Characterization of sildenafil citrate tablets of different sources by near infrared chemical imaging and chemometric tools

Guilherme P. Sabin [1], Valeria A. Lozano [2], Werickson F. C. Rocha [3], Wanderson Romão [4], Rafael S. Ortiz [5] and Ronei J. Poppi [1*]

[1] Institute of Chemistry, State University of Campinas, 13084-971 Campinas, SP, Brazil.

[2] Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, 2000 Rosario, Argentina.

[3] National Institute of Metrology, Quality and Technology (Inmetro), Directorate of Industrial and Scientific Metrology, Chemical Metrology Division, 25250-020, Xerém, Duque de Caxias, RJ, Brazil.

[4] Federal Institute of Education, Science and Tecnology of Espírito Santo, 29106-010, Vila Velha, ES, Brazil.

[5] Brazilian Federal Police, Ministry of Justice, Rio Grande do Sul Technical and Scientifical Division, 90160-092, Porto Alegre, RS, Brazil.

Abstract

In this paper the chemical imaging technique by near infrared spectroscopy was applied for characterization of drug formulations in tablets of sildenafil citrate of different sources. This work is original and has never been published because it was used the sildenafil citrate, including in the commercial one called "Viagra". In the first stage of the study, a chemometric method based on multivariate curve resolution was properly chosen for the development of the distribution map of concentrations of the active ingredient in tablets. In the second step, the normalized histograms of images from active ingredient were classified by hierarchical cluster analysis. Finally it was possible to recognize the patterns of homogeneity of images in sildenafil citrate tablet. This concept can be used to improve the knowledge of industrial products and processes, as well as, for characterization of counterfeit drugs.

Keywords: sildenafil citrate; chemical imaging; near infrared spectroscopy; multivariate curve resolution; classification; recognize the patterns.

Introduction

Patents are a form of legal protection of intellectual property that provide exclusive rights to make, use, import, sell and offer for sale the invention for up to 20 years. The economic logic of this protection mechanism is that the profits provided by the production license of a patented product guarantee the patent owner the reinvestment in research and development of new products[1-4]. Social factors, however, may eventually prevail over this economic development engine aspect, discussing the possibility of patent infringement. One of these factors is the great technological discrepancy in peripheral countries in relation to developed countries, and their low purchasing power to buy the next-generation products manufactured by the major economic centers.

Viagra® is the first drug approved to treat erectile dysfunction. Its mechanism is blocking the enzyme phosphodiesterase type 5 (PDE5), involved in the erection process. It has vasodilating properties and effects on blood pressure, and like nitrates, it works by the nitric oxide cyclic guanosine monophosphate pathway[5,6]. It is estimated that erectile dysfunction affects between 48 % and 52 % of men from 40 to 70 years.

The sildenafil citrate (active ingredient of the Viagra®) was registered in the European Union in 1991 by Pfizer. Nowadays, due to the expiration of the patent in Brazil, at least ten different companies are marketing this product, but in cheaper way. Additionally, the counterfeiting of Viagra® tablets has become an important and dangerous problem for pharmaceutical market, where the Brazilian Federal Police has reported many seizures mainly in south region of

Brazil (state of Parana). In 2007-2010 periods, the Federal Police reported a great numbers of seizures (371) being the counterfeit tablets market related to erectile dysfunction treatment responsible by 80 % of the seizures. Therefore, to control the quality of new pharmaceutical formulations and distinguish between authentic and counterfeit tablets is necessary the development of powerful analytical tools. Although analytical methods such as chromatography[7-10], voltammetry[11-13] and colorimetric determination[14] were reported in Viagra® tablets analysis, they are high time-consuming and require extensive sample preparation.

Due to its advantages such as non-destructive analysis, speed, and less consumption of chemicals, the near infrared spectroscopy (NIR) has been accepted in various fields of pharmaceutical industry[15,16]. NIR has the potential to provide increased process and product understanding which goes well with the process analytical technology (PAT) initiative of the Food and Drug Administration (FDA)[17].

Hyperspectral imaging shows a considerable promise for providing high-quality spectral information on active principle distribution within pharmaceutical formulations. The robust reliable combination of chemical (molecular spectroscopy) and physical (digital imaging) features have been successfully applied to diverse fields such as remote sensing[18,19], astronomy[20,21], agriculture[21,22], food[23] and pharmaceuticals[24-26].

Quantitative analysis of pharmaceutical samples using Near Infrared-Chemical Imaging (NIR-CI) has been performed in techniques as Partial Least Squares Regression (PLS) or PLS-Discriminant Analysis (PLS-DA)[24,26,27]. However, these techniques require a complete calibration set of samples with

the corresponding time consuming until the calibration model that is performed and adequately tested. Concerning the quantification purposes without needing a previous calibration model, Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) may be presented as a promising alternative by using second order advantage, i.e., analytes quantification in the presence of unknown interference. MCR-ALS is an algorithm that fits the requirements for image resolution[28-31].This method decompose the unfolded hyperspectral data cube, the matrix **X** ($xy \times \lambda$), into the product of two matrices, **C** ($xy \times K$), containing the concentration profiles and $\mathbf{S}^T$ ($K \times \lambda$), containing the spectral profiles for each K component (Figure 1 and equation 1). In this case, xy is the spatial image dimensions and λ the number of spectral data points.

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

where $\boldsymbol{E}$ ($xy \times \lambda$) corresponds to the experimental error matrix.

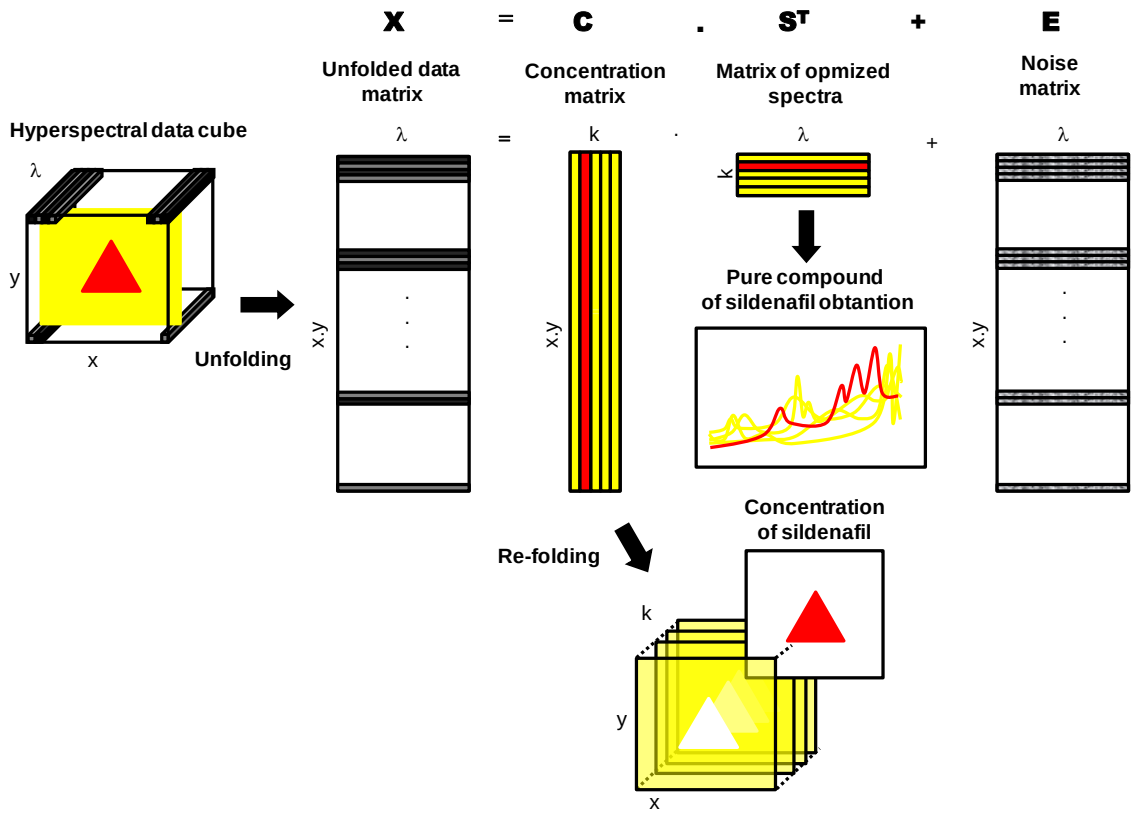

Figure 1. Scheme of MCR-ALS analysis application to the hyperspectral cube.

To initiate the iterative MCR-ALS procedure, an initial estimation is needed for the spectral profiles. Different methods are used for this purpose such as evolving factor analysis[32-34], the determination of the purest variables[35-38] or by using pure spectra of the components[39].

This work aims to build a concentration distribution map of sildenafil citrate in tablets of different sources where the chemical composition of all excipients constituents is not truly known using the multivariate curve resolution approach. In addition, the profile of the concentration distribution maps for each tablet were assessed for obtaining recognition pattern of images by analyzing hierarchical classification of their histograms.

Experimental section

In this work, tablets containing sildenafil citrate as active ingredient of six different formulations from different sources were studied. These formulations were named as **A** to **F**. The formulations from **A** to **E** were provided by Brazilian Federal Police and correspond to several trademarks of prohibited marketing. The **F** formulation was an authentic sample of Viagra® (Pfizer Ltda). For each formulation, it was performed an image acquisition of four tablets.

The acquisition of images was obtained by NIR-CI technique using Spotlight 400N FT-NIR Imaging by PerkinElmer. The mapping measurements were performed four times per sample type, spatial resolution 25 µm, 16 scans and spectral range between 6500 and 4000 cm^{-1}. The data array (80x80 pixels and 158 wavelengths) was obtained directly on surface of the tablet (after coating removal). The raw data were transformed to inverse logarithm of the reflectance values (pseudo-absorbance) and unfolded for further preprocessing by Multiplicative Scattering Correction (MSC). Since only the sildenafil citrate spectrum is known in advance, the tool choose for construction of the distribution maps was the MCR[31] as a quantitative way. The ALS optimization was initialized by loadings of Principal Component Analysis (PCA) using mean-centered data. Constraints as external spectral knowledge of the active ingredient as well as non-negativity and closure for concentration were used as a way to minimize rotation ambiguity. Thus, the standard spectrum of sildenafil citrate was compared among all loadings per sample and substituted by its most similar loading profile and a new optimization process using ALS was performed. In this direction, semi-quantitative information can be obtained for active ingredient by recovering spectral information after optimization stage. The distribution maps (images) were converted to histograms of pixel

concentration. Those histograms were normalized (mean concentration = 1) and classified by non-supervised Hierarchical Cluster Analysis (HCA).

Data analysis was performed in Matlab version 7.8 using routines developed in the laboratory and the MCR Toolbox provided by Romà Tauler[31].

Results and discussion

The first step in the development of the MCR-ALS modeling was the initialization of the model by using the loadings obtained by principal component analysis. The use of loadings for initialization of the MCR-ALS can be considered problematic for optimization of the **C** matrix (concentration), since the rotation ambiguity is present in this situation. In other words, the quantitative approach used in multivariate resolution, Equation 1, would be confounded by qualitative information of **T**-matrix (scores) and **P**-matrix (loadings) obtained by PCA analysis, Equation 2.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{T} + \mathbf{E} \qquad (2)$$

In this case, the **C** matrix would bring **T** matrix qualitative information because of the initializations by loadings. In this sense, the use of the purest **S** matrix and constraints are frequently a way for minimization of the ambiguities and recovery quantitative information. However, the purest spectra may present high condition number in relatively homogeneous images (i.e. high similarity among all spectral information) while loadings are always orthogonal with condition number equal to one. In MCR-ALS, **E** matrix (errors) can be seen as a lack of fit of the product of **C**-matrix (concentrations) and **S** matrix (pure spectra) for recovery the **X** matrix. For good results, it is desirable decreasing the

sensitivity of the **E** matrix. In this direction, a lower condition number of **S** matrix contributes to highest robustness for **C** matrix.

Due to use of loadings combined with sildenafil citrate pure spectrum for initialization of MCR-ALS, a high orthogonally was reached. On the other hand, the results of MRC-ALS optimization, Figure 2, show excellent recoveries for the active ingredient spectra for all drug formulations samples.

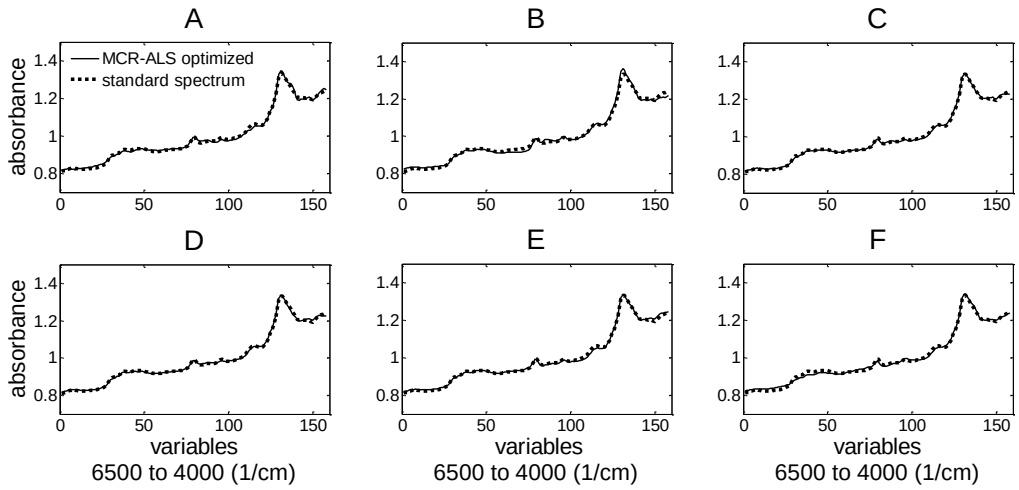

Figure 2. Overlap of the spectra of sildenafil citrate: standard and optimized by MCR-ALS for the drug formulations **A** to **F**.

In Figure 2, the dotted spectra were obtained directly from a standard of sildenafil citrate, while the line spectra were obtained by MCR-ALS. Note that the information recovered from optimization process in MCR is highly correlated with the known spectrum of the active ingredient standard, reaching selectivity for quantitative approach. Moreover, loadings assurance a lower similarity among all excipients and sildenafil citrate profiles, which help increase the sensitivity of the results.

This procedure is especially interesting for analysis where the components are not totaly known, but only part of the information is available. The lack of fit lower than 1% between the MCR-ALS resolution results and the original **X**

matrices added to good recovery of the sildenafil citrate spectrum assurance reliable images. The calculation of the lack of fit was performed by using the Equation 3:

$$lack\ of\ fit\ (\%) = \sqrt{\frac{\sum(x_{ij}^{*} - x_{ij})}{\sum x_{ij}^{2}}} \times 100 \qquad (3)$$

where x_{ij} is an element of the experimental matrix **X** and x_{ij}^{*} the element of the MCR-ALS reproduced matrix **X***.

In Figure 3, concentration maps of four different tablets for each source of the drug formulation (**A** to **F**) are showed. Note that on the drug **A**, there are a lot of big heterogeneous regions (i.e. excipient has big particle size) also observed on tablets through to optical microscopic view. In second line **B** a slightly different profile arises, smaller and denser point of active may be seen. In next two samples, respectively **C** and **D**, more homogeneous images are brought. Those drugs have very similar characteristics; maybe by represent the same brand, but different packaging, batch and seizure. Likewise **A**, the drug formulation **E** is recognized by low homogeneity that divides yours pixels at nearly absence or presence of the sildenafil information. The drug formulation **F**, the original drug formulation, presents a very homogeneous profile of concentration. Moreover, similar pattern of distributions among image replicates is possible to observe in this figure.

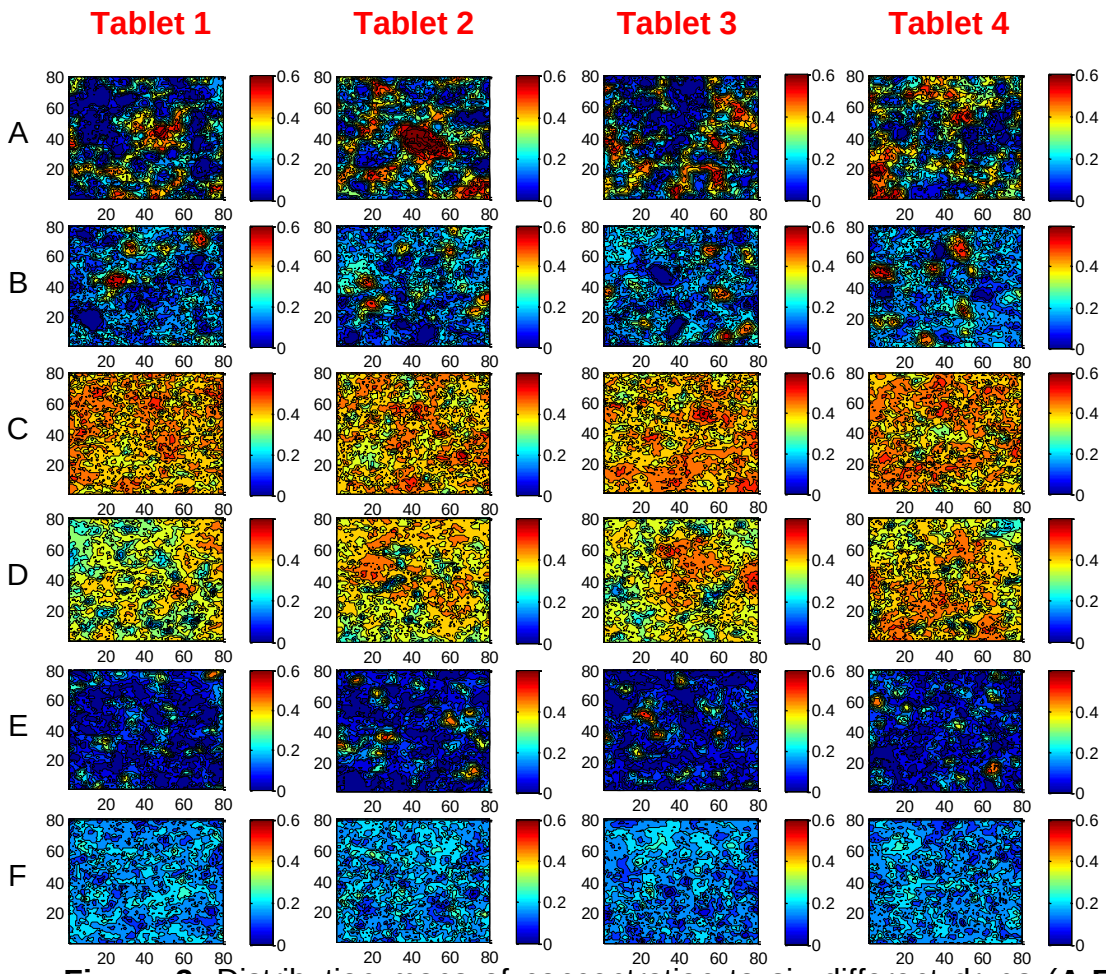

Figure 3. Distribution maps of concentration to six different drugs (**A-F**) obtained by MCR-ALS.

The images above can be translated into histograms of frequency distribution of concentrations. This type of result analysis removes the spatial components *xy* of the acquired information, but retains the ability to study the distribution profile, i.e., the homogeneity of the active ingredient information.

In Figure 4, histograms were normalized in order that the mean concentration is equal to one. Thus, it becomes possible to analyze profile of histograms independently of its mass fraction on the tablets surface.

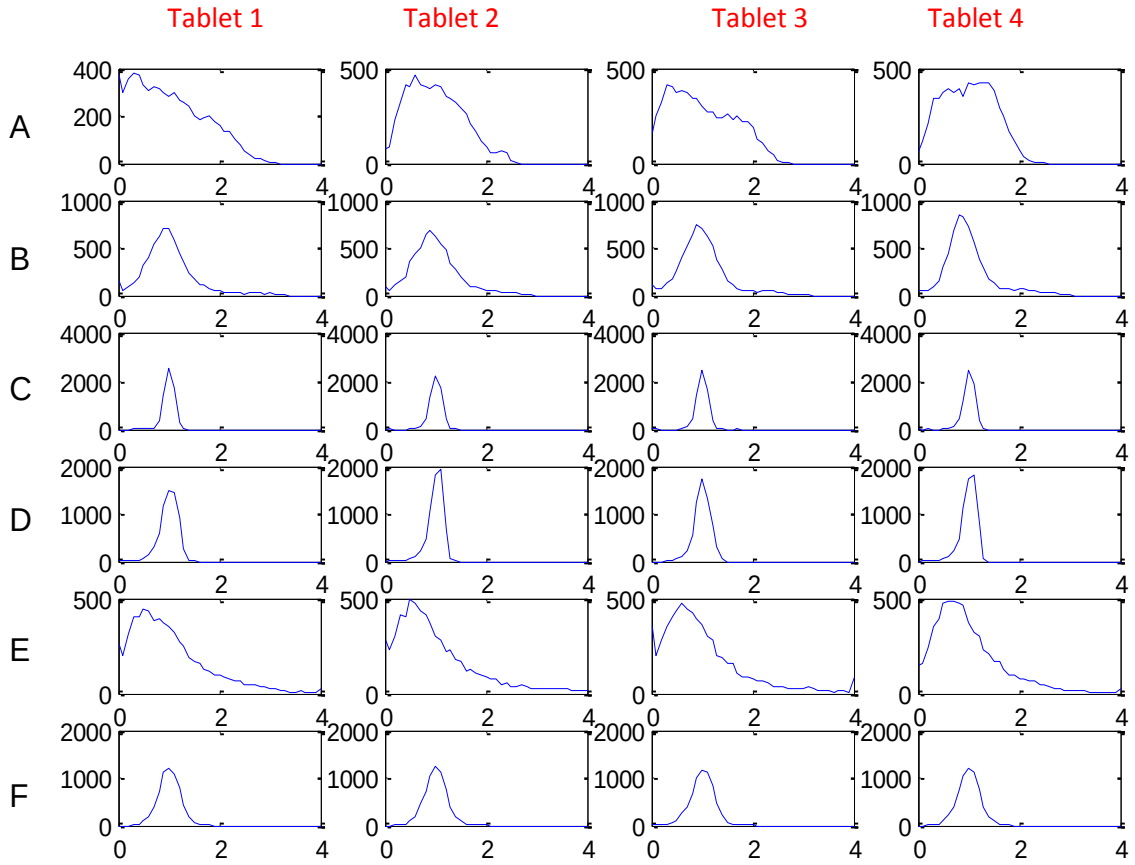

Figure 4. Histogram of the images centered at unit for the drug formulations **A** to **F**.

The histogram is a graphical technique for showing both the skewness and kurtosis of the data set. In this case, it is represented by the frequency distribution of concentrations. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Negative kurtosis would indicate a flat distribution, which is said to be platykurtic. Positive kurtosis would indicate a peaked distribution, which is said to be leptokurtic. Finally, the normal distribution has zero kurtosis, and it is said to be mesokurtic. In Figure 4, it is possible to classify the kurtosis from the concentration distribution of **D** and **F** drugs in mesokurtic, because among the six drugs analyzed they have a more homogeneous concentration distribution on the drug surface. For **A, B** and **E**

drugs the kurtosis can be classified as platykurtic, because they are less homogeneous. For **C** drugs the kurtosis can be classified as leptokurtic.

Skewness is a measure of the distribution symmetry, or more precisely, the lack of symmetry. From observation of the Figure 4, it can be inferred that in **A** and **E** formulations the sildenafil citrate concentration distribution on the surface is less symmetrical due to the particle size heterogeneity in those formulations. By the other hand, **B**, **C**, **D** and **F** formulations have more symmetrical distribution among the formulations studied, since the size of the particles are more homogeneous. Therefore, it can be concluded that by kurtosis and skewness histogram classification it is possible to study drug formulation homogeneity. Also they can be used as an indicative to characterize the formulations.

In order to provide a pattern recognition of the distribution of sildenafil citrate concentrations in tablets, it was used an established chemometric method called HCA. It is based on the multivariate distances among the samples by using an agglomerative procedure. The objective has been to provide a classification non-supervisioned of the different kind of drugs by similarity of distribution of the active. Thus, to show a way to recognition of images and its variability is an important parameter to study within and between samples. The dendrogram, Figure 5, describes different clusters for each drug formulation and it suggests an effective method to indentify the product or process of manufacture. This principle may be useful to identify different type of counterfeits including drugs that present the same composition, but different process of homogenization, particle size among other.

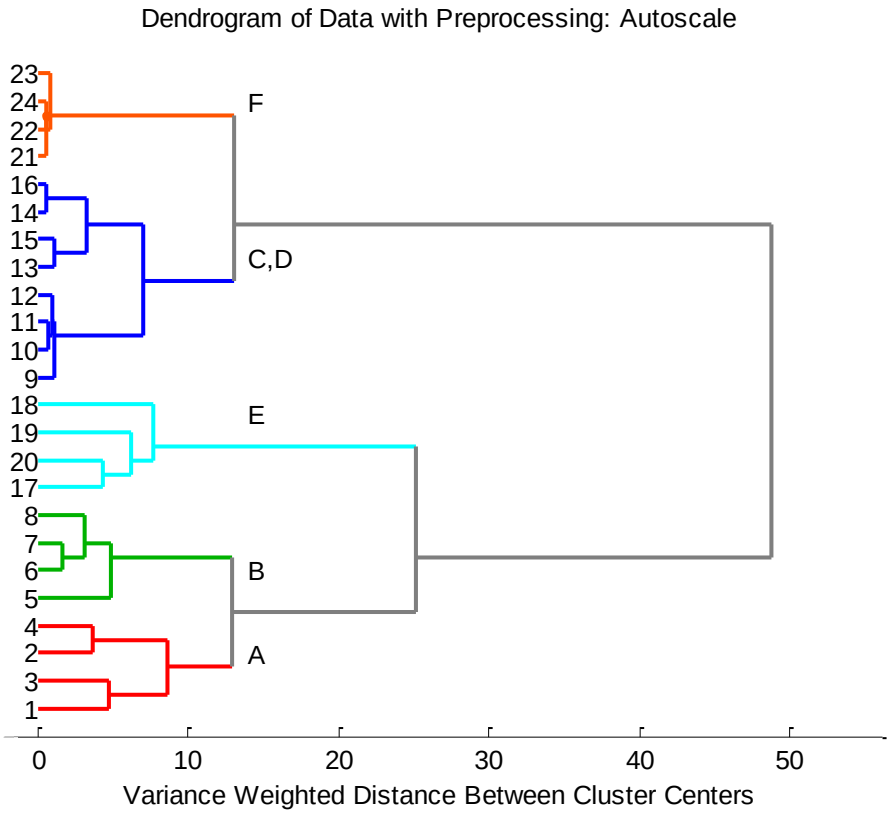

Figure 5. HCA of histograms for drugs **A-F**.

In Figure 5, the images are identified 1 to 24 divided in 6 groups of 4 replicates. The images of the same drug formulation have been clustered due the similarity of the profiles of histograms for each product. This pattern is repeated in all images analyzed. However, it is less evident for drugs **C** (Figure 5 samples 9 to 12) and **D** (Figure 5 samples 13 to 16). Those drugs have the same specification and brand but different package characteristics.

In this direction, the dendrogram obtained by HCA shows that this is an appropriate procedure for classification by multivariate similarity. However, more heterogeneous samples (flat histograms) are also less reproducible for the same image size. It is observed through the node of link of the clusters. In

contrast, images of the original product **F** (Figure 5 samples 21 to 24) showed a much better homogeneity and repeatability among images.

Conclusions

In this work, it was explored the technique of chemical mapping by near infrared spectroscopy in drug formulations with tablets of different sources in situations where the whole composition is not known. In such cases, it is very difficult to establish an appropriate calibration technique, so that only the information of sildenafil is considered independently of the excipients. The results obtained suggest an important way in which Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS), by using the advantage of second order, can be successfully applied. Furthermore, the images of the active ingredient have been adequately explored by unsupervised classification of their histograms. In conclusion, the pattern recognition for homogeneity of images was possible. Thus, the presented methodology may be useful both for development of product and process, as well as, identification of the counterfeit.

Reference

(1) Tinnemann, P.; Ozbay, J.; Saint V. A.; Willich S. N. *PLoS ONE* 2010, 5, e14059.

(2) Tsai, B. H. *Issues Stud.* 2010, 46, 173-203.

(3) Rastogi, T. *J. Intel. Prop. Rig.* 2010, 15, 302.

(4) Chapman, A. R. J. *Inter. Econ. Law* 2002, 5, 861-882.

(5) Pomara, G.; Morelli, G.; Pomara, S.; Taddei, S.; Ghiadoni, L.; Dinelli, N.; Travaglini, F.;Dicuio, M.; Mondaini, N.; Salvetti, A.; Selli, C. *J. Androl.* 2004, 25, 625-629.

(6) Gousse, A. E.; Lambert, M.; Kester, R. *Top. Spinal Cord Inj. Rehabil.* 2002, 8, 51-62.

(7) Yang, Y. J.; Song, D. M.; Jiang, W. M.; Xiang, B. R. *Anal. Letters* 2010, 43, 373-380.

(8) Abourashed, E. A.; Abdel-Kader, M. S.; Habib, A. A. M. *J. Planar. Chromatogr. - Mod. TLC* 2005, 18, 372-376.

(9) Cai, Y.; Cai, T. G.; Shi, Y.; Cheng, X. L.; Ma, L. Y.; Ma, S. C.; Lin, R. C.; Feng, W. *J. Liq. Chromatogr. Related Technol.* 2010, 33, 1287-1306.

(10) Berzas, J. J.; Rodríguez, J.; Villaseñor, M. J.; Contento, A. M.; Cabello, M. P. *Chromatographia* 2002, 55, 601-606.

(11) Tyszczuk, K.; Korolczuk, M. *Bioelectrochem.* 2010, 78, 113-117.

(12) Stefan-van Staden, Raluca-Ioana; van Staden, J. F.; Aboul-Enein, H. Y. *J. Solid State Electrochem.* 2010, 14, 997.

(13) Issa, Y. M.; El-Hawary, W. F.; Youssef, A. F. A.; Senosy, A. R. *Spectrochim. Acta A*, 2010, 75, 1297-1303.

(14) Amin, A. S.; Moustafa, E.; El-Dosoky, R. M. S. *J. AOAC Intern.* 2009, 92, 125-130.

(15) Rocha, W. F. C.; Rosa, A. L.; Martins, J. A.; Poppi, R. J. *J. Braz. Chem. Soc.* 2010, 21, 1904-1912.

(16) Rocha, W. F. C.; Rosa, A. L.; Martins, J. A.; Poppi, R. J. *J. Mol. Struct.* 2010, 982, 73-78.

(17) US Food and Drug Administration (FDA), Guidance for Industry, PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, FDA (CDER), 2004. Acessed in <www.fda.gov/cder/OPS/PAT.htm>.

(18) Kokaly, R. F.; Asner, G. P.; Ollinger, S. V.; Martin, M. E.; Wessman, C. A. *Remote Sens. Environ.* 2009, 113, 78–91.

(19) Sabins, F.F. *Geol. Rev.* 1999, 14, 157-183.

(20) Gao, B. C.; Montes, M. J.; Davis, C. O.; Goetz, A. F. H. *Remote Sens. Environ.* 2009, 113, 17-24.

(21) Verrelst, J.; Schaepman, M. E.; Malenovsk, Z.; Clevers, J. G. P. W. *Remote Sens. Environ.* 2010, 114, 647-656.

(22) Sommer, S.; Hill, J.; Mégier, J. *Agric. Ecosyst. Environ.* 1998, 67, 197–209.

(23) Gowen, A. A.; O'Donnell, C. P.; Cullen, P. J.; Downey, G.; Frias, J. M. *Trends Food Sci. Technol.* 2007, 18, 590-598.

(24) Gendrin, C.; Roggo, Y.; Collet, C. *Talanta* 2007, 73, 733-741.

(25) Roggo, Y.; Edmond, A.; Chalus, P.; Ulmschneider, M. *Anal. Chim. Acta* 2005, 535, 79-87.

(26) Ravn, C.; Skibsted, E.; Bro, R. J. *Pharm. Biomed. Anal.* 2008, 48, 554-561.

(27) Amigo, J. M.; Cruz, J.; Bautista, M.; Maspoch, S.; Coello, J.; Blanco, M. *TrAC, Trends Anal. Chem.* 2008, 27, 696-713.

(28) de Juan, A.; Tauler, R. *Anal.Chim. Acta* 2003, 500, 195-210.

(29) Tauler, R.; Smilde, A. K.; Kowalski, B. R. J. *Chemom.* 1995, 9, 31-58.

(30) Tauler, R. *Chemom. Intell. Lab. Syst.* 1995, 30, 133-146.

(31) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.* 2005, 76, 101-110.

(32) Windig, W.; Stephenson, D. A. *Anal.Chem.* 1992, 64, 2735-2742.

(33) Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbuhler, A. D. *Talanta* 1986, 33, 943-951.

(34) Maeder, M. *Anal. Chem.* 1987, 59, 527-530.

(35) Sanchez, F. C.; Toft, J.; Vanden Bogaert, B.; Massart, D. L. *Anal. Chem.* 1996, 68, 79-85.

(36) Sanchez, F. C.; Vandeginste, B. G. M.; Hancewicz, T. M.; Massart, D. L. *Anal. Chem.* 1997, 69, 1477-1484.

(37) Windig, W.; Guilment, J. *Anal. Chem.* 1991, 63, 1425-1432.

(38) Sabin G. P., Breitkreitz M. C., de Souza A. M., da Fonseca P., Calefe L., Moffa M., Poppi R.J. *Anal.Chim. Acta.* 2011, 706(1), 113-119.