Article

Analysis of Seized Cocaine Samples by using Chemometric Methods and FTIR Spectroscopy

Nathália V. S. Rodrigues,[a] Eduardo M. Cardoso,[b] Marcus V. O. Andrade,[b] Cláudio L. Donnici[a] and Marcelo M. Sena[a,c]*

[a]Departamento de Química, Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte-MG, Brazil

[b]Polícia Federal, Superintendência Regional de Minas Gerais, 30430-940 Belo Horizonte-MG, Brazil

[c]Instituto Nacional de Ciência e Tecnologia em Bioanalítica (INCT Bio), 13083-970 Campinas-SP, Brazil

O objetivo deste trabalho foi desenvolver uma metodologia quimiométrica para descrever o perfil químico de amostras de cocaína apreendidas no estado de Minas Gerais (Brasil). A identificação de adulterantes e a quantificação do teor de cocaína foram feitas por cromatografia gasosa acoplada à espectrometria de massa (GC-MS). Espectros de 91 amostras foram obtidos por espectroscopia no infravermelho usando refletância total atenuada (ATR-FTIR), e utilizados na construção de um modelo de análise de componentes principais (PCA). A primeira componente principal (PC1) discriminou as amostras de maior pureza das mais diluídas/adulteradas, nas quais foram identificadas lidocaína, cafeína e benzocaína. PC2 discriminou a forma química da cocaína, cloridrato ou base. Além disso, dois modelos supervisionados discriminantes (mínimos quadrados parciais para análise discriminante, PLS-DA) foram desenvolvidos para classificar as amostras em função de sua diluição (abaixo e acima de 15% m/m) e de sua forma química, apresentando taxas de acerto que variaram entre 83 e 97%. Os modelos de classificação constituem uma ferramenta simples, rápida e não destrutiva, de grande valor para peritos forenses e investigadores criminais.

The aim of this article was to develop a chemometric methodology for determining the chemical profile of cocaine samples seized in Minas Gerais State, Brazil. The adulterant detection and the cocaine determination were performed by gas chromatography-mass spectrometry (GC-MS). Spectra of 91 samples were obtained by attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) and used to build an exploratory principal component analysis (PCA) model. The first principal component (PC1) discriminated samples of more purity from the more diluted/adulterated ones, which were characterized by the presence of lidocaine, caffeine and benzocaine. PC2 discriminated the two chemical forms of cocaine, hydrochloride and base. In addition, two supervised discriminant partial least-squares models (partial least-squares discriminant analysis, PLS-DA) were developed for classifying the samples according to dilution (above and below 15% m/m) and chemical form, with a rate of success that varied between 83 and 97%. The classification models constitute a simple, rapid and non-destructive tool, of great value for both forensic experts and criminal investigators.

Keywords: illicit drugs, MID infrared, principal component analysis, cocaine, chemometrics

Introduction

Cocaine is an alkaloid extracted from plants of the genera *Erythroxylum* (*E. novagranatense* and *E. coca*). United Nations Office on Drugs and Crime (UNODC) considers cocaine, after heroin, the second most problematic drug worldwide in terms of negative health consequences and probably the most problematic drug in terms of trafficking-related violence. Estimates suggest that 440 t of pure cocaine were consumed in the world in 2009, which would be in line with a total production estimate of 1,111 t, wholesale purity-adjusted seizures

*e-mail: marcsen@ufmg.br

of 615 t and global losses of 55 t. The value of the global market of cocaine is estimated as US$ 85 billion. Although Brazil has a relatively low prevalence rate of 0.7% of the population aged 15-64 because of its large population, it has the highest number of cocaine users (900,000) in South America. In addition, Brazil is considered one of the most important secondary distribution countries in the cocaine market, and the most prominent transit country for Europe in the Americas. In recent years, seizures of cocaine have increased significantly in Brazil, going from 8 t in 2004 to 24 t in 2009 seized only by Federal Police, which corresponds to 3% of the total of the world.[1]

Considering the aforementioned circumstances, researches with the aim of elucidating the chemical profile of seized drugs are of great relevance for determining their characteristics of origin, processing, traffic and purity. Combined with the results of criminal investigations, these data can be used for establishing connections between drug gangs and suppliers, tracing of drug distribution networks, and identifying which chemical substances should be priority of control in each region of the country. The most important information that can be obtained from seized cocaine by chemical analyses are its content, its form and the characterization of the adulterants. A general phenomenon in recent years has been the decline of cocaine purity parallel with an increase in the presence of adulterants. Adulterants are psychoactive substances used to compensate for some of the pharmacological effects of the drug lost by lower levels of purity, which can also lead to additional health problems for the users.[1,2] The most common adulterants listed by UNODC are lidocaine, benzocaine, levamisole, caffeine, procaine, boric acid, hydroxyzine and phenacetin.[1] In addition to the adulterants, seized cocaine samples may also contain diluents or cutting agents, such as lactose, glucose, carbonates/bicarbonates, silicates, starch and sulfates.[3] Another important chemical aspect is that cocaine may be trafficked and used as salt (cocaine hydrochloride) or as base (coca paste, free base, merla or crack). The base is soluble in ethyl ether, insoluble in water and volatile, making it appropriate for smoking, while the cocaine salt is soluble in water, insoluble in ether and decomposes if smoked directly, being used preferably intravenously or snorted.[4]

A promising and simple alternative to extract information directly from seized cocaine samples is the combination of vibrational spectroscopic techniques with chemometric methods. In spite of its potential, a relative small number of articles have applied chemometric methods for the analysis of illicit drugs. In these articles, the most used vibrational technique is the Raman spectroscopy, which can present the disadvantage of the interference of the fluorescent background. It has been used for the discrimination of simulated street drugs, such as cocaine, heroin and MDMA (ecstasy).[5-7] More recently, Raman spectra have also been used for quantitative analysis of simulated street drugs.[8] In addition, other two articles also employed attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) and diffuse reflectance near infrared spectroscopy (NIRS) for quantitative analysis of simulated samples of methylamphetamine and diacetylmorphine (heroin), respectively.[9,10] Nevertheless, none of these articles analyzed real seized samples. To the best of our knowledge, only two articles employed vibrational techniques jointly with chemometrics for analyzing these kind of samples. One article used NIRS to perform analysis of real seized samples, with the aim of discriminating ecstasy tablets,[11] and another article determined heroin by using this same analytical technique and partial least-squares (PLS) regression.[12] All of these articles also described the difficulties caused by the complexity of the chemical matrix due to the presence of a wide range of contaminants/adulterants, which demands the use of multivariate chemometric methods. In particular, the combined use of FTIR and principal component analysis (PCA) has proved to be a rapid and non-destructive analysis tool of great importance in criminal investigations, and has been recently used in other subjects of forensic interest, such as the discrimination of ballpoint pen inks[13] and spray paints.[14]

The aim of this article was to chemically characterize cocaine samples seized between 2008 and 2010 by the Federal Police of the Minas Gerais State (Brazil), which is the government organization responsible for international and interstate drug traffic control. Minas Gerais is the second most populous state in Brazil, with 19.6 million inhabitants, located in the Southeast region. The qualitative analysis of adulterants and the quantitative determination of cocaine were both performed by gas chromatography coupled with mass spectrometry (GC-MS). ATR-FTIR spectra of powder samples were obtained and used in a PCA model, searching for chemical similarities and pattern recognition. Finally, these spectra were used in two supervised classification models (partial least-squares discriminant analysis (PLS-DA)), which were able to discriminate cocaine samples as a function of their content and chemical form.

PCA[15,16] allows a summarized interpretation of a data set, revealing latent structures through new independent vectors called principal components (PCs). The combined analysis of score and loading plots allows finding out spectral regions related to sample groups. Scores provide the PC structure related to the samples, while loadings provide this same structure related to the variables. Loadings for spectral

data are better interpreted by observing them in separated plots, one at a time, as a function of the wavelength.

PLS[15,17] is a multivariate calibration method used for building regression models based on a latent variable (LV) decomposition. PLS-DA is a variant of PLS regression used for supervised classification. For each class to be predicted, a model is built correlating the matrix of independent variables (spectra) to a class membership vector, whose elements are called dummy variables, i.e., they have values of 1 if a sample belongs to a class and 0 otherwise. The closer each predicted element is to 1, the more likely a sample is to be a member of this class. In this research, the commonly used threshold value of 0.5 was adopted. The Kennard-Stone algorithm[18] was used to split the dataset into training and test sets. It is a uniform mapping algorithm used to select the most representative samples for the training set, based on their Euclidean distance.

As it is usual in the chemometric analysis of spectra, the data set was previously mean centered. In addition, standard normal variate (SNV) was also used as a preprocessing method aiming to eliminate multiplicative baseline deviations and other physical artifacts. This kind of non-linear deviation, also known as drift, is caused by the light scattering promoted by the diversity of the particle size of the samples, and since it is not related to the sample chemical composition, it should be corrected. SNV consists in subtracting from each data matrix element its row mean value, followed by the division by its respective row standard deviation. This is exactly equivalent to the autoscaling of the transposed data matrix. Multiplicative scatter correction (MSC) is an alternative to SNV, which corrects the baseline deviations based on a regression on the mean spectrum. Contrary to SNV, MSC is a set dependent transformation, which makes it less robust in the presence of outliers.[19] Thus, considering the large spectral diversity present in seized cocaine samples, SNV was preferred in this research. Other authors have empirically verified the best performance of SNV over MSC for the analysis of infrared and Raman spectra of forensic samples.[14]

Experimental

Samples and wet tests

Ninety one cocaine samples seized by the Federal Police in the Minas Gerais State, between January 2008 and April 2010, were analyzed. Wet tests were performed in test tubes for identifying the chemical form of cocaine (salt or base), according to the methodologies recommended by the Brazilian Federal Police,[20] UNODC[21] and Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG).[22] These tests were based on the solubility of the samples in water and ethyl ether, and included additions of concentrated solutions of NH_4OH or HCl. Solutions of 1% $AgNO_3$ and 10% $BaCl_2$ were also used for detecting chloride and sulphate, respectively. All the reagents used were of analytical grade (Merck, Darmstadt, Germany). Throughout this research, deionized water (Milli-Q) was employed.

Qualitative GC-MS analysis of adulterants

According to the recommended methodology,[20-22] an Agilent 6890N GC system with a mass selective detector (MSD) Agilent 5973 and an Agilent 7693A autosampler was used. A capillary column Agilent HP-5MS (0.25 mm × 0.25 μm × 30 m) was employed with ultra high purity helium as carrier gas at a constant flow of 0.8 mL min^{-1}. The oven temperature was initially set as 200 °C, and then increased to 300 °C. The run time was 5.00 min. The transfer line to the mass spectrometer was set to 250 °C. MSD was operated in the electron ionization (EI) mode, with electron energy at 70 eV. The MS data were recorded in the full scan mode, in the *m/z* range from 40 to 550, using a detector voltage of 1576.5 V. The solvent delay was 1.60 min. An injection volume of 1.00 μL was used in the split mode with a split ratio of 50:1.

Quantitative GC-MS analysis of cocaine

Cocaine determination was based on the recommended methods.[20-22] A cocaine base standard sample (80.91%) was provided by the Instituto Nacional de Criminalística (INC) of the Federal Police (Brasília, Brazil). A stock solution of cocaine was prepared in methanol, and from successive dilutions, five working solutions in the range between 0.10 and 0.50 mg mL^{-1} were obtained. A 3.95 mg mL^{-1} dipentyl phthalate solution, used as internal standard, was also prepared. The calibration curve was constructed in duplicate and the injections were repeated three times. Approximately 30 mg of each seized cocaine sample were dissolved in methanol and successively diluted to obtain solutions in the analytical range. The same GC-MS system described in the previous section was used. A capillary column J&W Scientific DB-5MS (0.20 mm × 0.33 μm × 25 m) was employed with ultra high purity helium as carrier gas at a constant flow of 0.8 mL min^{-1}. The oven temperature was programmed as follows: the initial temperature of 200 °C was increased at a constant rate of 40 °C min^{-1} to 300 °C, and then held for 3.75 min. The transfer line and the injector temperatures were 250 and 280 °C, respectively. The injection volume

was 0.20 µL in the splitless mode. MSD was operated in the electron ionization (EI) mode, with electron energy at 70 eV. The MS data were recorded in the selected ion monitoring (SIM) mode, selecting ions at *m/z* of 82.10 and 303.10 for cocaine and at *m/z* 149.00 for the internal standard.

FTIR measurements and data handling

Mid infrared spectroscopy measurements were conducted in a Nicolet 380 FTIR spectrometer (Thermo Fisher Scientific Inc., Madison, USA), and obtained with an attenuated total reflection ATR accessory equipped with a diamond crystal of one reflection (Smart Orbit Diamond ATR, Thermo Fisher, USA). About 1 mg of each seized cocaine sample was pulverized and placed in the accessory. Spectra were recorded from 4000 to 525 cm⁻¹, with a resolution of about 2 cm⁻¹ and 32 scans. In addition, spectra of standards of cocaine base (80.01%), cocaine hydrochloride (90.68%), lidocaine hydrochloride, caffeine, benzocaine and boric acid (all above 98.5%) were also recorded. All these standards were provided by the INC-Federal Police, Brazil. All of the sample spectra were obtained in triplicates. The mean spectra of each sample were used in all the constructed models. Data were handled using MATLAB software, version 7.9 (The MathWorks, Natick, USA) and PLS Toolbox, version 5.2 (Eigenvector Technologies, Manson, USA).

Results and Discussion

Based on the wet tests mentioned above, 66 seized samples (72.5%) were identified as cocaine base, while 25 samples (27.5%) were identified as cocaine hydrochloride. Qualitative GC-MS analysis confirmed the presence of cocaine in all the samples and detected the presence of four adulterants in 22 samples. The number of samples in which caffeine, lidocaine, benzocaine and boric acid were found is shown in Table 1. These four substances are white powders. While caffeine is used as a stimulant, the other three substances have similar anesthetic effects as cocaine.[1] The other most common adulterants

Table 1. Qualitative composition of adulterants detected in seized cocaine samples

Adulterant(s)	Number of samples
Caffeine only	2
Lidocaine only	3
Benzocaine only	1
Boric acid only	2
Caffeine plus lidocaine	11
Caffeine plus benzocaine	1
Caffeine plus Lidocaine plus benzocaine	2

cited by UNODC,[1] levamisole, procaine, hydroxyzine and phenacetin were not found in the samples. Also, none of the samples showed positive result for the sulfate test. Finally, the content of cocaine was determined for the seized samples, which showed values between 1.0 and 73.6% m/m, according to the histogram displayed in Figure 1.
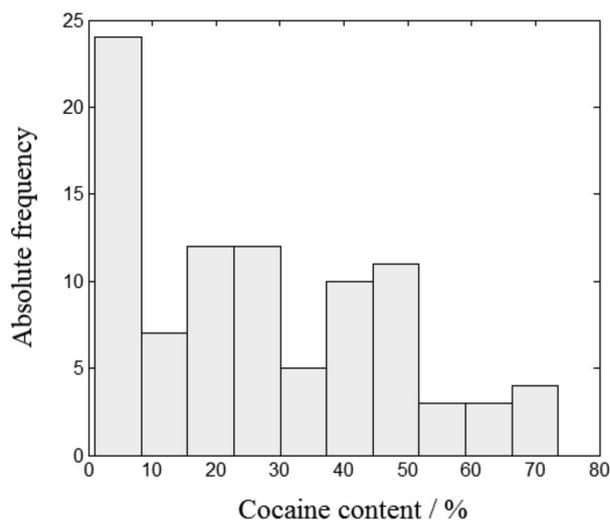
Figure 1. Histogram of the values of cocaine content determined in the seized samples.
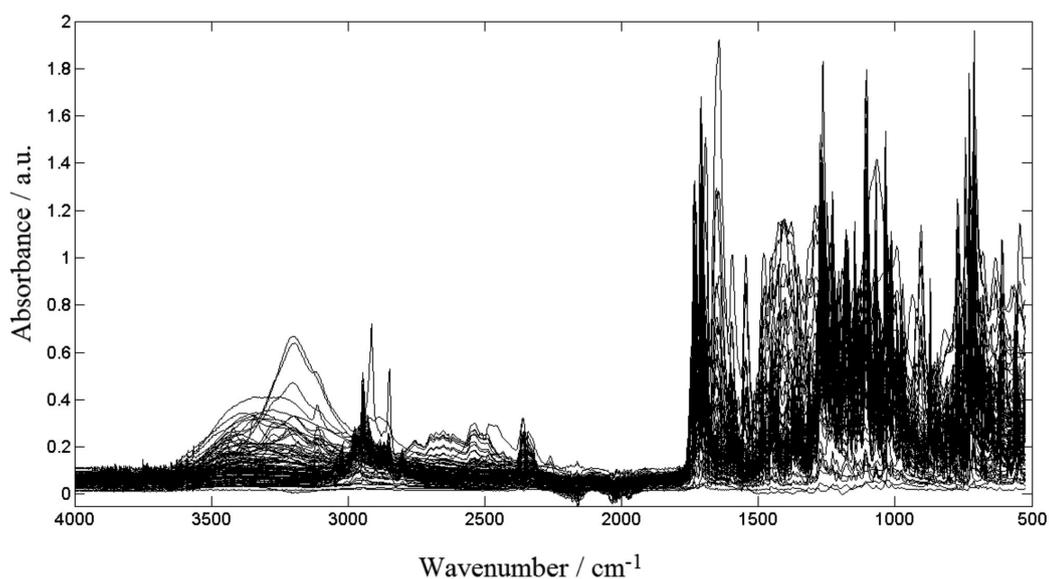
ATR-FTIR spectra of cocaine and adulterants

The ATR-FTIR spectra for the 91 seized cocaine samples are shown in Figure 2. As can be seen, they are very complex and present strongly overlapped bands, making difficult a qualitative analysis of these samples through a visual inspection. ATR-FTIR spectra of standards of cocaine base and hydrochloride are shown offset from the baseline in Figure 3. A search in the literature[23-25] allows identifying the characteristic peaks of the two forms of cocaine, taking into account that small variations in peak intensity and position occur between infrared spectra obtained by ATR and transmittance modes. Cocaine hydrochloride (Figure 3a) showed strong bands at 1728 and 1712 cm⁻¹ (stretching vibration of the two carbonyl groups), 1265, 1230 and 1105 cm⁻¹ (acetate C−O stretching), 1071, 1026 and 729 cm⁻¹ (mono substituted benzene stretching and the last one an out-of-plane bending). The most distinctive band is around 2540 cm⁻¹, attributed to the N−H stretching due to the hydrochloride salt formation. In addition, three faint bands were observed at 953, 924 and 897 cm⁻¹, which are difficult to assign, but probably attributable to bending vibrations out-of-plane.[23] Cocaine base (Figure 3b) showed similar strong bands with small shifts at 1734, 1707, 1273, 1227, 1107, 1068, 1035 and 712 cm⁻¹, this last one a C−H out-of-plane bending of the

benzyl group. The most pronounced difference in relation to hydrochloride spectrum is the absence of the band due to the N–H bond around 2500 cm[-1], and the presence of bands of medium intensity between 3000 and 2800 cm[-1] (the most intense peak at 2945 cm[-1]).

Figure 4 shows ATR-FTIR spectra of standards of the four detected adulterants. The region from 1800 to 1500 cm[-1] was considered more adequate to deduce the presence of adulterants, while the region from 960 to 860 cm[-1] more suitable to identify diluents.[23] The lidocaine hydrochloride spectrum (Figure 4a) showed the characteristic peaks[23] at 1655 cm[-1] (carbonyl), 1541 cm[-1] (in-plane bending in the N–H mode), 1490-1460 cm[-1] (hydrochloride), 1271 and 1249 cm[-1]

(tertiary amine), 1036, 715 and 604 cm[-1] (aromatic ring); in addition, the distinctive observed bands between 2600 and 2300 cm[-1] refers to the N–H vibrations of hydrochloride salt[26] and the bands between 3600 and 2800 nm to the amide bond. The caffeine spectrum (Figure 4b) showed the characteristic strong peaks at 1693 and 1644 cm[-1] (stretching vibrations of the two carbonyl groups), and peaks of lesser intensity at 1598 and 1547 cm[-1] (stretching vibrations of C=N and N–C=O), 1236 and 758 cm[-1] and a peak of stronger intensity at 744 cm[-1].[23,25] The benzocaine spectrum (Figure 4c) showed the characteristic bands at 3450-3200 cm[-1] (NH$_2$), 2983 cm[-1] (aromatic C–H), 1679 cm[-1] (carbonyl), 1272 cm[-1] (ν_E (CC,CO)), 1171 and 845 cm[-1] (C–H),

Figure 2. ATR-FTIR spectra of the 91 seized cocaine samples.
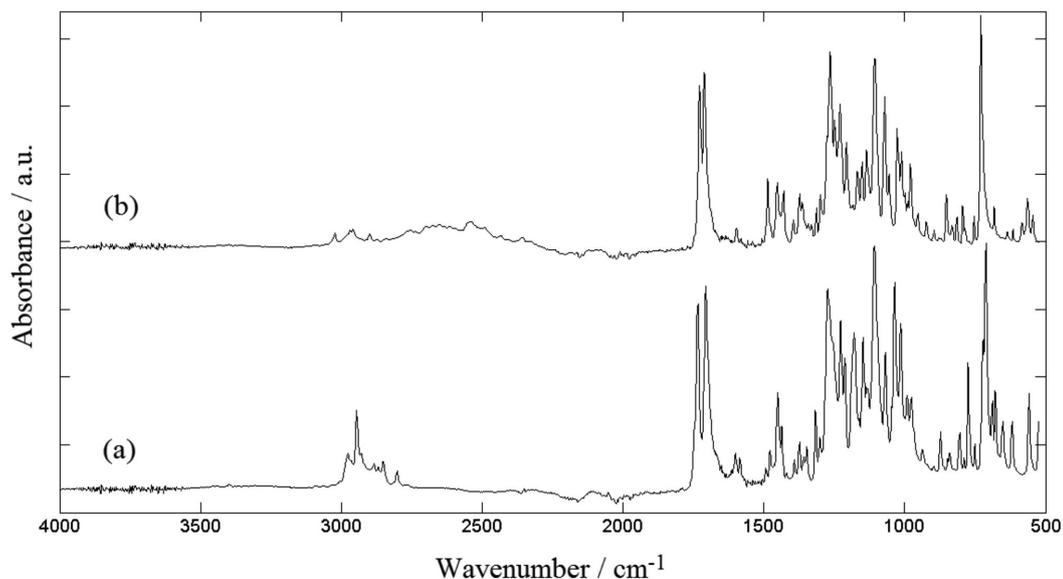
Figure 3. ATR-FTIR spectra of the standards of (a) cocaine base and (b) cocaine hydrochloride.
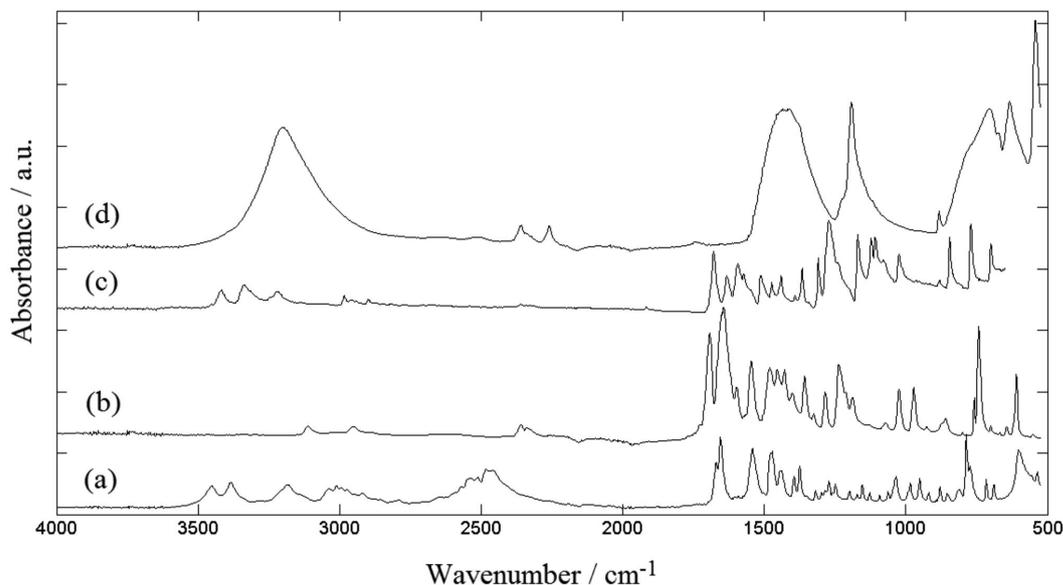
Figure 4. ATR-FTIR spectra of the standards of four adulterant substances identified in the cocaine seized samples: (a) lidocaine hydrochloride, (b) caffeine, (c) benzocaine and (d) boric acid.

770 and 700 cm^{-1} (C=O).[27] The boric acid spectrum (Figure 4d) showed the characteristic broad bands centered at 3200 cm^{-1} (O−H stretching), between 1550 and 1300 cm^{-1} (B−O stretching), between 1210 and 1170 cm^{-1} (B−O−H bending), between 720 and 680 cm^{-1} (O−H twisting), between 645 and 620 cm^{-1} (BO$_3$ bending), and a narrow and more intense peak at 543 cm^{-1} (O−B−O bending).[28]

An exploratory PCA model for the ATR-FTIR data

By observing Figures 2-4, it can be noted that some parts of the spectra, such as the regions above 3600 cm^{-1} and between 1800 and 2300 cm^{-1}, present no significant absorption from the cocaine forms or the main adulterants. Thus, it was decided to delete these spectral regions since they can contribute only with noise or other artifacts, such as the small absorbances near 2000 and 2100 cm^{-1} due to the diamond ATR crystal.[29] A preliminary PCA model allowed to identify the presence of a peak around 2350 cm^{-1} in some samples, which was modeled in PC5 (ca. 5%). This peak was attributed to the presence of gas phase carbon dioxide,[30] and thus, it was also decided to delete the region between 2300 and 2400 cm^{-1}. Then, the models were built using the spectral regions from 525 to 1800 cm^{-1}, and from 2400 to 3600 cm^{-1}, with a step of about 1.9 cm^{-1}. The spectral data were organized in a 93 × 1285 matrix, containing 91 spectra of the seized samples plus the spectra of the standards of cocaine base and cocaine salt in the rows. This matrix was preprocessed by SNV and mean centered. PCA analysis provided a model that accounted for 62.99% of the total data variance with the first two

PCs. The scores plot of PC1 × PC2 is shown in Figure 5. PC1 (42.84%) discriminates samples as a function of their cocaine content, and separates the more adulterated ones (the region of more positive scores, flagged on the right of Figure 5) from the more pure ones (predominantly negative scores on PC1). All the 19 samples in this region contained the adulterants lidocaine, caffeine and/or benzocaine, as detected by GC-MS, and maximum cocaine contents around 15%. PC2 (20.15%) discriminates samples as a function of their cocaine chemical form, and separates cocaine hydrochloride (positive scores on PC2, upper left part of Figure 5) from cocaine base (negative scores on PC2, lower left part). Almost all of the higher content (above around 20%) of cocaine hydrochloride samples lie around the standard of cocaine hydrochloride, indicated by an arrow in the upper left quadrant of Figure 5. Conversely,
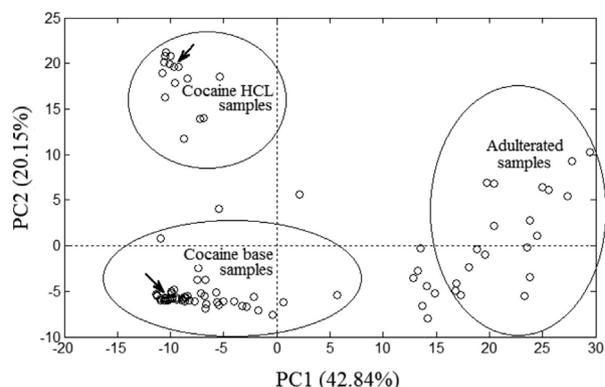

Figure 5. PCA model: scores of PC1 × PC2. The groups of more adulterated (in general), less adulterated cocaine base and less adulterated cocaine hydrochloride samples are highlighted. The arrows indicate samples of the standards of cocaine base and cocaine hydrochloride.

a group of samples highlighted in the lower left quadrant around the standard of cocaine base (also indicated by an arrow) corresponds to samples of this form of cocaine and of higher content. Also, most of the samples outside of the three highlighted groups are cocaine base.

To complement the interpretation of the PCA model, the loadings (Figure 6) must be analyzed in order to identify the spectral regions associated with the observed discriminations. This analysis corroborated the interpretation of PC1 as a factor that discriminated the samples adulterated by lidocaine/caffeine/benzocaine, while PC2 discriminated the cocaine forms. Since the more adulterated samples showed positive scores on PC1, the wavenumbers with positive loadings were associated to bands/peaks of the adulterants. In the loadings plot of PC1 (Figure 6a), a broad band between 3500 and 3000 cm^{-1} was assigned to the N−H stretching vibrations of lidocaine and benzocaine (Figures 4a and 4c). Note that none of the spectra of the two forms of cocaine has significant absorption in this region (Figure 3). Other discriminant peaks are at 1653 cm^{-1} (carbonyls of lidocaine and caffeine at 1655 and 1644 cm^{-1}, respectively) with a shoulder at 1689 cm^{-1} (other carbonyl of caffeine), 1545 cm^{-1} (N−C of caffeine at 1547 cm^{-1} and N−H of lidocaine at 1541 cm^{-1}), 1471 cm^{-1}, assigned to lidocaine, 744 cm^{-1}, which corresponds to the second strongest peak of caffeine, and 607 cm^{-1} (aromatic C−H of lidocaine). Reciprocally, wavenumbers with negative loadings on PC1 were related to the less adulterated/more pure samples. The most characteristic negative loadings peaks are at 2945 cm^{-1}, related to cocaine base, at 1734 and 1700 cm^{-1} (two carbonyls), 1265 and 1107 cm^{-1} (acetate C−O), 1036 and 712 cm^{-1} (mono substituted benzene), all related to the main peaks of the two forms of cocaine. The region around 2500 cm^{-1} showed no significant contribution on PC1, because it is assigned to N−H vibrations of hydrochloride, which are present in both the cocaine salt and the adulterant lidocaine.

By comparing the loadings plot of PC2 (Figure 6b) with the spectra of standards of cocaine salt and base
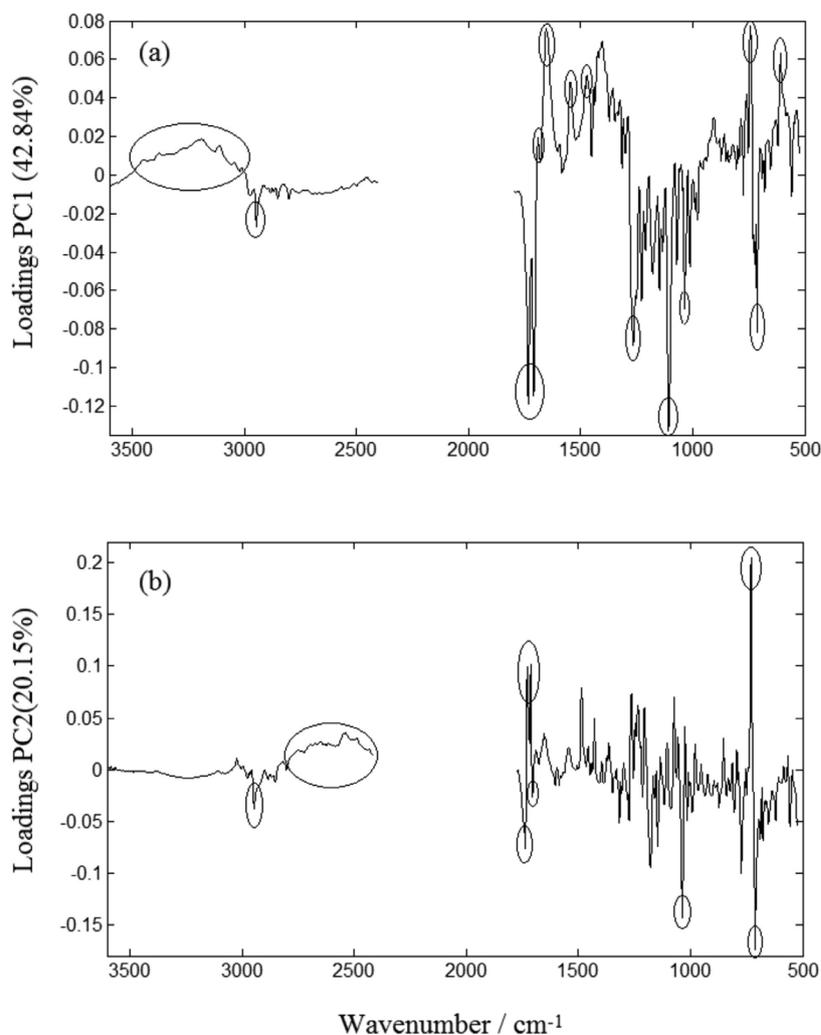
Figure 6. Loadings of the PCA model: (a) loadings of PC1 *vs.* wavenumber and (b) loadings of PC2 *vs.* wavenumber. The main discriminant peaks are marked.

(Figure 3), it can be noted the characteristic absorptions associated with cocaine hydrochloride on the positive values, such as the broad band from 2800 to 2400 cm^{-1} centered at 2540 cm^{-1} (hydrochloride N−H stretching), and the peaks at 1728 and 1712 cm^{-1} (two C=O), and at 729 cm^{-1} (mono substituted benzene stretching). Reciprocally, the absorptions assigned to cocaine base show negative loadings on PC2, such as the peaks at 2945, 1736, 1705, 1036 and 712 cm^{-1}. Other PCs should describe the presence in smaller amounts of other adulterants, diluents or decomposition products. The loadings for PC3 and PC4 are shown in Figure 7. PC3 (Figure 7a) accounted for 14.55% of variance and was characterized by positive sharp intense peaks in the carbonyl region, the most intense two were at 1694 and 1652 cm^{-1}. These peaks were attributed to bicarbonate and/or carbonate vibrations[31] and were associated with a group of crack samples, which showed positive scores on PC3. PC4 (Figure 7b) accounted for 9.33% of variance and was

characterized by the presence of a negative large band of loadings centered at 3350 cm^{-1}, associated with the only two samples adulterated with boric acid, which showed more negative scores on PC4. The first four PCs accounted for 86.9% of the total data variance. Other PCs accounted for up to no more than 3% each one.

Another way to observe the identified clusters of samples is by using hierarchical cluster analysis (HCA).[15,16] HCA was performed on the scores of a PCA model, with two PCs, and using Mahalanobis distance and the Ward's method as the similarity criterion. The resulting dendrogram is shown in Figure 8, in which it is possible to observe three clusters corresponding to the same groups obtained by PCA (Figure 5).

PLS-DA models

After constructing an exploratory and descriptive unsupervised PCA model, the next step was to build
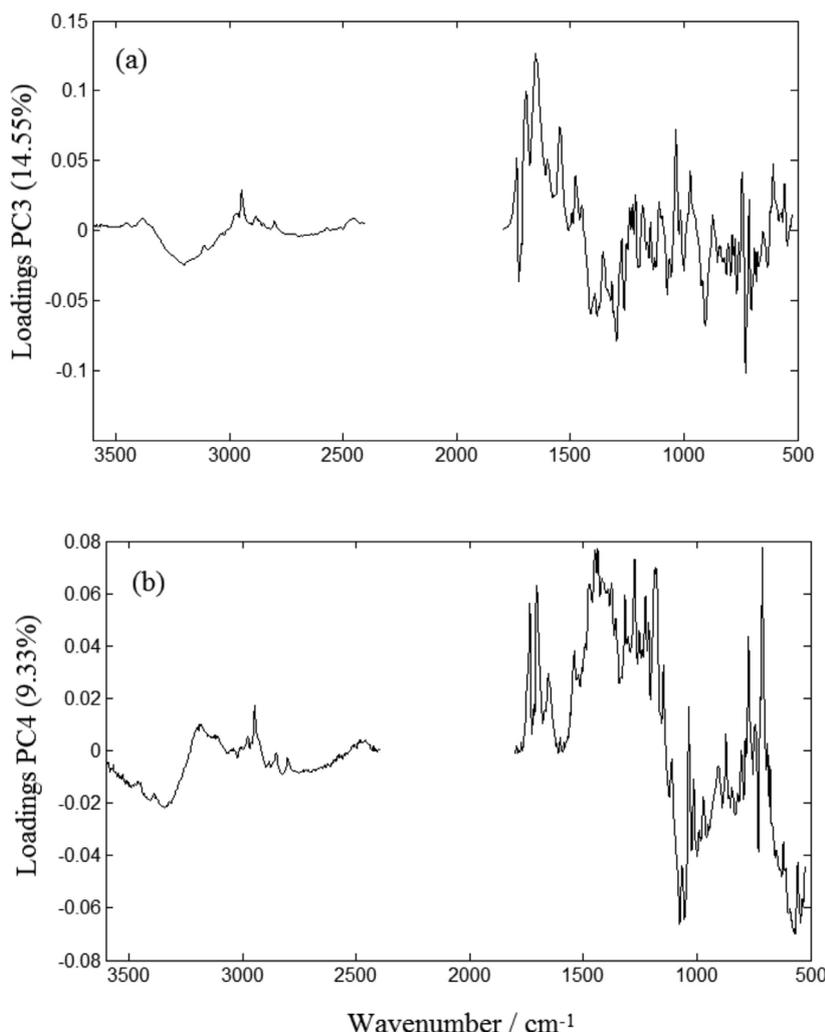
Figure 7. Loadings of the PCA model: (a) loadings of PC3 *vs.* wavenumber and (b) loadings of PC4 *vs.* wavenumber.
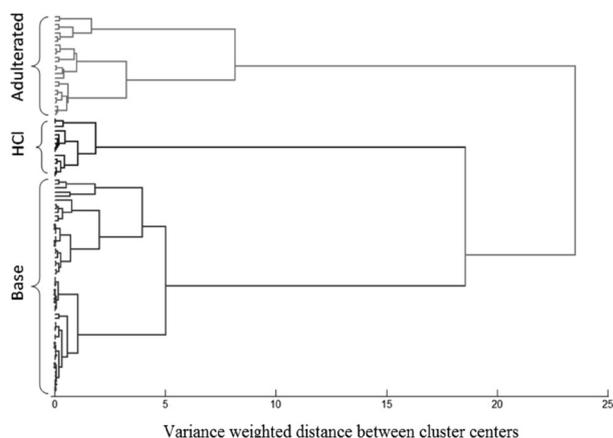
Figure 8. Agglomerative dendrogram obtained by HCA using Mahalanobis distance and the Ward's method.

supervised classification PLS-DA models based on ATR-FTIR spectra. The aim was to develop a rapid and non-destructive tool for classifying the seized cocaine samples as a function of their chemical form and content. Previously, an attempt to develop a multivariate calibration PLS model for predicting cocaine content was carried out, trying to correlate the spectra with the results obtained by GC-MS. Nevertheless, reasonable results were not obtained since many of the samples presented errors of prediction above 20%. This is certainly caused by the large chemical variety of the samples, leading to the detection of a high number of outliers (the snowballing effect).[32] Thus, it was decided to build a PLS-DA model to discriminate samples as of high or low content, adopting a threshold value of 15% m/m of cocaine. This value was chosen because it represents a gap, i.e., a small number of samples showed cocaine contents between 15 and 20%, and because it was considered representative to discriminate concentrated and diluted samples by the Brazilian Federal Police. So, a vector of dependent variables was built containing 1 and 0 for samples of high and low content, respectively. Another PLS-DA model was constructed to discriminate samples of hydrochloride from cocaine base, for which the values of 0 and 1 were arbitrarily assigned, respectively.

The data were divided in a training set (57 spectra) used to build the model, and a test set (34 spectra) used to validate it, by using the Kennard-Stone algorithm. As in the previous PCA model, the spectra were preprocessed by SNV and mean centered. The best number of LVs was chosen by leave-one-out cross-validation. The results are presented through the confusion matrices (Tables 2 and 3), visualization tools typically utilized in supervised learning, in which each column represents the instances in a predicted class, while each row represents the instances in an actual class. The model for discriminating cocaine samples of high content was estimated using one LV and accounted for 38.6 and 64.7% of the variance in X and Y blocks, respectively. This model (Table 2) presented sensitivity (percentage of true positive) of 95% and specificity (percentage of true negative) of 83%. The coefficients of regression are plotted in Figure 9a, in which it is possible to note that the main variables that contribute to discriminate more concentrated samples (positive values) are related to spectral peaks of the two forms of cocaine: 2945, 1732, 1709, 1265, 1107, 729 and 712 cm^{-1}. The main variables that contribute to discriminate more diluted samples (negative values) are related to peaks of the main adulterants lidocaine (1655, 1471 and 607 cm^{-1}) and caffeine (1545 and 744 cm^{-1}).

Table 2. Confusion matrix for the discrimination of seized samples of high/low cocaine content obtained with PLS-DA

		Predicted			
		Training set		Test set	
Actual	Content	Low	High	Low	High
	Low	18	3	6	2
	High	3	33	0	26

The model for discriminating the cocaine chemical form was estimated using eight LVs and accounted for 93.4 and 86.6% of the variance in X and Y blocks, respectively. This model (Table 3) presented sensitivity of 97% and specificity of 88%. The coefficients of regression are plotted in Figure 9b, in which it is possible to note some of the characteristic peaks of cocaine hydrochloride with negative values (2800-2400, 1728 and 1100 cm^{-1}) and some of the peaks assigned to cocaine base with positive values (2945 and 1276 cm^{-1}).

Table 3. Confusion matrix for the discrimination of the chemical form of the cocaine in the seized samples obtained with PLS-DA

		Predicted			
		Training set		Test set	
Actual	Form	Hydrochloride	Base	Hydrochloride	Base
	Hydrochloride	16	1	6	2
	Base	0	40	2	24
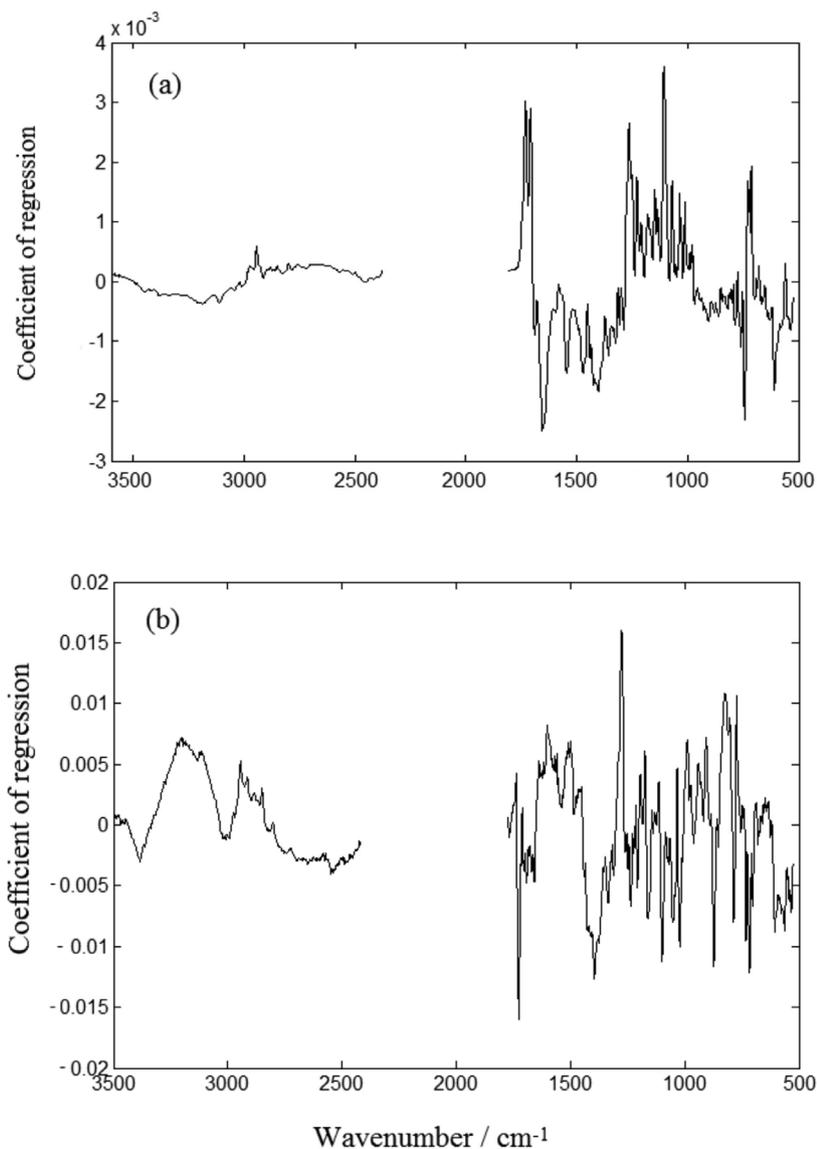
Figure 9. Coefficients of regression of the PLS-DA models for discriminating cocaine samples as a function of their (a) content and (b) chemical form.

Conclusion

This article provided an exploratory study that characterized the chemical composition of the cocaine samples seized by the Federal Police of the Minas Gerais State (Brazil) during the period from 2008 to 2010, based on the chemometric analysis of their ATR-FTIR spectra. Pattern recognition analysis by PCA allowed inferring that the main pattern in these data, as described by PC1 (42.8% of the variance), is the correlated presence of the adulterants lidocaine, caffeine and to a lesser extent benzocaine, which were directly related to the lower purity samples. The second main pattern, as described by PC2 (20.2%), is the discrimination of the two chemical forms of cocaine, base and salt. This type of information is very useful for the Police and can help to establish common origins for

otherwise unrelated cocaine seizures and to investigate the frequent changes in the illicit market.

Moreover, two supervised classification PLS-DA models were developed for discriminating concentrated from diluted, and hydrochloride from base cocaine samples. They showed rates of true positive between 95 and 97%, and of true negative between 83 and 88%. In general, limits of percentages of 5% for both false positive and false negative are recommended for the validation of qualitative methods.[33] But in the case of seized drug analysis, in which the complexity of the matrices makes qualitative tests especially prone to produce false results,[21] higher limits are considered acceptable. These models constitute a simple, rapid and non-destructive tool, of great importance to speed criminal investigations. In some countries, such as Spain, the sentence imposed by the court for cases

of cocaine traffic varies according to its percent content in the samples.[23] In these cases, a threshold value for discriminating the seized drug samples can be established according to the legislation.

Acknowledgments

References

1. United Nations Office on Drugs and Crime (UNODC); *World Drug Report*; UNO: New York, USA, 2011, p. 85-126.

2. Shannon, M.; *Ann. Emerg. Med.* **1988**, *17*, 1243.

3. Carvalho, D. G.; Mídio, A. F.; *Braz. J. Pharm. Sci.* **2003**, *39*, 71.

4. Goldstein, R .A.; DesLauriers, C.; Burda, A.M.; *DM-Dis. Mon.* **2009**, *55*, 6.

5. Ryder, A. G.; *J. Forensic Sci.* **2002**, *47*, 275.

6. Noonan, K. Y.; Tongue, L. A.; Fenton, O. S.; Damiano, D. B.; Frederick, K. A.; *Appl. Spectrosc.* **2009**, *63*, 742.

7. O'Connell, M. L.; Ryder, A. G.; Leger, M. N.; Howley, T.; *Appl. Spectrosc.* **2010**, *64*, 1109.

8. Fenton, O. S.; Tongue, L. A.; Moot, T. H.; Frederick, K. A.; *Spectrosc. Lett.* **2011**, *44*, 229.

9. Goh, C. Y.; van Bronswijk, W.; Priddis, C.; *Appl. Spectrosc.* **2008**, *62*, 640.

10. Melucci, D.; Monti, D.; D'Elia, M.; Luciano, G.; *J. Forensic Sci.* **2012**, *57*, 86.

11. Sondermann, N.; Kovar, K. A.; *Forensic Sci. Int.* **1999**, *106*, 147.

12. Moros, J.; Galipienso, N.; Vilches, R.; Garrigues, M.; de la Guardia, S.; *Anal. Chem.* **2008**, *80*, 7257.

13. Kher, A.; Mulholland, M.; Green, E.; Reed, B.; *Vib. Spectrosc.* **2006**, *40*, 270.

14. Muehlethaler, C.; Massonnet, G.; Esseiva, P.; *Forensic Sci. Int.* **2011**, *209*, 173.

15. Brereton, R .G.; *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; Wiley: Chichester, USA, 2003.

16. Sena, M. M.; Frighetto, P. J.; Valarini, R. T. S.; Tokeshi, H.; Poppi, R. J.; *Soil Till. Res.* **2002**, *67*, 171.

17. Brereton, R. G.; *Analyst* **2000**; *125*, 2125.

18. Kennard, R. W.; Stone, L. A.; *Technometrics* **1969**, *11*, 137.

19. Dhanoa, M. S.; Lister, S. L.; Sanderson, R.; Barnes, R. J.; *J. Near Infrared Spectrosc.* **1994**, *2*, 43.

20. Departamento de Polícia Federal; *Instrução Técnica No. 006/2006*; GAB/DITEC: Brasília, Brasil, 27 julho, 2006.

21. United Nations Division of Narcotic Drugs (UNODC); *Recommended Methods for Testing Cocaine*; Manual for Use by National Narcotics Laboratories, Vienna, Austria, 1986.

22. Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Recommendations, 6th Rev.; United States Department of Justice: Washington, DC, USA, 2011.

23. López-Artíguez, M.; Cameán, A.; Repetto, M.; *J. Forensic Sci.* **1995**, *40*, 602.

24. Koulis, C. V.; Reffner, J. A.; Bibby, A. M.; *J. Forensic Sci.* **2001**, *46*, 822.

25. http://etd.lib.metu.edu.tr/upload/12606293/index.pdf accessed in March 2013.

26. Neville, G. A.; Regnier, Z. R.; *Can. J. Chem.* **1969**, *47*, 4229.

27. Balci, K.; Akyuz, S.; *Vib. Spectrosc.* **2008**, *48*, 215.

28. Servoss, R. R.; Clark, H. M.; *J. Chem. Phys.* **1957**, *26*, 1175.

29. Küpper, L.; Heise, H. M.; Butvina, L. N.; *J. Mol. Struct.* **2001**, *563*, 173.

30. http://webbook.nist.gov/cgi/cbook.cgi?ID=C124389&Type=IR-SPEC&Index=1 accessed in November 2012.

31. Miller, F. A.; Wilkins, C. H.; *Anal. Chem.* **1952**, *24*, 1253.

32. Valderrama, P.; Braga, J. W. B.; Poppi, R. J.; *J. Agric. Food Chem.* **2007**, *55*, 8331.

33. Trullols, E.; Ruisánchez, I.; Rius, F. X.; *TrAC, Trends Anal. Chem.* **2004**, *23*, 137.