

One-Class Classification Models for the Authentication of Analgesic Tablet Reference Medicine Using Differential Scanning Calorimetry and Visible-Near Infrared Spectroscopy

David Douglas S. Fernandes,^{1a} Cleildo P. Santana,^{1b} Felipe Hugo A. Fernandes,^{1c,d}
Hilthon A. Ramos,^{1e} Ana Cláudia D. Medeiros^{1c} and Germano Veras^{1*,e}

^aLaboratório de Automação e Instrumentação Analítica e Quimiometria,
Centro de Ciências Exatas e da Terra, Universidade Federal da Paraíba, Campus I,
Loteamento Cidade Universitária, 58051-900 João Pessoa-PB, Brazil

^bLaboratório de Química Analítica e Quimiometria, Centro de Ciências e Tecnologia,
Universidade Federal de Minas Gerais, R. Baraúnas, 351, Bodocongó,
58429-500 Belo Horizonte-MG, Brazil

^cLaboratório de Desenvolvimento e Ensaios de Medicamentos,
Centro de Ciências Biológicas e da Saúde, Universidade Estadual da Paraíba,
R. Baraúnas, 351, Bodocongó, 58429-500 Campina Grande-PB, Brazil

^dFACISA, R. Baraúnas, 351, Bodocongó, 58429-500 Campina Grande-PB, Brazil

^eLaboratório de Química Analítica e Quimiometria, Centro de Ciências e Tecnologia,
Universidade Estadual da Paraíba, R. Baraúnas, 351, Bodocongó,
58429-500 Campina Grande-PB, Brazil

The aim of this study was to develop a fast, simple and accurate analytical method for the classification of reference tablet analgesic drugs containing dipyron, orphenadrine and caffeine using differential scanning calorimetry (DSC) and visible-near infrared spectroscopy (VNIRS) associated with one-class chemometric classification algorithm. The training set is based on reference medicine with 15 samples as target class. Three different brands with 10 samples each and five reference medicine samples, obtaining 35 samples, were used as the test set. Chemometric models based on principal component analysis (PCA) and data-driven soft independent modelling of class analogy (DD-SIMCA) were used to obtain the results. Two DD-SIMCA models obtained 100% sensitivity, specificity, and accuracy using DSC and VNIRS, both with a significance level of 0.01. This method using one-class classification as a chemometric tool proved to be a good alternative for quality control of pharmaceutical samples.

Keywords: tablet, associate drugs, thermal analysis, vibrational spectroscopy, chemometric classification tools

Introduction

Many countries and their populations are harmed by the marketing of counterfeit medicines. The World Health Organization (WHO)¹ at the 65th World Health Assembly, approved a Member State Mechanism to prevent and control substandard/spurious/falsely-labelled/falsified/counterfeit medicines as a strategy to protect public health

and to promote access to affordable, safe, effective and quality medicines. It is estimated that more than 10% of medicines in low- and middle-income countries are substandard or falsified to an estimated revenue from falsified drugs of \$30.5 billion dollars.^{1,2} Special attention should be paid to analgesic medicines because more than 35% of all samples studied have failed in quality tests.¹ Two of the objectives to combat these problems are: to identify key challenges to develop national and regional capacities with appropriate methodologies for detection and control of “substandard/spurious/falsely-labelled/falsified/counterfeit

*e-mail: germano@servidor.uepb.edu.br

Editor handled this article: Ivo M. Raimundo Jr. (Associate)

medical products"; and to strengthen regulatory capacity and quality control laboratories for developing countries.

Pisani *et al.*³ identified market risks for falsified or substandard products. The authors interviewed regulators, policy-makers, pharmaceutical manufacturers, physicians, pharmacists, patients and academics in selected middle-income countries, namely China, Indonesia, Turkey and Romania. The responses of the questionnaires enabled identification of three large groups of market scenarios: (i) industry protects profit margins by reducing costs; (ii) industry protects profit margins by avoiding unprofitable products or markets; (iii) industry and healthcare providers promote profitable products. Based on responses, the authors concluded that organizations and governments must consider developing industrial, environmental and trade policies on the quality of medicines. In other words, fair prices would prevent consumers from seeking products without quality assurance but which have lower prices. The results obtained by the authors should now be extrapolated to other countries, mainly those without regulation of the commercialization of medicines and quality public healthcare.

Sweileh⁴ performed a bibliometric analysis of scientific production regarding substandard and falsified drugs. The author considers substandard and falsified drugs a crime against humanity based on observance of the significant growth of the problem over the three decades evaluated (1960 to 1990). The countries that were found to have published most articles about substandard and falsified drugs were mainly the developed countries (United States of America, United Kingdom, Germany, Belgium, France, Switzerland, Netherlands, Australia and Italy) and only two middle-income countries (India and China).

Rahman *et al.*,⁵ on the other hand, identified by country those articles that reported incidents involving damage to health due to falsified medicines between 1972 and 2017. A total of 81 articles with 48 incidents related to falsified medicines were founded. In this study, 56.3% of related cases were obtained in developing countries (low and middle income) and 47.3% in developed countries, indicating that the distribution of this problem was regardless of the economic or social development of the countries named. In a similar study, Koczwara and Dressman⁶ found 41 articles related to counterfeit medicines between the years 2007 and 2016. In their study, articles with counterfeit or falsified medicines were only found from low or middle-income countries, except for two higher income countries (Japan and USA), but these with internet marketing to the whole world.

McManus and Naughton⁷ conducted a systematic review of substandard, falsified, unlicensed and unregistered medicine sampling studies in the years 2013 to 2018. In

this paper, the researchers assessed the type of drug-related problems and six categories were found: four associated to the amount of active ingredients (inadequate, missing or other substance and excess) and two in terms of more global problems such as impurity and dissolution failure.

The quality control methods for drugs are, as presented above, based on quantitative analysis, verifying the quantified active constituents and comparing these with the label information, as described in the pharmacopoeias, among them United States⁸ and British⁹ pharmacopoeias. Missing, however, is a more complete drug analysis that considers not only the correct presence of the active ingredients listed, but also the interactions among these constituents and the excipients and the consistency in quality of the drugs across different batches or brands. This information is not yet available from national control agencies, such as FDA (Food and Drugs Administration) in the USA or ANVISA (Agência Nacional de Vigilância Sanitária) in Brazil.

The procedure of quantification of an available drug by observing only the active ingredients was not significantly problematic up to the end of the 1960s. Excipients, for example, were considered as inert substances³ and analytical instruments did not have the capability to obtain a large amount of data. Furthermore, considering that quantitative analysis is expensive and cannot assess the level of interaction between substances in drugs and therefore record the consistency of drugs across different batches or brands, there is a need for analytical screening techniques such as differential scanning calorimetry (DSC) and near infrared spectroscopy (NIRS).

Rebiere *et al.*¹⁰ published a review on analytical techniques that identified specific information about the organic and inorganic composition, the presence of an active substance or impurities, or the crystalline arrangement of a compound formulation that provides useful information for identifying problems such as counterfeiting or lack of quality in drugs.

Khanmohammadi *et al.*¹¹ proposed a successive projection algorithm-partial least squares (SPA-PLS) chemometric model to quantify codeine and paracetamol in pharmaceutical tablets. The analytical technique used was thermogravimetric analysis (TGA); similar results were obtained for the reference high-performance liquid chromatography (HPLC) method. Thus, the thermogravimetric analysis (TGA) analytical method coupled with the SPA-PLS chemometric model provided a simple, rapid and reliable method of analysis without the need for sample preparation or extraction.

Modern analytical instruments obtain a huge amount of data, which can generate a very great amount of sample

data, and consequently evaluate a drug comprehensively. Preliminary assessment of drug quality requires pattern recognition techniques as a chemometric tool to treat this volume of analytical data.¹²⁻¹⁵

Lawson¹² developed a low cost and rapid analytical method and easy interpretation of pharmaceutical ingredient screening results for the rapid identification of substandard and falsified paracetamol drugs using attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy. In terms of results, the determination of paracetamol was obtained using the partial least square regression (PLSR) equation obtained from the calibration data using the range of 1524-1493 cm^{-1} ; 12% of the tablet samples were identified as substandard.

Storme-Paris *et al.*¹³ studied different excipients in drugs with ciprofloxacin and fluoxetine as the active pharmaceutical ingredients using NIRS as an analytical method. In terms of chemometric models, principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA) were used in six cases studied, and satisfactory results were obtained.

Rodionova *et al.*¹⁴ analyzed counterfeit fluconazole capsules using NIRS as an analytical technique and PLSR as a chemometric tool to quantify the active pharmaceutical ingredient, and SIMCA was used for authentication with a specificity equal to 94%.

Rebiere *et al.*¹⁵ used a multi-analytic approach to determine the manufacturing process factors of omeprazole drugs. The analytical methods used were gas chromatography-mass spectroscopy (GC-MS), NIRS, nuclear magnetic resonance (NMR), and X-ray powder diffraction. The chemometric models were hierarchical cluster analysis (HCA), PCA, SIMCA and PLS-discriminant analysis (PLS-DA). The authors concluded that NMR and XRPD were adequate in differentiating samples from 9 of the 11 manufacturers.

Santos *et al.*¹⁶ used HCA to discriminate between counterfeit and authentic sildenafil and tadalafil drugs with differential scanning calorimetry (DSC) coupled to HCA as the chemometric technique. The results showed different heat flow profiles between authentic and counterfeit drugs and similarity between the active pharmaceutical ingredients and each drug.

Despite all these publications involving the classification of pharmaceutical samples, few have been published using one-class classification methods. These methods distinguish objects of a particular class from all other objects and are used to detect adulteration or authentication of samples. The most commonly used one-class classification technique is data-driven soft independent modelling of class analogy (DD-SIMCA).¹⁷⁻¹⁹

Pomerantsev and Rodionova¹⁷ proposed the DD-SIMCA chemometric method to determine extreme or outlier samples by observing appropriate thresholds. This method was applied to simulated and to real data; in this case it was for the identification of outliers in a taurine pharmaceutical substance packed in closed polyethylene (PE) bags, using NIR spectroscopy. Zontov *et al.*¹⁸ proposed an easy way of establishment and employment of data to build a chemometric model using DD-SIMCA.

Ciza *et al.*¹⁹ compared the performances of different portable NIR and Raman spectrometers for the detection of a group of falsified drugs: artemether-lumefantrine, paracetamol and ibuprofen. The chemometric models used were HCA, DD-SIMCA, and hit quality index (HQI). The authors concluded that portable NIR and Raman spectrometers are promising tools for the identification of substandard and falsified drugs.

Analgesics, which have a large consumer market, are among the drugs that can be studied by pattern recognition techniques. These drugs are among the most falsified, either in labeling, quantification of active ingredients, or in changing excipients in the pharmaceutical formulation, without bioequivalence and bioavailability studies, or drugs without legal registration to operate on the market. Given these findings, the aim of our study was to develop a new analytical method for rapid, simple and accurate classification of dipyron, orphenadrine and caffeine-containing drugs by identifying and grouping them. This study used DSC and visible-near infrared spectroscopy combined with chemometric DD-SIMCA one-class classification technique.

Experimental

Samples

Tablets containing dipyron (300 mg), caffeine (50 mg) and orphenadrine (35 mg) were analyzed utilizing 50 drug samples purchased in the Northeastern Brazilian states of Ceará and Paraíba. In this work four different brands (C1 to C4) of tablets were analyzed. In C2 to C4, ten different lots were analyzed; and in C1, twenty different lots were analyzed. C1 is a class with samples of reference drugs. C2 to C4 are samples of generic drugs.

Differential scanning calorimetry (DSC)

The DSC curves were obtained on a TA Instruments Calorimeter, model DSC Q20 (New Castle, USA), using aluminum crucibles with about 2 ± 0.1 mg of samples under nitrogen atmosphere, at the flow of 50 mL min^{-1} .

Rising temperature experiments were conducted in the temperature range from 30 to 400 °C and a heating rate of 10 °C min⁻¹. Indium (mp 156.6 °C) was used as standard for equipment calibration. Data were analyzed using the software TA Instruments Universal Analysis 2000, 4.7 A.

Visible-near infrared spectroscopy (VNIRS)

A small fraction of each subsample was placed in a sample-holder for diffuse reflectance analysis, without any previous sample treatments or use of chemical reagents. The XDS Rapid Content™ Analyzer (FOSS, Hilleroed, Denmark) was the chosen instrument, with 0.5 nm spectral resolution, equipped with holographic net and Si and PbS detection systems. Sample spectra were obtained from both sides of each of the tablets in the spectral range from 400 to 2500 nm.

Chemometric study

The data were partitioned into training (15 samples) and test (35 samples) sets using the Kennard-Stone (KS) algorithm.²⁰ Then, both data, DSC and VNIRS, were preprocessed using baseline offset correction. The training set was used to build the data driven soft independent modelling of class analogy (DD-SIMCA) models. The test set of samples was employed to evaluate the quality of the chemometric models built.

The samples were separated using the KS algorithm as the sample selection technique with two sets: a training set containing 15 samples of C1 brand, and a test set with 5 samples of the C1 brand and other brands (10 each), making 35 samples.

Data preprocessing, sample selection using the KS algorithm and the DD-SIMCA analysis were performed in the MatLab® environment R2011a.²¹

Results and Discussion

Exploratory analysis

Figure 1a illustrates all tablet samples showing three events, with two endothermic peaks and one exothermic peak. The first event can be attributed to the volatile constituents of the sample losses, while the second and third peaks are related to the phase transition process and component decomposition. For the first event, samples from four brands showed similar characteristics. The endothermic peaks showed an average temperature of 107.64, 108.65, 104.03 and 106.61 °C. The small variations in these temperatures can be attributed to the fact that water loss is an event which occurs similarly in samples with similar composition. The exothermic peak shows thermal decomposition processes with average temperatures of 222.90, 224.57, 220.81 and 224.72 °C. The process was probably due to the early decomposition of dipyrone,

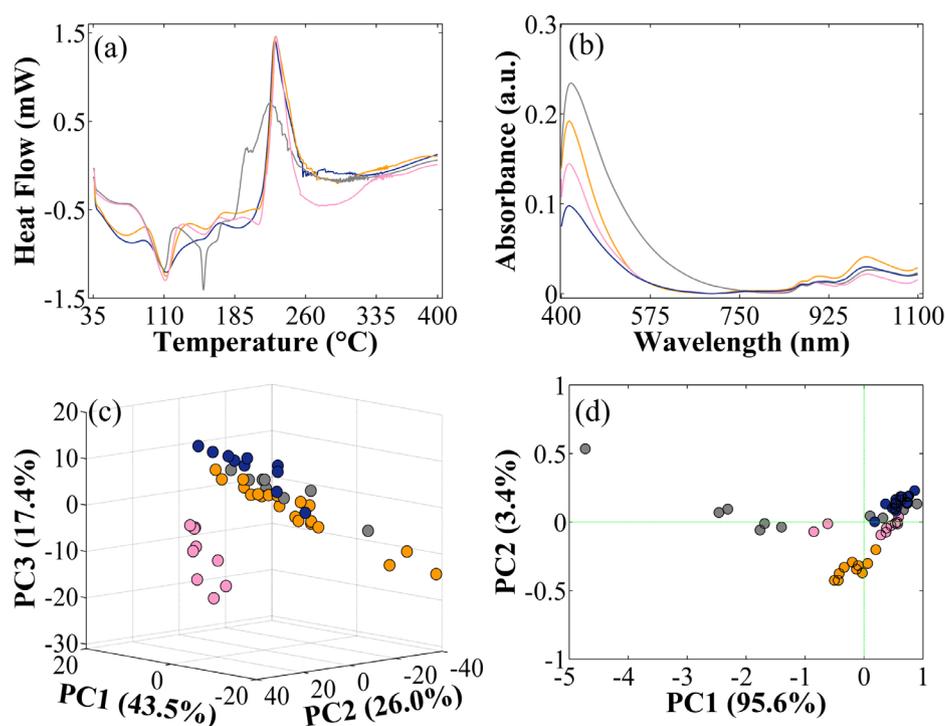


Figure 1. Analytical data: (a) average DSC curves; (b) average VNIRS spectra, and PCA scores associated to: (c) DSC data and (d) VNIRS data. Brands: C1 (blue), C2 (orange), C3 (gray) and C4 (pink).

since this drug shows an exothermic peak, characteristic of decomposition at 245.55 °C. For illustration, Figure 1b shows the average VNIRS spectra of C1 (blue), C2 (orange), C3 (gray) and C4 (pink) in the 400 to 1100 nm range. There is a similarity between the molecular absorption bands that are displayed in the same spectral regions.

In Figure 1c, score plot of PC1 (43.5%) \times PC2 (26.0%) \times PC3 (17.4%) of the DSC data, indicating that no clusters had been formed among the study drug samples (C1, C2, C3, and C4). For this reason, supervised pattern recognition tools were used for the classification of the remaining drugs. The PCA scores are illustrated in Figure 1d, VNIRS data with PC1 (95%) *versus* PC2 (3%), corroborate with the information displayed in the heat flow curve, Figure 1b, where C3 has a well-defined cluster due to the composition of the excipients.

After building the training models containing only the samples from the reference drug, the test set was evaluated. The model performance is shown in Figure 2. In Figures 2a and 2d, all training samples behave similarly to regular samples of the reference target. All training samples that lie within the bounds of the blue line set, considering a α -value of 0.05, are shown in blue circles, as can be seen in the acceptance plots in Figures 2a and 2d. No sample with anomalous behavior was found. Both models (0.01 and 0.05) showed a sensitivity of 100%, which indicates the absence of false negative samples. All reference analgesic tablets were recognized within the region corresponding to the acceptance area.

The DD-SIMCA model at the 0.01 significance level showed a high ability (Figure 2c) to recognize reference samples and differentiate between the outlier samples and the other brands (C2, C3 and C4). The DD-SIMCA model with significance level (0.05) was similar; however, it showed only one case of false negative classification. Only one sample of the training set was not recognized as belonging to the target class. It is important to note that misclassified samples cannot be included in non-target samples; they are positioned near the boundary of the acceptance area.

Table 1 shows the figures of merit for the DD-SIMCA models in the prediction phase. Both models achieved acceptable results. The former showed a correct rating (sensitivity = 100%, specificity = 100% and accuracy = 100%) and (sensitivity = 80%, specificity = 100% and accuracy = 97.1%) for the 0.01 and 0.05 significance levels, respectively.

DD-SIMCA stands out as presenting robust estimators by virtue of its approach, which is data-driven. This establishes that the empirical distribution method of the dataset contributes to improving the model's help on the data. The models allowed the recognition of reference samples with high sensitivity. Thus, based on the results obtained in this study, we suggest that the combination of VNIRS data with class models based on the DD-SIMCA method are able to recognize reference analgesic tablet samples. These results may be useful for quality control in pharmaceutical industries that manufacture a wide range of reference and non-reference drugs.²²

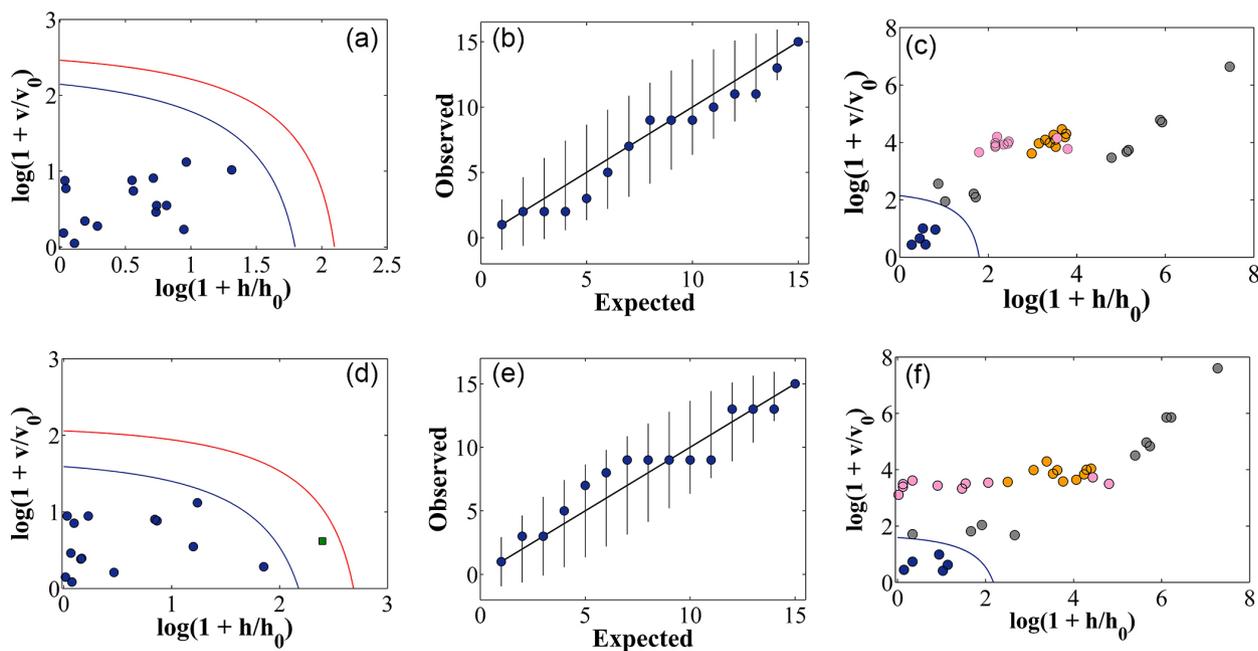


Figure 2. (a) Acceptance graph of the training samples for VNIRS data, (b) extremes graph and (c) acceptance graph of the test samples using statistical significance of 0.01, (d) acceptance plot for the training samples, (e) extremes plot, and (f) acceptance plot for the test samples using a statistical significance of 0.05.

Table 1. Classification results DD-SIMCA for analgesic tablets using DSC curves and VNIRS spectra

	DSC (0.01) ^a		DSC (0.05) ^a		VNIRS (0.01) ^a		VNIRS (0.05) ^a	
	Training	Test	Training	Test	Training	Test	Training	Test
Factors	3	3	3	3	2	2	1	1
Reference	15/15	5/5	15/15	4/5	15/15	5/5	14/15	5/5
No-reference	–	30/30	–	30/30	–	30/30	–	30/30
Sensitivity / %	100	100	100	80.0	100	100	93.3	100
Specificity / %	–	100	–	100	–	100	–	100
Accuracy / %	100	100	100	97.1	100	100	93.3	100

^aSignificance level. Factors: latent variables or principal component. DSC: differential scanning calorimetry; VNIRS: visible-near infrared spectroscopy.

Conclusions

A fast, simple and accurate method was developed and validated using DSC curves and VNIRS spectra coupled with class pattern recognition models, namely DD-SIMCA to differentiate reference from non-reference drugs. The best result was obtained for DD-SIMCA from the VNIRS spectra at 0.01 level of significance, which achieved 100% sensitivity and accuracy, respectively, on the training set and 100% sensitivity, specificity, and accuracy on the test set, respectively. A similar result was obtained with DSC. Taking into account the time for analysis, cost, and the cheapest instrumentation, however, the VNIRS was the best method to identify reference drugs containing dipyron, orphenadrine and caffeine. From this perspective, the results of this study may be used as screening and quality control of reference drugs.

Acknowledgments

Germano Veras acknowledges the support of CNPq, Brazil (Research Fellowship 313579/2021-0) and David Douglas de Sousa Fernandes thanks to Fundação de Apoio à Pesquisa do Estado da Paraíba (Research Fellowship 1271/2021). The authors gratefully acknowledge use of the TA Instruments Calorimeter of the CERTBIO/UEPB. The English text of this paper has been revised by Sidney Pratt, Canadian, MAT (The Johns Hopkins University), RSAdip - TESL (Cambridge University).

Author Contributions

David D. S. Fernandes was responsible for data curation, formal analysis, investigation, software; Cleildo P. Santana for investigation, methodology; Felipe Hugo A. Fernandes for data curation investigation, methodology; Hilthon A. Ramos for writing-original draft writing-review and editing; Ana Cláudia D. de Medeiros for conceptualization, methodology; Germano Veras for corresponding author, conceptualization, data curation, investigation, project

administration, supervision, writing - original draft, review and editing.

References

- World Health Organization (WHO); *Substandard/Spurious/Falsely-Labelled/Falsified/Counterfeit Medical Products*; WHO: Geneva, Switzerland, 2012. [Link] accessed in July 2022
- Nayyar, G. M. L.; Breman, J. G.; Mackey, T. K.; Clark, J. P.; Hajjou, M.; Littrell, M.; Herrington, J. E.; *Am. J. Trop. Med. Hyg.* **2019**, *100*, 1058. [Crossref]
- Pisani, E.; Nistor, A. L.; Hasnida, A.; Parmaksiz, K.; Xu, J.; Kok, M. O.; *Wellcome Open Res.* **2019**, *16*, 4. [Crossref]
- Sweileh, W. M.; *Global Health* **2021**, *17*, 114. [Crossref]
- Rahman, M. S.; Yoshida, N.; Tsuboi, H.; Tomizu, N.; Endo, J.; Miyu, O.; Akimoto, Y.; Kimura, K.; *Trop. Med. Int. Health* **2018**, *23*, 1294. [Crossref]
- Koczwara, A.; Dressman, J.; *J. Pharm. Sci.* **2017**, *106*, 2921. [Crossref]
- McManus, D.; Naughton, B. D.; *BMJ Global Health* **2020**, *5*, e002393. [Crossref]
- The United States Pharmacopoeia (USP)*, USP 36-NF 31, The United States Pharmacopoeial Convention, Rockville, 2013, p. 2283-4029.
- British Pharmacopoeia*, vol. 1, The Stationery Office, 2009.
- Rebiere, H.; Guinot, P.; Chauvey, D.; Brenier, C.; *J. Pharm. Biomed. Anal.* **2017**, *142*, 286. [Crossref]
- Khanmohammadi, M.; Soleimani, M.; Morovvat, F.; Garmarudi, A. B.; Ghasemi, M.; Khalafbeigi, K.; *Thermochim. Acta* **2012**, *530*, 128. [Crossref]
- Lawson, G.; Ogwu, J.; Tanna, S.; *PLoS One* **2018**, *10*, e0202059. [Crossref]
- Storme-Paris, I.; Rebiere, H.; Matoga, M.; Civade, C.; Bonnet, P. A.; Tissier, M. H.; Chaminade, P.; *Anal. Chim. Acta* **2010**, *658*, 163. [Crossref]
- Rodionova, O. Y.; Titova, A. V.; Demkin, N. A.; Balyklova, K. S.; Pomerantsev, A. L.; *Talanta* **2019**, *195*, 662. [Crossref]
- Rebiere, H.; Grange, Y.; Deconinck, E.; Courselle, P.; Acevska, J.; Brezovska, K.; Maurin, J.; Rundlöf, T.; Portela, M. J.; Olsen,

- L. S.; Offerlé, C.; Bertrand, M.; *J. Pharm. Biomed. Anal.* **2022**, *208*, 114444. [Crossref]
16. dos Santos, M. K.; Mariotti, K. C.; Kahmann, A.; Anzanello, M. J.; Ferrão, M. F.; Gomes, A. A.; Limberger, R. P.; Ortiz, R. S.; *J. Pharm. Biomed. Anal.* **2019**, *166*, 304. [Crossref]
17. Pomerantsev, A. L.; Rodionova, O. Y.; *J. Chemom.* **2014**, *28*, 429. [Crossref]
18. Zontov, Y. V.; Rodionova, O. Y.; Kucheryavskiy, S. V.; Pomerantsev, A. L.; *Chemom. Intell. Lab. Syst.* **2017**, *167*, 23. [Crossref]
19. Ciza, P. H.; Sacre, P. Y.; Waffo, C.; Coïc, L.; Avohou, H.; Mbinze, J. K.; Ngono, R.; Marini, R. D.; Hubert, P.; Ziemons, E.; *Talanta* **2019**, *202*, 469. [Crossref]
20. Galvão, R. K. H.; Araújo, M. C. U.; José, G. E.; Pontes, M. J. C.; Silva, E. C.; Saldanha, T. C. B.; *Talanta* **2005**, *67*, 736. [Crossref]
21. *MATLAB*, version: 7.12; The MathWorks, Inc.: Natick, MA, USA, 2011.
22. Gomes, A. A.; Khvalbota, L.; Machyňáková, A.; Furdíková, K.; Zini, A. C.; Špánik, I.; *Spectrochim. Acta, Part A* **2021**, *257*, 119770. [Crossref]

Submitted: March 31, 2022

Published online: July 15, 2022

