on an ILS model is invariant with respect to the number of chemical components included in the analysis. The reduced model is given as:

$$\mathbf{c} = \mathbf{Ap} + \mathbf{e_c} \quad (9)$$

Where **c** is the $m \times 1$ vector of concentrations for the analyte that is being analyzed, **p** is $n \times 1$ vector of calibration coefficients and $\mathbf{e_c}$ is the $m \times 1$ vector of concentration residuals not fit by the model. During the calibration step, the least-squares estimate of **p** is:

$$\hat{\mathbf{p}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \cdot \mathbf{c} \quad (10)$$

Where $\hat{\mathbf{p}}$ is the estimated calibration coefficients. Once $\hat{\mathbf{p}}$ is calculated, the concentration of the analyte of interest can be predicted with the equation below.

$$\hat{c} = \mathbf{a}' \cdot \hat{\mathbf{p}} \quad (11)$$

Where $\hat{c}$ is the scalar estimated concentration and **a** is the spectrum of the unknown sample. The ability to predict one component at a time without knowing the concentrations of interfering species has made ILS one of the most frequently used calibration methods. However, the identity of interfering species still need to be known to prepare a good calibration sample set.

The major disadvantage of ILS can be seen in equation 10 where the matrix, which must be inverted, has dimensions equal to the number of wavelengths in the spectrum and this number can not exceed the number of calibration samples. This is a big restriction since the number of wavelengths in a spectrum will generally be more than the number of calibration samples and the selection of wavelengths that provide the best fit for the model is not a trivial process. Several wavelength selection strategies, such as stepwise wavelength selection and all possible combination searches, are available to build an ILS model that fits the data best. Here we used the same genetic algorithm described in GCLS to build genetic inverse least squares (GILS) models with one difference. This difference is in the way the mating and single point crossover operations are carried out. Because the number of wavelengths is restricted in response matrix A in the ILS, the size of the largest gene is restricted to one less than the number of calibration samples in the concentration vector. However, if the single point crossover is set to take place in any point of a gene, then the mating step could produce new genes that have a larger number of wavelengths than the

number of calibration samples even though all the genes in the initial gene pool were set to have smaller number of wavelengths than the size of the concentration vector. In order to avoid this problem, the crossover operation is only performed in the middle of each gene in GILS so that the new generations will never have larger sizes than the number of calibration samples. The rest of the algorithm is the same as the one used in GCLS.

Experimental

Materials In this work, a vitamin tablet (Benexol® film-coated tablet, Roche Pharm. Ind., Turkey, Batch no. 10169) consisting of 250 mg pyridoxine HCl and 250 mg thiamine HCl per tablet was investigated. Roche Pharm. Ind., Turkey kindly donated the active vitamin compounds. Stock solutions of 100 mg/100 ml of thiamine HCl and pyridoxine HCl were prepared by using 0.1 M HCl. The standard solutions in 25-ml volumetric flasks containing 8—40 μg/ml for both vitamins were obtained from their stock solutions by appropriate dilution. The concentration profiles of calibration and validation samples were designed in a way that minimizes colinearity problem since a binary system has been studied. For the commercial vitamin, ten tablets were accurately weighed and powdered in a mortar. An amount equivalent to one tablet was dissolved in 0.1 M HCl in a 100 ml calibrated flask by sonication. The solution was filtered into a 100 ml calibrated flask through Whatman no. 42 filter paper and diluted to an appropriate volume with the same solvent.

Methods Sample spectra were measured on a Shimadzu UV-1600 double beam UV–visible spectrophotometer from 215 to 330 nm with 0.1 nm intervals. Quartz cells with 1 cm pathlengths were used. The CLS and the three new genetic algorithms based multivariate calibration methods (GCLS, GILS and GR) were written in MATLAB programming language using Matlab 5.3 (MathWorks Inc, Natick, MA, U.S.A.). The text files for calibration, validation and prediction sets were generated with the use of Microsoft Excel (MS office 97, Microsoft Corporation, CA, U.S.A.).

Results and Discussion

Two separate experiments were carried out. In the first case, a total of 20 samples were selected to be included in the calibration set and 10 samples were used to construct the validation set. In addition, 10 samples of commercial tablets having 24 μg/ml of each constituent were used to build prediction set. In a second design, two of the commercial tablet spectra were added to the calibration set in order to better represent actual samples in the calibration step.

UV spectra of pure thiamine (24 μg/ml), pyridoxine (24 μg/ml) in 0.1 M HCl along with the binary mixture of the two components between 220 and 330 nm wavelength range are shown in Fig. 1. As seen from the figure, thiamine shows a maximum absorbance around 245 nm and pyridoxine has a maximum absorbance around 290 nm. Their mixture spectrum, however, indicates strong overlap over the entire region which indicates that the use of multivariate methods would be needed to resolve these components. Throughout the genetic multivariate calibration process, it is expected that these overlaps will be resolved and reveal the information necessary to build successful calibration models otherwise almost impossible with univariate calibration methods.

Several calibration models were generated with the four methods and Table 1 shows the results of binary mixtures in the first experiment for calibration and validation sets. Here, the CLS method was applied to whole spectrum data set and in the case of genetic algorithm based methods (GCLS, GILS and GR) the algorithms were set to rum 10 times with 20 gene and 100 iterations in each run. The results given in table for GCLS, GILS and GR are from the runs that generate the lowest SEC and SEP combination. Then these models were used later to predict the actual tablet samples and compared with each other based on their success of predicting actual samples as shown in Table 2.

A close examination of the results given in Table 1 indicates that CLS, GCLS and GR generates approximately the same results whereas GILS produces almost twice better results than other three. However, this could be very mislead-

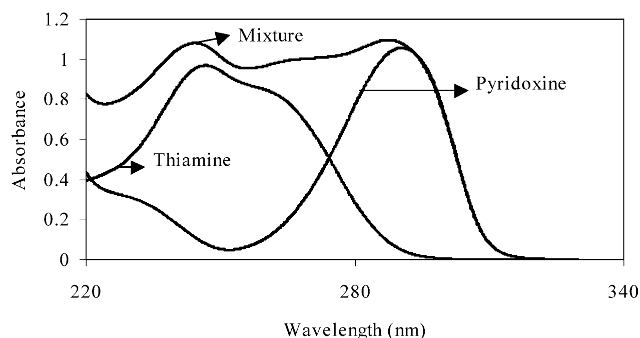

Fig. 1. UV Spectra of Pure Thiamine (24 μg/ml) and Pyridoxine (24 μg/ml) in 0.1 M HCl along with a Binary Mixture of the Two Components between 220 and 330 nm Wavelength Range

Table 1. Results of Mixtures in Calibration and Validation Sets Containing Thiamine and Pyridoxine Obtained with Four Multivariate Calibration Methods (CLS, GCLS, GILS and GR) When the Calibration Models Were Built Only with Sythetic Samples

Component	Data sets	Parameters	Methods			
			CLS	GCLS	GILS	GR
Thiamine HCL	Calibration	SEC (μg/ml)	0.19	0.19	0.13	0.20
		Average recovery (%)	100.22	100.25	100.00	100.13
		RSD	1.48	1.40	0.69	1.57
	Validation	SEP (μg/ml)	0.22	0.22	0.13	0.23
		Average recovery (%)	100.47	100.49	100.01	100.58
		RSD	1.71	1.62	0.81	1.80
Pyridoxine HCl	Calibration	SEC (μg/ml)	0.38	0.37	0.17	0.36
		Average recovery (%)	100.97	100.90	100.23	100.45
		RSD	2.13	2.00	0.86	1.65
	Validation	SEP (μg/ml)	0.34	0.32	0.16	0.28
		Average recovery (%)	100.69	100.65	100.23	100.95
		RSD	2.08	1.92	0.70	1.52

SEC: standard error of calibration, SEP: standard error of prediction, RSD: relative standard deviation.

Table 2. Results of 10 Commercial Vitamin Tablets Containing Thiamine (250 mg) and Pyridoxine (250 mg) Obtained with Four Multivariate Calibration Methods (CLS, GCLS, GILS and GR) When the Calibration Models Were Built Only with Synthetic Samples

Actual (250 mg/tablet)	Predicted (mg/tablet)							
	Thiamine HCl				Pyridoxine HCl			
	CLS	GCLS	GILS	GR	CLS	GCLS	GILS	GR
Mean	238.88	244.96	249.11	245.99	243.84	248.80	251.76	245.77
SD	3.19	3.17	8.16	2.90	3.19	3.02	5.87	3.16
RSD	1.34	1.29	3.28	1.18	1.31	1.22	2.33	1.28
SEP	11.52	5.87	7.79	4.86	6.86	3.11	5.84	5.18

SD: standard deviation, RSD: relative standard deviation, SEP: standard error of prediction.

Table 3. Results of Mixtures in Calibration and Prediction Sets Containing Thiamine and Pyridoxine Obtained with Four Multivariate Calibration Methods (CLS, GCLS, GILS and GR) When the Calibration Models Were Built with Synthetic Samples and Spectra of Two Tablets

Component	Data sets	Parameters	Methods			
			CLS	GCLS	GILS	GR
Thiamine HCl	Calibration	SEC (μg/ml)	0.39	0.18	0.07	0.23
		Average recovery (%)	100.19	100.19	100.02	100.11
		RSD	2.35	1.42	0.32	1.68
	Validation	SEP (μg/ml)	0.29	0.22	0.07	0.24
		Average recovery (%)	101.14	100.54	99.95	100.73
		RSD	1.77	1.60	0.30	1.65
Pyridoxine HCl	Calibration	SEC (μg/ml)	0.43	0.36	$<$0.01	0.34
		Average recovery (%)	100.85	100.79	100.00	100.45
		RSD	2.41	2.00	$<$0.01	1.50
	Validation	SEP (μg/ml)	0.36	0.33	$<$0.01	0.29
		Average recovery (%)	101.09	100.08	100.00	100.76
		RSD	2.04	1.86	$<$0.01	1.47

SEC: standard error of calibration, SEP: standard error of prediction, RSD: relative standard deviation.

Table 4. Results of Commercial Vitamin Tablets Containing Thiamine (250 mg) and Pyridoxine (250 mg) Obtained with Four Multivariate Calibration Methods (CLS, GCLS, GILS and GR) When the Calibration Models Were Built with Synthetic Samples and Spectra of Two Tablets

Actual (250 mg/tablet)	Predicted (mg/tablet)							
	Thiamine HCl				Pyridoxine HCl			
	CLS	GCLS	GILS	GR	CLS	GCLS	GILS	GR
Mean	240.66	250.93	247.93	251.47	245.13	249.75	249.25	251.68
SD	3.44	3.10	5.23	3.16	3.29	3.10	5.82	3.22
RSD	1.43	1.23	2.11	1.26	1.34	1.24	2.34	1.28
SEP	9.88	3.04	5.31	3.30	5.76	2.91	5.50	3.45

SD: standard deviation, RSD: relative standard deviation, SEP: standard error of prediction.

ing conclusion if one considers the results given in Table 2 where it generates the lowest means and highest SEP's for both components. It is evident that the hard modelling method CLS is unable to predict the composition of actual tablets as good as the genetically modified multivariate methods. In addition, the GILS method produced the largest relative standard deviation (RSD) for the tablets even though it generated the lowest RSD values for the synthetic mixtures which might be the indication of slight overfit for the model. In terms of the overall performance of the four methods it can be said that the genetically modified methods performs better than CLS for actual tablet samples.

In order to see whether the calibration models generated with only synthetic data are suitable or not, a second experiment was carried out in which the two of the tablet spectra are added to the calibration set. The expectation was that the added tablet spectra in calibration set will generate better and more robust calibration models. The results of this experiment are shown in Table 3 and 4 for synthetic and actual tablets, respectively. Upon the comparison of the result given in Table 2 and 4, it is evident that predictions of actual tablets were improved for all the methods except GILS investigated in this study. Particularly the prediction of thiamine in tablets with GCLS and GILS were significantly better than those obtained in the first experiment.

Figures 2 and 3 show the plot of actual *vs*. predicted thiamine and pyridoxine concentrations, respectively for the calibration and validation sets obtained with the four meth-

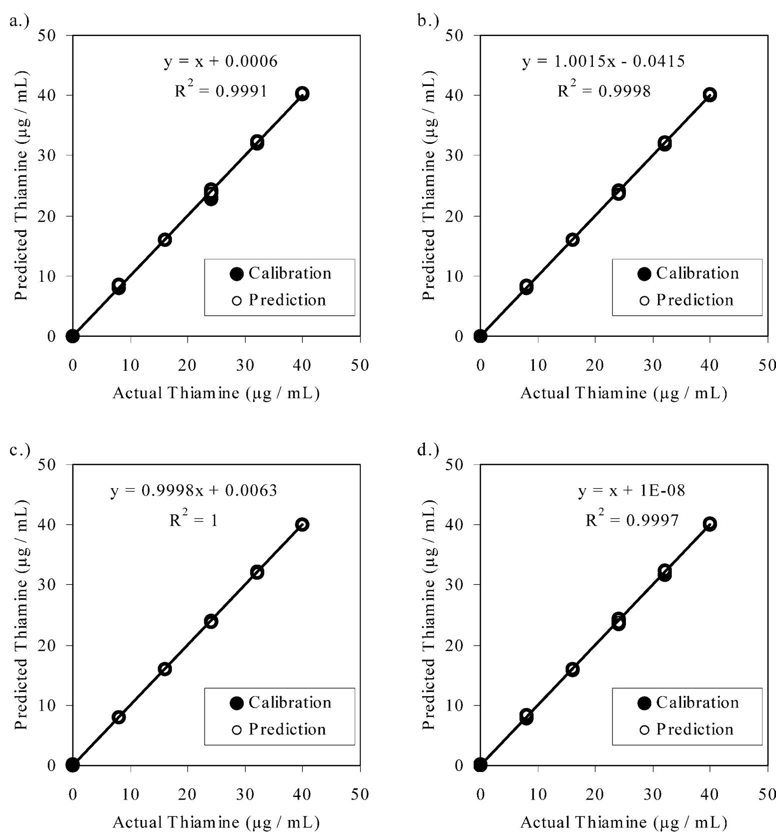

Fig. 2. Plots of Actual *vs.* Predicted Thiamine Concentrations for the Calibration and the Prediction Sets Obtained with the Four Multivariate Calibration Methods

a.) CLS, b.) GCLS, c.) GILS and d.) GR method.

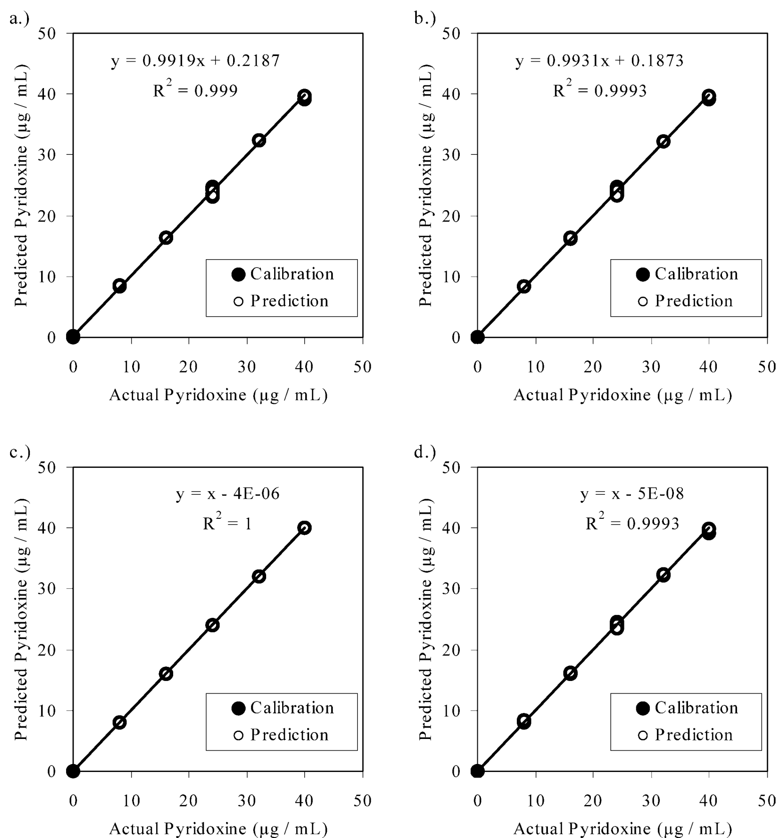

Fig. 3. Plots of Actual *vs.* Predicted Pyridoxine Concentrations for the Calibration and the Prediction Sets Obtained with the Four Multivariate Calibration Methods

a.) CLS, b.) GCLS, c.) GILS and d.) GR method.

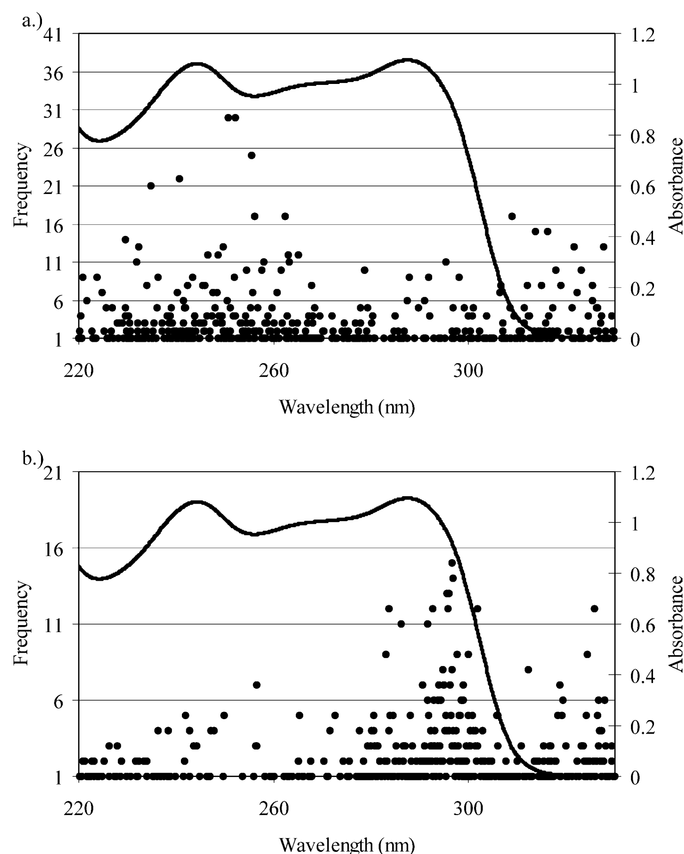

Fig. 4. Distribution of the Selected Wavelengths by Genetic Algorithm in GR Method for a Total of 50 Runs with 20 Genes and 100 Iterations in Each Run a.) Thiamine and b.) pyridoxine.

ods in the second experiment. The R square (R^2) values of regression were ranged between 0.9990 and 1.0000 indicating very good fit between actual and predicted concentrations.

In order to determine whether the genetic algorithm selected wavelengths corresponds the particular component absorbance region, the GR method was also set to run 50 times with 20 genes and 100 iterations in the second experiment. The overall distribution of the selected wavelengths for each component along with a mixture spectrum are shown in Fig. 4 for thiamine and pyridoxine, respectively. As can be seen from the figure, the genetic regression method select the wavelengths that correspond the each component absorption range even though the algorithm starts with the whole spectrum information at the beginning of each run and each wavelength has equal chance of being selected. The explanation of this is in the evolutionary nature of genetic algorithm where the wavelengths that are suited for the particular component survives in the long run of iterations and other do not. This gives an advantage to the genetic algorithm based methods where only the information related to the particular component are used to construct the model thereby reducing the noise in the overall information.

Conclusion

This study illustrates the application of the hard modelling technique CLS and three genetic algorithm based multivariate calibration methods to simultaneous determination of pharmaceuticals in synthetic and actual tablet formulations. It can be said that all four method generate acceptable results in the given concentration range of the components. These methods coupled with spectrophotometry could be an alternative to other methods such as chromatography, which is more expensive and time consuming.

References

1) Ortega Barrales P., DomínguezVidal A., Fernández de Córdova M. L., Molina Díaz A., *J. Pharm. Biomed. Anal.*, **25**, 619—630 (2001).
2) Dinc E., Baleanu D., *J. Pharm. Biomed. Anal.*, **31**, 969—978 (2003).
3) Dinc E., Baleanu D., *Talanta*, **59**, 707—717 (2003).
4) Dinc E., Kökdil G., Onur F., *J. Pharm. Biomed. Anal.*, **22**, 915—923 (2000).
5) García L., Blázquez S., San Andrés M. P., Vera S., *Anal. Chim. Acta*, **434**, 193—199 (2001).
6) Fotsing L., Fillet M., Bechet I., Hubert Ph., Crommen J., *J. Pharm. Biomed. Anal.*, **15**, 1113—1123 (1997).
7) Halvatzis S. A., Timotheou-Potamia M., *Anal. Chim. Acta*, **227**, 405—419 (1989).
8) Haaland D. M., Thomas E. V., *Anal. Chem.*, **60**, 1193—1202 (1988).
9) Geladi P., Kowalski B. R., *Anal. Chim. Acta*, **185**, 1—17 (1986).
10) Wentzell P. D., Andrews D. T., Kowalski B. R., *Anal. Chem.*, **69**, 2299—2311 (1997).
11) Esbensen K., Geladi P., Wold S., *Chem. Intell. Lab. Syst.*, **2**, 37—52 (1987).
12) Özdemir D., Williams R. R., *Appl. Spectrosc.*, **53**, 210—217 (1999).
13) Mosley R. M., Williams R. R., *Appl. Spectrosc.*, **52**, 1197—1202 (1998).
14) Özdemir D., Mosley R. M., Williams R. R., *Appl. Spectrosc.*, **52**, 599—603 (1998).
15) Özdemir D., Mosley R. M., Williams R. R., *Appl. Spectrosc.*, **52**, 1203—1209 (1998).
16) Paradkar R. P., Williams R. R., *Appl. Spectrosc.*, **51**, 92—100 (1997).
17) Hörchner U., Kalivas J. H., *Anal. Chim. Acta*, **311**, 1—13 (1995).

18) Lucasius C. B., Kateman G., *Chem. Intell. Lab. Syst.*, **19**, 1—33 (1993).
19) Holland J. H., "Adaptation in Natural and Artificial Systems," MIT Press, Cambridge, 1992.
20) Lawerence D., "Handbook of Genetic Algorithms," Van Nostrand Reinhold, New York, 1991.