



## Attenuated Total Reflectance with Fourier Transform Infrared Spectroscopy (ATR/FTIR) and Different PLS Algorithms for Simultaneous Determination of Clavulanic Acid and Amoxicillin in Powder Pharmaceutical Formulation

Aline L. H. Müller,<sup>a</sup> Érico M. M. Flores,<sup>a,d</sup> Edson I. Müller,<sup>a,d</sup> Fabiana E. B. Silva<sup>b</sup> and Marco F. Ferrão<sup>\*,c,d</sup>

<sup>a</sup>Departamento de Química, Universidade Federal de Santa Maria, 97105-900 Santa Maria-RS, Brazil

<sup>b</sup>Departamento de Farmácia, Universidade Federal do Pampa, 97500-970 Uruguaiana-RS, Brazil

<sup>c</sup>Departamento de Química Inorgânica, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, 91501-970 Porto Alegre-RS, Brazil

<sup>d</sup>Instituto Nacional de Ciência e Tecnologia de Bioanalítica, 13083-970 Campinas-SP, Brazil

Um método para determinação simultânea de ácido clavulânico (CA) e amoxicilina (AMO) foi desenvolvido utilizando a técnica de espectroscopia no infravermelho médio com transformada de Fourier acoplada ao acessório de reflexão total atenuada (ATR/FTIR). Utilizaram-se 27 e 8 amostras para os conjuntos de calibração e previsão, respectivamente. Os modelos de calibração foram desenvolvidos utilizando os algoritmos por mínimos quadrados parciais (PLS), PLS por intervalo (iPLS), PLS por sinergismo (siPLS) e PLS por exclusão (biPLS). Os melhores modelos foram aqueles que utilizaram o algoritmo biPLS. Obtiveram-se erro padrão relativo de previsão (RSEP) de 3,8 e 5,1% para CA e AMO, respectivamente. Os resultados obtidos pela metodologia proposta foram comparados com os obtidos por cromatografia líquida de alta eficiência (HPLC) e nenhuma diferença significativa foi observada. O método proposto utilizando ATR/FTIR associado a métodos de análise multivariados foi satisfatório para determinação de CA e AMO em produtos farmacêuticos.

A method for simultaneous determination of clavulanic acid (CA) and amoxicillin (AMO) was developed using Fourier transform mid infrared technique coupled with attenuated total reflectance (ATR/FTIR) accessory. 27 samples were used as calibration set and 8 samples were used for prediction set. Calibration models were developed using partial least squares (PLS), interval PLS (iPLS), synergy PLS (siPLS) and backward PLS (biPLS). Multiplicative scatter correction and the mean centering were used and produced the best models. Relative standard error of prediction (RSEP) of 3.8% for CA and 5.1% for AMO were obtained using biPLS algorithm for ATR/FTIR data. Results obtained by the proposed methodology were compared with those using high performance liquid chromatography (HPLC) and no significant differences were obtained. The proposed method using ATR/FTIR combined to multivariate analysis methods was suitable for the simultaneous determination of CA and AMO in commercial pharmaceutical products.

**Keywords:** clavulanic acid, amoxicillin, ATR/FTIR, pharmaceutical products, backward interval PLS

### Introduction

Amoxicillin ((2*S*, 5*R*, 6*R*)-6-[(*R*)-(-)-2-amino-2-(*p*-hydroxyphenyl) acetamido]-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo [3.2.0] heptane-2-carboxylic acid trihydrate, C<sub>16</sub>H<sub>19</sub>N<sub>3</sub>O<sub>5</sub>·3H<sub>2</sub>O, MW 419.46) is an analog

of ampicillin, derived from the basic penicillin nucleus, 6-aminopenicillanic acid. Clavulanic acid (potassium (*Z*)-(2*R*,5*R*)-3-(2-hydroxyethylidene)-7-oxo-4-oxa-1-azabicyclo[3.2.0]-heptane-2-carboxylate, C<sub>8</sub>H<sub>8</sub>KNO<sub>5</sub>) is originally produced by fermentation of *Streptomyces clavuligerus*. It inactivates a wide variety of β-lactamases by blocking the active sites of these enzymes. Amoxicillin is commonly prescribed in association with clavulanic acid

\*e-mail: ferrao@iq.ufrgs.br

for the treatment of infection caused by bacteria that are resistant to amoxicillin alone.<sup>1</sup> The main analytical methods that have been reported for the determination of these two drugs in biological fluids are high performance liquid chromatography (HPLC),<sup>2-5</sup> capillary electrophoresis<sup>6-9</sup> and luminescence spectroscopy.<sup>10</sup> For the pharmaceutical preparations of CA and AMO contents, the pharmacopoeias use HPLC technique as official assay for the procedure of quality control.<sup>11</sup> However, this technique can present some drawbacks related to the throughput as consumption and generation of chemical residues and excessive time for analysis.

In this way, alternative techniques for drugs quality control that use measurements in real time, minimum sample preparation and free-solvent methods have been proposed in last years.<sup>12-17</sup> The use of Fourier transform infrared spectroscopy can supply with these characteristics, mainly when associated to reflection techniques, such as attenuated reflection. Attenuated total reflectance/Fourier transform infrared spectroscopy (ATR/FTIR) has been suggested as a sensible technique to analyze pharmaceutical formulations such as solids, semi-solids and liquids.<sup>18</sup>

Infrared spectroscopy combined with multivariate analysis allows the improvement of the quality of the results obtained for complex mixtures by overcoming problems related to overlapped signals. Partial least-squares (PLS) regression is the most popular algorithm used for multivariate calibration for quantitative analysis and allows minimizing problems such as the loss of resolution in the analytic signal.<sup>19,20</sup> Generally, PLS performs the calibration using information from the full spectrum to build a regression model to determine the property of interest, called full-spectrum method. However, recent applications have been proposed using methods for selection spectral region with suitable algorithms to improve the performance of PLS regression.<sup>21-25</sup> In the case of FTIR, specific regions are selected and there are production of models with lower prediction error. In practice, these methods are based on the identification of a subset of complete data that will produce the lowest prediction error. An optimized region can be found by reducing or increasing or by subtracting or adding new variables.<sup>21</sup>

Interval PLS model (iPLS) is based on the division of the spectrum into smaller intervals followed by the construction of a PLS regression model for each interval. The root mean square error of cross-validation (RMSECV) is calculated for each interval and compared with the value obtained for the full spectrum model. Regions that present the smallest value of RMSECV is then chosen.<sup>26</sup> One of the main advantages of this method is the possibility of representing a local regression model in a graphical display, focusing on a choice

of better intervals allowing a comparison among interval models and the full spectrum model.<sup>21</sup>

Synergic interval PLS model (siPLS) is an evolution of the iPLS algorithm described by Munck *et al.*<sup>27</sup> Synergic iPLS algorithm also splits the spectrum in a number of intervals (variables-wise) and develops PLS regression models for all possible combinations of two, three or four intervals.

For backward interval PLS (biPLS) algorithm, as well as for iPLS algorithm, the spectrum is split into a given number of intervals.<sup>28</sup> Finally, PLS models are calculated with each interval left out. The first left out interval is the one that when left out gives the poorest performing model with respect to RMSECV, i.e. each time excluding the interval whose removal results in the lowest RMSECV. The procedure is continued until the last interval or can be stopped when the number of retained wavelengths is lower than a predefined threshold.

In this study, the main objective was to investigate the feasibility of ATR/FTIR spectroscopy associated to iPLS, siPLS and biPLS algorithms to obtain regression models for simultaneous determination of CA and AMO in pharmaceutical formulations, as an alternative to official method (HPLC). Partial least square was employed to data modeling using full spectra information, while iPLS, siPLS and biPLS were used to select variable intervals. Results obtained by the proposed method were compared with those obtained by recommended procedures by pharmacopoeias.

## Experimental

### *Materials and sample preparation*

Clavulanic acid standard (in lithium salt form) was acquired from United States Pharmacopoeia (USP), and amoxicillin certified reference material from Brazilian Pharmacopoeia. Methanol and monobasic sodium phosphate were HPLC grade and purchased from Merck (Darmstadt, Germany). Clavulanic acid (as potassium clavulanate) and amoxicillin trihydrate pharmaceutical grade were purchased from manipulation drugstore and used for the preparation of synthetic samples. A total of 16 formulations (synthetic samples) containing CA, AMO and excipients mixture (78% of microcrystalline cellulose, 20% of talc, 1% of colloidal silicon dioxide and 1% of magnesium stearate) were prepared in laboratory. 19 commercial tablet formulations from different laboratories were acquired in local drugstore. The concentration of these samples was previously determined by HPLC. Thus, samples were divided in calibration and prediction subsets (Table 1). Calibration set was constructed

with 12 synthetic samples and 15 commercial samples and prediction set was constructed with 4 synthetic samples and 4 commercial samples. Synthetic and commercial samples were ground and mixed in a cryogenic mill Spex Certiprep (model 6750 Freezer Mill, Metuchen, EUA). A time period of 2 min was enough to mixing each samples, that was ground up to particle size less than 80  $\mu\text{m}$ .

**Table 1.** Samples used in the calibration and prediction set

Set	SN <sup>a</sup>	Minimum	Maximum
Calibration	27		
AC / (mg g <sup>-1</sup> )		69.4	185.3
AMO / (mg g <sup>-1</sup> )		375.7	737.2
Prediction	8		
AC / (mg g <sup>-1</sup> )		80.5	167.6
AMO / (mg g <sup>-1</sup> )		387.5	686.1

<sup>a</sup>SN: number of samples.

### Spectra collection

Spectra were collected in the range from 4000 to 650  $\text{cm}^{-1}$  using a PerkinElmer Model Spectrum One FTIR spectrometer with deuterated triglycine sulfate (DTGS) detector and KBr beam splitter. This instrument is equipped with an attenuated total reflection (ATR) sampling universal accessory supplied with a top plate for ZnSe crystal. For the ATR spectrum acquisition, each sample was previously weighted before spectrum acquisition (about  $25 \pm 5$  mg of powder sample) and placed onto the ATR crystal. Crystal was cleaned with acetone after acquisition of each spectrum. Spectra were obtained in triplicate with 16 scans, 4  $\text{cm}^{-1}$  resolution and normalized. Spectra of the CA and AMO were normalized with an ordinate limit of up to 1.0 of absorbance using the tool available in the software of the spectrometer (Spectrum, 5.01 version, Perkin-Elmer, 2003), and finally the medium spectrum was obtained.

### Softwares

Hierarchical cluster analysis (HCA) algorithm from Pirouette<sup>®</sup> 3.11 software (<http://www.infometrix.com> EUA) was used to divide the samples set in calibration and prediction samples. Data were handled using Matlab<sup>®</sup> software 6.5 version (The Math Works, Natick, USA) using PLS multivariate calibration models from PLS Toolbox 2.0 version. The iToolbox for Matlab<sup>®</sup> (<http://www.models.kvl.dk> USA) was used to variable selections and develop multivariate models.<sup>26</sup> For evaluation of the models obtained from iPLS, siPLS and biPLS algorithms, spectrum was divided in 10, 20, 30, 40 and 50 intervals. Multiplicative scatter correction (MSC)

pretreatment and autoscaling (A) and mean centering (MC) preprocessing were applied. F test (95% confidence level) was used to compare the prediction errors of constructed models. Comparison between results obtained by the proposed methodology and reference methodology was performed using *t*-test paired (95% confidence level).

### HPLC reference method

The reference method used for the determination of CA and AMO in pharmaceutical formulations was HPLC that is recommended by United States Pharmacopoeia, USP 31.<sup>11</sup> High performance liquid chromatography system used was a Agilent 1100 Series system equipped with pump (model G1311A), detector (model G133A ALS) and loop sampling of 20  $\mu\text{L}$ . Detector was set at 220 nm and peak areas were integrated automatically using a Chemstation<sup>®</sup> software program (Agilent Technologies Inc. CA, USA). Separation was carried out at ambient temperature using a Zorbax<sup>®</sup> SBC-18 column (250 mm  $\times$  4.5 mm i.d., 5  $\mu\text{m}$  particle size) and guard cartridge system was used to safeguard the analytical column. Mobile phase was a mixture of methanol:sodium phosphate buffer (5:95 v/v). Commercial and synthetic samples were dissolved in water with the aid of mechanical stirring, transferred to a suitable volumetric flask in order to obtain a solution containing about 0.5  $\text{mg mL}^{-1}$  of amoxicillin in agreement with USP 31. All these determinations were carried out in triplicate for synthetic and commercial samples.

### Chemometric models

PLS algorithm was applied to ATR/FTIR data to built calibration models for CA and AMO quantification in pharmaceutical preparation. For variable selection iPLS, siPLS and biPLS algorithms were used. Interval PLS models were built with spectra division into 10, 20, 30, 40 and 50 intervals. Synergic interval PLS and biPLS models were also built with spectra division into 10, 20, 30, 40 and 50 intervals and combinations of 2 and 3 intervals were used for siPLS models.

Root mean square error was calculated according to the following equation 1:<sup>20</sup>

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where  $\hat{y}_i$  is the prediction value for test set sample  $i$ ,  $y_i$  the reference value for test set sample  $i$  and  $n$  is the number of observation in the test set (for cross-validation was used  $n - 1$ ).

Root mean square error of cross-validation (RMSECV)<sup>29</sup> was used to select the number of latent variables and variable selections. Root mean square error of calibration (RMSEC) was employed to evaluate the error of the proposed calibration models and root mean square error of prediction (RMSEP) was used to evaluate the prediction ability of different PLS models and used to select the best model.

The performance of the calibration models was calculated using relative standard error of prediction (RSEP), using the following equation 2:<sup>30-32</sup>

$$\text{RSEP} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i)^2}} 100 \quad (2)$$

where  $\hat{y}_i$  is the prediction value for test set sample  $i$ ,  $y_i$  the reference value for test set sample  $i$ .

Correlation coefficients between the prediction and measured values were calculated for the calibration set, which were calculated as equation 3, where  $\bar{y}_i$  is the mean of the reference measurement results for all samples in the training set.<sup>22</sup>

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (3)$$

Systematic error (bias) and standard deviation of validation (SDV) were calculated from equations 3 and 4, respectively:

$$\text{bias} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (4)$$

$$\text{SDV} = \sqrt{\frac{\sum_{i=1}^n [(y_i - \hat{y}_i) - \text{bias}]^2}{n - 1}} \quad (5)$$

Thereafter, the  $t$ -test was applied, according to the following equation 6:<sup>33</sup>

$$t_{\text{sis}} = \frac{|\text{bias}| \sqrt{n}}{\text{SDV}} \quad (6)$$

Systematic error was considered not significant for  $t$  systematic ( $t_{\text{sis}}$ ) values lower than critical value ( $t_{\text{crit}}\alpha = 0.5\%$  and  $n-1$  degree of freedom).

Accuracy was calculated for the proposed ATR/FTIR method and reports the agreement between the reference

value obtained by HPLC and the prediction value obtained by the calibration model (in this case evaluated by RMSEP value).

Results obtained by the proposed method were compared with reference method (using HPLC) by  $t$ -test paired ( $\alpha = 0.5\%$ ). The results obtained by ATR/FTIR were also in agreement to the interval allowed by United States Pharmacopoeia (90-120% declared value).<sup>11</sup>

## Results and Discussion

### Treatment data

PLS models were obtained using full spectrum and MSC pre-treatment, autoscaling (A) and mean center data (MC) preprocessing. MSC pre-treatment showed the lower RMSEP values and better correlation coefficient. No significant difference was observed for RMSEP values when A and MC were used. MC preprocessing combined with MSC pretreatment were chosen because showed the lower RMSEP value.

### Results of PLS models

Initially, models were built using ATR/FTIR full-spectrum data. For PLS algorithm, the number of latent variables is a critical parameter in calibration model. The optimum number of latent variables is determined by the lowest root mean square error cross-validation (RMSECV). The full-spectrum PLS models that showed the lowest RMSECV used eight and seven latent variables for CA and AMO, respectively. These models showed RMSEP of 18.05 and 39.60 mg g<sup>-1</sup> for CA and AMO, respectively (Tables 2 and 3).

### Results of clavulanic acid iPLS models

Spectral interval selection was first carried out by iPLS. Spectra were split into smaller equidistant regions and models were developed for each subinterval. The RMSECV values obtained for each subinterval were compared with full-spectrum RMSECV value.<sup>26</sup> Interval number 19 showed the minor RMSECV values when compared with the others 19 intervals.

In Table 2, it is possible to compare the prediction performance of all models obtained for CA using ATR. Root mean square error of the prediction value obtained using iPLS algorithm did not show significant difference when compared with PLS full-spectrum ( $F$  test, 95% confidence level). The best iPLS model for CA determination used the interval number 19 (spectra divided in 20 intervals), 3 latent

**Table 2.** Statistical results for the CA better calibration models and CA full-spectrum PLS model. The best models iPLS, siPLS and biPLS are showed in bold

Model	VN <sup>a</sup>	Intervals	LVs <sup>b</sup>	R <sub>cal</sub>	RMSEP / (mg g <sup>-1</sup> )
PLS	3351	full	8	0.980	18.05
iPLS10	335	9	3	0.915	24.58
<b>iPLS20</b>	<b>168</b>	<b>19</b>	<b>3</b>	<b>0.929</b>	<b>11.05</b>
iPLS30	112	30	5	0.975	21.86
iPLS40	84	40	6	0.978	20.94
iPLS50	66	43	2	0.900	18.76
si2PLS10	670	4, 10	5	0.974	13.33
si3PLS10	1005	3, 4, 7	6	0.961	35.97
si2PLS20	336	2, 20	7	0.990	23.56
si3PLS20	504	2, 9, 20	9	0.992	25.20
si2PLS30	224	3, 30	7	0.993	27.94
si3PLS30	336	6, 10, 21	5	0.979	34.34
<b>si2PLS40</b>	<b>168</b>	<b>13, 29</b>	<b>4</b>	<b>0.965</b>	<b>12.21</b>
si3PLS40	252	4, 39, 40	7	0.993	22.67
si2PLS50	132	17, 44	5	0.958	17.58
biPLS10	1675	2, 4, 5, 6, 7	6	0.967	29.03
biPLS20	671	2, 7, 9, 18	11	0.990	24.90
<b>biPLS30</b>	<b>446</b>	<b>10, 21, 23, 27</b>	<b>8</b>	<b>0.967</b>	<b>5.09</b>
biPLS40	503	6, 13, 18, 19, 28, 36	8	0.980	29.39
biPLS50	469	5, 6, 8, 17, 23, 35, 45	7	0.988	23.50

<sup>a</sup>VN: total variable numbers; <sup>b</sup>LV: latent variables.

**Table 3.** Statistical results for the AMO better calibration models and AMO full-spectrum PLS model. The best models iPLS, siPLS and biPLS are showed in bold

Model	VN <sup>a</sup>	Intervals	LVs <sup>b</sup>	R <sub>cal</sub>	RMSEP / (mg g <sup>-1</sup> )
PLS	3351	full	7	0.967	39.60
iPLS10	335	9	3	0.897	49.98
iPLS20	168	16	4	0.919	57.78
iPLS30	112	26	5	0.909	55.91
iPLS40	84	31	5	0.932	49.16
iPLS50	66	41	4	0.824	50.58
s2iPLS10	670	4, 10	6	0.954	36.62
s3iPLS10	1005	3, 4, 9	4	0.927	43.16
s2iPLS20	336	7, 11	8	0.997	83.60
s3iPLS20	504	7, 8, 11	7	0.992	74.87
<b>s2iPLS30</b>	<b>224</b>	<b>11, 26</b>	<b>7</b>	<b>0.966</b>	<b>35.14</b>
s3iPLS30	336	10, 11, 16	10	0.999	79.79
s2iPLS40	168	22, 31	10	0.995	54.56
s3iPLS40	252	21, 22, 31	12	0.999	54.14
s2iPLS50	132	12, 43	8	0.968	55.69
biPLS10	1005	3, 4, 9	4	0.927	43.16
biPLS20	503	4, 7, 17	4	0.938	44.57
biPLS30	557	1, 21, 23, 25, 26	10	0.985	33.86
biPLS40	418	22, 28, 31, 34, 38	8	0.982	38.28
<b>biPLS50</b>	<b>268</b>	<b>12, 35, 38, 42</b>	<b>8</b>	<b>0.970</b>	<b>27.22</b>

<sup>a</sup>VN: total variable numbers; <sup>b</sup>LV: latent variables.

variables, 168 variable numbers and showed RMSEP equal to 11.05 mg g<sup>-1</sup>. Therefore, iPLS models can reduce noise by selecting specific spectral regions. However, useful spectral information can be lost.

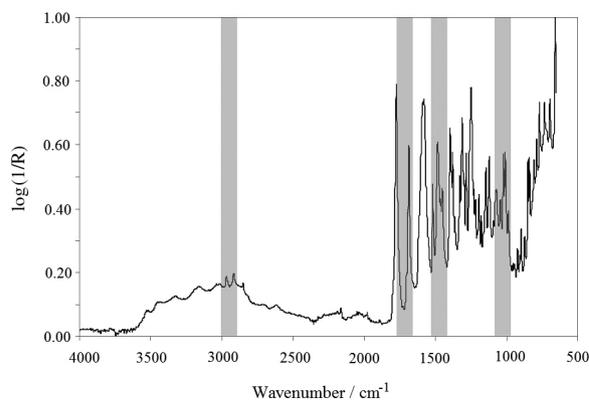
#### *Results of clavulanic acid siPLS model*

The operation of synergy interval PLS algorithm was similar to the one used for iPLS. First, spectra were split into

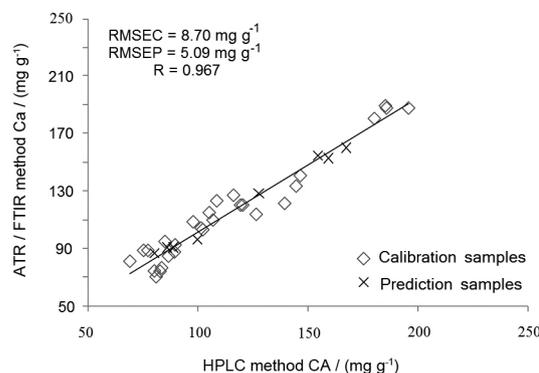
a number of intervals (variable-wise) and PLS regression model for all possible combinations of two and three were obtained. Root mean square error cross-validation was then calculated for each combination of intervals. Combination of intervals with the lowest RMSECV was chosen and compared with full-spectrum RMSECV values.<sup>26</sup> Table 2 shows the statistical results for CA obtained using siPLS. The best models using siPLS algorithm were those that divided the spectra in 10, 40 and 50 intervals and used combinations of 2 and 3 intervals. The lower RMSEP value for siPLS was obtained when the spectra were split in 40 intervals and the interval number 13 and 29 were combined (Table 2 in bold). Comparison between RMSEP value of the best siPLS model and PLS full-spectrum RMSEP value did not show significant difference (*F* test, 95% confidence level). Then, siPLS model was not suitable for CA determination in pharmaceutical formulations because did not show good prediction ability.

#### Results of clavulanic acid biPLS models

Backward interval PLS algorithm used in this work was previously developed by Leardi *et al.*<sup>28</sup> As in the iPLS model, the spectra were split in a selected number of intervals. However, PLS models are calculated with each interval left out, i.e. if one chooses 30 intervals then each model is based on 29 intervals, leaving out one interval at a time. The first left out interval is the one that when left out gives the poorest performing model with respect to RMSECV. This procedure was continued until one interval remains. Table 2 shows the statistical results for best CA models using biPLS (Table 2 in bold). It can be observed that the best biPLS model was that which divided the spectrum in 30 intervals and it selected the interval numbers 10, 21, 23 and 27 (Figure 1). This model showed equivalent correlation coefficient, RMSEP three times lower and reduction of 87% of the variable numbers when compared with PLS full-spectrum model. Backward interval PLS model showed results with good correlation between reference and prediction values as shown in Figure 2. The selected intervals included the regions of 2890 to 3002  $\text{cm}^{-1}$  (interval 10) that corresponds to  $\text{CH}_2$  symmetric stretching, 1658 to 1770  $\text{cm}^{-1}$  (interval 21) that corresponds to C=C stretching, 1434 to 1546  $\text{cm}^{-1}$  (interval 23) that corresponds to =C-H out-of-plane bending of vibration (C=C) and 986 to 1098  $\text{cm}^{-1}$  that corresponds to =C-H in-plane bending of vibration (C=C) (interval 27).<sup>34</sup> These groups are constituents of CA chemical structure. In a general way, the elimination of intervals that did not present correlation with CA chemical structure allowed the reduction of RMSEP values. The comparison between the



**Figure 1.** Spectral region selected by biPLS algorithm for the determination of CA.



**Figure 2.** Values obtained by reference method *versus* values obtained by biPLS model for the CA determination using intervals 10, 21, 23 and 27, and 8 latent variables.

results obtained for CA using the best biPLS calibration model and results obtained by reference method are shown in Table 4. Backward interval PLS model developed using intervals 10, 21, 23 and 27 resulted in low relative standard error of prediction (RSEP = 3.8%), suggesting that the used method is accurate. The systematic error provided by model was not significant (bias = 0.5176 and  $t_{\text{sis}} < t_{\text{crit}}$ ) due to the non-tendency for the prediction errors. Backward interval PLS model was the most suitable for CA determination in pharmaceutical formulations because showed better prediction ability (lower RMSEP value).

#### Results of amoxicillin iPLS models

In the same way as for the CA determination, the models for AMO were obtained using iPLS algorithm. In Table 3, it is possible to compare the prediction performance of all models obtained for AMO using ATR. The best iPLS model was that used spectra divided in 30 intervals, 5 latent variables and selected the interval number 26 (Table 3 in bold). Root mean square error of prediction values obtained for iPLS models and PLS full-spectrum model

**Table 4.** Results obtained for the prediction samples using biPLS algorithm (mg g<sup>-1</sup>)

Samples	Clavulanic acid		Amoxicillin	
	HPLC method / (mg g <sup>-1</sup> )	ATR/FTIR method / (mg g <sup>-1</sup> )	HPLC method / (mg g <sup>-1</sup> )	ATR/FTIR method / (mg g <sup>-1</sup> )
1 <sup>a</sup>	89.02	91.05	520.97	552.74
2 <sup>a</sup>	80.45	87.20	509.27	496.52
3 <sup>a</sup>	100.02	96.57	387.54	404.91
4 <sup>a</sup>	86.14	90.61	520.17	510.96
5 <sup>b</sup>	127.66	128.60	686.07	636.71
6 <sup>b</sup>	167.59	160.15	580.70	576.31
7 <sup>b</sup>	154.61	153.90	493.63	536.78
8 <sup>b</sup>	159.38	152.50	473.16	465.96

<sup>a</sup>Synthetic samples; <sup>b</sup>commercial samples.

were compared and significant difference was not obtained (*F* test, 95% confidence level). It is possible because the information was spread on the whole spectral range and a variable selection *per* interval could automatically reduce the information and induce an increase of RMSEP compared with PLS full-spectrum model.<sup>24</sup>

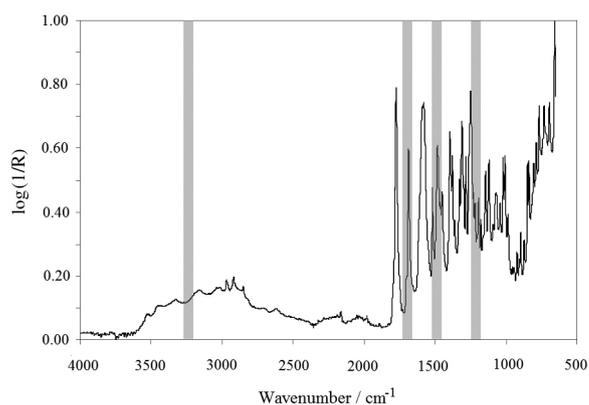
#### Results of amoxicillin siPLS models

For determination of AMO in pharmaceutical formulation, siPLS algorithm was also used. Table 3 shows the statistical results for AMO obtained with use of siPLS algorithm. The best models using siPLS algorithm were that divided the spectra in 10 and 30 intervals using combinations of 2 intervals (Table 3 in bold). The lower RMSEP was obtained when the spectra were split in 30 intervals and the interval numbers 11 and 26 were combined. However, significant difference was not obtained when compared with PLS full-spectrum RMSEP value (*F* test, 95% confidence level).

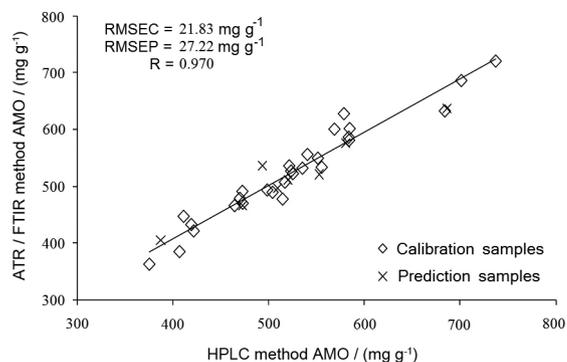
#### Results of amoxicillin biPLS models

In Table 3, it is possible to observe the statistical results for best models obtained using biPLS algorithm (Table 3 in bold). The best biPLS model was that which divided spectra in 50 intervals and used interval numbers 12, 35, 38 and 42 (Figure 3). It showed significant reduction of RMSEP (*F* test, 95% confidence level) and used about 92% less variables when compared with PLS full-spectrum model. This biPLS model showed results with good correlation between reference and prediction values (correlation coefficient of 0.970) as shown in Figure 4. The selected intervals included the regions of 3197 to 3254 cm<sup>-1</sup> (interval 12) that corresponds to O–H stretching of the phenol, 1656 to 1723 cm<sup>-1</sup> (interval 35) that corresponds to NH<sub>2</sub> in-plane bending of vibration, 1455 to 1522 cm<sup>-1</sup> (interval 38) that corresponds to bending of

vibration N–H of the secondary amine or C=C stretching of the aromatic and 1187 to 1254 cm<sup>-1</sup> that corresponds to C–N stretching of the primary amine (interval 42).<sup>34</sup> All these groups are present in the AMO chemical structure. In a general way, the elimination of intervals that do not contain relevant information of chemical structure reduced RMSEP values. Comparison between results obtained using the best biPLS model and results obtained by reference method are shown in Table 4. Backward interval PLS model



**Figure 3.** Spectral region selected by biPLS algorithm for determination of the AMO.



**Figure 4.** Values obtained by reference method *versus* values obtained by biPLS model for the AMO determination using intervals 12, 35, 38 and 42, and 8 latent variables.

developed used interval numbers 12, 35, 38 and 42 and showed low relative standard error of prediction (RSEP = 5.12%), suggesting that the proposed method is accurate. Systematic error calculated for the model was not significant (bias = 2.80 and  $t_{\text{sis}} < t_{\text{crit}}$ ) due to the non-tendency for the prediction errors. Backward interval PLS model was the most suitable for AMO determination in pharmaceutical formulations because it showed good prediction ability (lower RMSEP value) when compared with the models PLS, iPLS and siPLS.

### Evaluation of the proposed methods

The proposed methods were validated using some parameters usually recommended by different guidelines: accuracy, precision or repeatability and linearity.<sup>35</sup>

### Accuracy

Accuracy was evaluated by comparison of results obtained by proposed method and reference method (HPLC) for samples not used in calibration model (prediction samples).<sup>36</sup> Paired *t*-test was performed between the results obtained by proposed and reference method (Table 4) to establish the accuracy of the ATR/FTIR combined multivariate analysis method. This *t*-test allows checking if the results between the methods were significantly different. The acceptance criterion is defined by the formula:  $t_{\text{exp}} \leq t_{\text{tab}}$  (where  $t_{\text{exp}}$  is *t* experimental and  $t_{\text{tab}}$  is *t* tabulated) and the results for  $t_{\text{tab}}$  showed value of 2.36 for *t*-test at 95% of confidence and seven degrees of freedom. In the CA determination, it was obtained  $t_{\text{exp}} = 0.29$  and in the AMO determination  $t_{\text{exp}} = 0.11$ . The proposed method did not showed significant differences when compared with reference method. Thus, the proposed method could substitute the reference method in the quality control of formulations in study.

### Precision

The parameter precision was evaluated by repeatability or intra-day precision. Repeatability was calculated using results from the same analyst on the same day. In this work the repeatability was calculated by three replicates of three different concentrations, using the following equation 7:

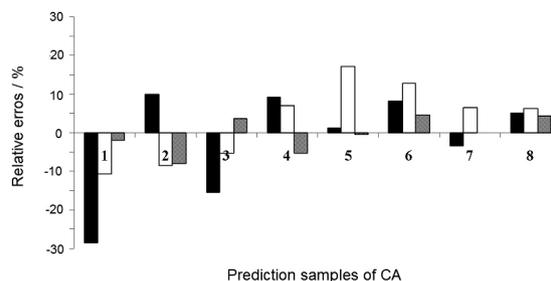
$$\text{precision} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - \hat{\bar{y}}_i)^2}{n(m-1)}} \quad (7)$$

where *m* is the realized replicate number, *n* the sample number,  $\hat{y}_i$  is the prediction mean values of each replicate  $\hat{y}_{ij}$ .<sup>37</sup> Thus, values of 5.5 mg g<sup>-1</sup> for CA and 19.8 mg g<sup>-1</sup>

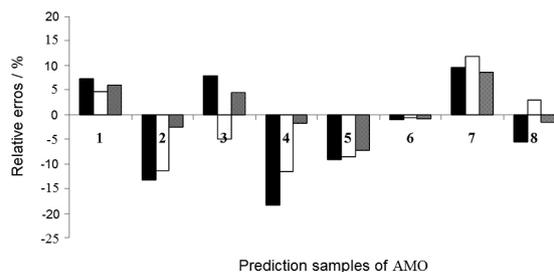
for AMO were obtained and relative standard deviations (RSD%) were of 4.2 and 3.7% respectively. The RSD values obtained were lower than the widely accepted 5% for this type of determination.<sup>14</sup>

### Discussion of the results

Comparing the results obtained by PLS, iPLS, siPLS and biPLS models for simultaneous determination of the CA and AMO in powder pharmaceutical formulations, the biPLS model showed better prediction capacity (lower RMSEP) and relative error (%) (Figures 5 and 6). Such fact could be explained by the following reasons: PLS full-spectrum models were performed with full spectrum region (3351 variable numbers), thus regions did not correlate with CA and AMO were used and high RMSEP values were obtained. Interval PLS algorithm divided spectra in intervals and this procedure can reduce noise by selecting specific spectral information, but only one interval cannot be enough to show useful spectral information. Using the siPLS algorithm, it was possible to combine 2 and 3 intervals, then completing spectral information that were not obtained in iPLS models. However, the combination of number of intervals higher than 3 intervals was not used because requires a greater computational time. For biPLS algorithm, irrelevant information was excluded and only information linked with the structures of interest were



**Figure 5.** Comparison among relative error obtained for the best iPLS, siPLS and biPLS models of CA determination. Black column iPLS error (%), white column siPLS error (%) and gray column biPLS error (%).



**Figure 6.** Comparison among relative errors obtained for the best iPLS, siPLS and biPLS models of AMO determination. Black column iPLS error (%), white column siPLS error (%) and gray column biPLS error (%).

used to build model. With this algorithm, the combination of 4 intervals (best models showed RMSEP of 5.09 and 27.22 mg g<sup>-1</sup> for CA and AMO, respectively) was possible.

## Conclusions

In the present study, it was verified that ATR/FTIR spectroscopy associated multivariate analysis is a suitable method for simultaneous quantification of clavulanic acid and amoxicillin in pharmaceutical formulations. Variable selection algorithms had ability to produce better models when compared with the PLS full-spectrum model. Performance of biPLS model was better for both drugs when compared with PLS, iPLS and siPLS algorithms. The obtained recoveries were 95.8 to 107.7% for CA and 92.8 to 108.7% for AMO. These results are in agreement with range allowed to the content of CA and AMO in power mixtures according to the USP requirements (90-120%) for solid preparations. In a general way, the proposed method was characterized to be less onerous, simple, solvent free, fast and reliable, allowing potential applications to simultaneous determinations of CA and AMO in solid pharmaceutical formulations for routine analyses.

## Acknowledgments

The authors are grateful to INCT-Bioanalítica, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) for financial support and Agência Nacional de Vigilância Sanitária (ANVISA) for providing the equipments used in this work.

## References

- Sweetmenn, S. C.; Martindale, W.; *Martindale: The Complete Drug Reference*, 33<sup>rd</sup> ed., Pharmaceutical Press: London, 2003.
- Tsou, T. L.; Wu, J. R.; Young, C. D.; Wang, T. M.; *J. Pharm. Biomed. Anal.* **1997**, *15*, 1197.
- Dousa, M.; Hosmanova, R.; *J. Pharm. Biomed. Anal.* **2005**, *37*, 373.
- Aghazadeh, A.; Kazemifard, G.; *J. Pharm. Biomed. Anal.* **2001**, *25*, 325.
- Foroutan, S. M.; Zarghi, A.; Shafaati, A.; Khoddam, A.; Movahed, H.; *J. Pharm. Biomed. Anal.* **2007**, *45*, 531.
- Pajchel, G.; Pawłowski, K.; Tyski, S.; *J. Pharm. Biomed. Anal.* **2002**, *29*, 75.
- Bobrowska-Grzesik, E.; *Mikrochim. Acta* **2001**, *136*, 31.
- Pasamontes, A.; Callao, M. P.; *Anal. Chim. Acta* **2003**, *485*, 195.
- Pasamontes, A.; Callao, M. P.; *Anal. Chim. Acta* **2004**, *515*, 159.
- Peña, A. M.; Espinosa-Mansilla, A.; Valenzuela, M. I. A.; Goicoechea, H. C.; Olivieri, A. C.; *Anal. Chim. Acta* **2002**, *463*, 75.
- The United States Pharmacopeia Convention, *The Official Compendia of Standards, USP 33-NF 26*, Rockville, 2008.
- Parisotto, G.; Ferrão, M. F.; Furtado, J. C.; Molz, R. F.; *Braz. J. Pharm.* **2007**, *43*, 89.
- Qu, N.; Zhu, M.; Mi, H.; Dou, Y.; Ren, Y.; *Spectrochim. Acta, Part A* **2008**, *70*, 1146.
- Blanco, M.; Alcalá, M.; *Eur. J. Pharm. Sci.* **2006**, *27*, 280.
- Blanco, M.; Alcalá, M.; Bautista, M.; *Eur. J. Pharm. Sci.* **2008**, *3*, 409.
- Boyer, C.; Brégère, B.; Crouch, S.; Gaudin, K.; Dubost, J. P.; *J. Pharm. Biomed. Anal.* **2006**, *40*, 433.
- Sulub, Y.; LoBrutto, R.; Vivilecchia, R.; Wabuyele, B. W.; *Anal. Chim. Acta* **2008**, *611*, 143.
- Stuart, B.; *Infrared Spectroscopy: Fundamentals and Applications*, John Wiley: Chichester, 2004.
- Geladi, P.; Kowalki, B. R.; *Anal. Chim. Acta* **1986**, *185*, 1.
- Brereton, R. G.; *Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Chemical Analysis Series*, John Wiley & Sons Ltda: Chichester, 2003.
- Borin, A.; Poppi, R. J.; *Vib. Spectrosc.* **2005**, *37*, 27; Marder, L.; Tomedi, P.; Ferrão, M. F.; Jablonski, A.; Davanzo, C. U.; *J. Braz. Chem. Soc.* **2006**, *17*, 594; Aguiar, J.; Borin, A.; Poppi, R. J.; *J. Braz. Chem. Soc.* **2010**, *21*, 436.
- Chen, Q.; Zhao, J.; Liu, M.; Cai, J.; Liu, J.; *J. Pharm. Biomed. Anal.* **2008**, *46*, 568.
- Xiaobo, Z.; Jiewen, Z.; Yanxiao, L.; *Vib. Spectrosc.* **2007**, *44*, 220.
- Pataca, L. C. M.; Borges, W. N.; Marcucci, M. C.; Poppi, R. J.; *Talanta* **2007**, *71*, 1926.
- Silva, F. E. B.; Ferrão, M. F.; Parisotto, G.; Müller, E. I.; Flores, E. M. M.; *J. Pharm. Biomed. Anal.* **2009**, *49*, 800.
- Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B.; *Appl. Spectrosc.* **2000**, *54*, 413.
- Munck, L.; Nielsen, J. P.; Moller, B.; Jacobsen, S.; Sondergaard, I.; Engelsen, S. B.; Norgaard, L.; Bro, R.; *Anal. Chim. Acta* **2001**, *446*, 169.
- Leardi, R.; Norgaard, L.; *J. Chemom.* **2004**, *18*, 486.
- Wold, S.; *Technometrics* **1978**, *20*, 397.
- Blanco, M.; Coello, J.; Eustaquio, A.; Iturriaga, H.; Maspocho, S.; *Analyst* **1999**, *124*, 1089.
- Blanco, M.; Coello, J.; Iturriaga, H.; Maspocho, S.; Pagès, J.; *Anal. Chim. Acta* **1999**, *389*, 207.
- Costa, P. A. F.; *Anal. Chim. Acta* **2009**, *631*, 206.
- Annual Book of ASTM Standards, *Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-00*, ASTM International, West Conshohocken: Pennsylvania, USA, 2000.

34. Silverstein, R. M.; *Spectrometric Identification of Organic Compounds*, 6<sup>th</sup> ed., John Wiley & Sons: New York, 1998.
35. Valderrama, P.; Braga, J. W. B.; Poppi, R. J.; *J. Braz. Chem. Soc.* **2007**, *18*, 259; Rocha, W. F. C.; Rosa, A. L.; Martins, J. A.; Poppi, R. J.; *J. Braz. Chem. Soc.* **2010**, *21*, 1929.
36. International Conference on Harmonization (ICH), *ICH Harmonised Tripartite Guideline: Validation of Analytical Procedures: Text and Methodology*, Q2 (R1), USA, 2005;
37. Braga, J. W. B.; Poppi, R. J.; *Quim. Nova* **2004**, *27*, 1004.

[http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Quality/Q2\\_R1/Step4/Q2\\_R1\\_\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf).

*Submitted: February 8, 2011*

*Published online: July 28, 2011*

**FAPESP has sponsored the publication of this article.**