Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 202 (2018) 359–367

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Advanced chemometrics manipulation of UV-spectroscopic data for determination of three co-formulated drugs along with their impurities in different formulations using variable selection and regression model updating

Maha A. Hegazy [a], Shereen A. Boltia [a,*], Ahmed S. Fayed [a], Awadh Musaed [b]

[a] Analytical Chemistry Department, Faculty of Pharmacy, Cairo University, Kar Al aini Street, 11562 Cairo, Egypt
[b] Analytical chemistry department, Faculty of pharmacy, Aden University, Republic of Yemen

ARTICLE INFO

Article history:
Received 28 December 2017
Received in revised form 28 April 2018
Accepted 9 May 2018
Available online xxxx

Keywords:
Ambroxol
Guaifenesin
Theophylline
PLS
iPLS
siPLS

ABSTRACT

Multivariate calibration models manipulating UV-spectroscopic data of three anti-productive cough drugs namely ambroxol, guaifenesin and theophylline were constructed for the intent of simultaneous determination in presence of their impurities; guaiacol and caffeine. Both interval partial least squares (iPLS) and synergy interval partial least square (siPLS) algorithms were adopted for variables selection to extract useful information and improve the models' performance. The optimal spectral range and their combinations were assigned according to the lowest value of Root Mean Square Error of Prediction (RMSEP), Standard Error of Prediction (SEP) and Correlation Coefficient (R^2). The results obtained from full spectrum PLS were compared with those obtained by iPLS and siPLS. The siPLS method exhibited better performance. The combination of four subintervals, 2, 9, 13, and 16, showed the best effect, with RMSEP of 0.1039, 0.3548 and 0.207 μg/mL, for ambroxol, guaifenesin and theophylline, respectively and correlation coefficient of 0.9999, 0.9975 and 0.9994 for ambroxol, guaifenesin and theophylline, respectively. The proposed methods were used for the simultaneous determination of the three drugs in presence of their impurities in bulk powder and in pharmaceutical formulation.

1. Introduction

Ambroxol hydrochloride (AMB) is chemically designated as *trans*-4-[(2-amino-3,5-dibromobenzyl)amino] cyclohexanol hydrochloride [1], it is bromhexine metabolite and used with as mucolytic [2]. Guaifenesin (GUA) is used for productive cough with reports of increasing volume and reducing viscosity of tenacious sputum [2], its chemical name is (2*RS*)-3-(2-methoxyphenoxy) propane-1,2-diol [1]. Theophylline (THEO) is a xanthine [2] with a chemical name; 1,3-dimethyl-3,7-dihydro-1*H*-purine-2,6-dione [1], it is used as bronchodilator for managing reversible airway obstruction [2].

Literature survey revealed that several methods were reported for determination of each AMB, GUA, and THEO along with other drugs or with each other as binary mixtures, these reports includes; determination of AMB and GUA by high performance liquid chromatography (HPLC) [3–9], thin layer chromatography (TLC) and multivariate calibration [4]. On the other hand, GUA and THEO had been determined by HPLC [10–12], multivariate spectrophotometric methods [10,13], univariate spectrophotometric methods [12,14]. While, only two reports were found for determination of AMB, GUA and THEO by univariate spectrophotometric methods [15] and HPLC and multivariate calibration [16].

Guaiacol (GUAIA) which is 2-methoxyphenol is reported to be GUA impurity A according to BP [1]. Caffeine (CAFF), is 1,3,7-trimethyl-3,7-dihydro-1*H*-purine-2,6-dione is reported in BP as impurity A of THEO [1].

No reports were found for determination of the three dugs along their impurities, so the aim of our work is to investigate the advantage of synergistic selection of variables prior to multivariate calibration in resolution of very complex mixture with spectrally similar bands. Wavelength selection was performed by interval and synergy interval partial least squares.

2. Theoretical Background

2.1. Partial Least Squares

PLS is a well-established full spectrum factor based multivariate calibration algorithm having the advantage of correlating the scores and loadings of both response and concentration matrices [17].

* Corresponding author.
E-mail address: sahab333@yahoo.com (S.A. Boltia).

https://doi.org/10.1016/j.saa.2018.05.038
1386-1425/

2.2. Interval Partial Least Squares (iPLS)

iPLS splits the response matrix into a number of equidistant intervals which is variablewise and a correlation between samples spectra and quality parameters y in PLS models to find the lowest RMSECV and highest correlation coefficients. Local PLS models for each spectral subintervals of equal width are developed. The prediction performance of these local models is compared with the full spectrum model "global model", by validation parameters mainly RMSECV, r^2, slope and intercept to ensure a comprehensive model overview. The global PLS model is usually more complex than the optimal local models because of the presence of noisy and non-correlated variables in it. iPLS method gives an overview of the data and helps in the interpretation.

2.3. Synergy Interval Partial Least Squares (siPLS)

siPLS divides the data sets into a number of equidistant intervals which is variable wise, and then calculates all possible PLS models from combination of two, three or four intervals. A large number of models are processed depending on the number of intervals and the selected number of intervals to be combined. The results are represented automatically as number of PLS components, intervals combinations, and RMSECV for best models according to original number of intervals. Selection of a model with proper number of latent variables "7 in this situation" from the automatic represented best models, done by application on the independent validation set.

2.4. Regression Model Updating

Multivariate calibration models can be used for prediction of new samples as long as they are in the calibration samples range and not subjected to new sources of variance. If so, calibration model need to be updated for prediction of new samples containing unmodeled source of variation.

Model update is based on including new samples with new source of data variance into developed calibration model, expressed as follow:

$$X_{upd} = \begin{bmatrix} X \\ X_{new} \end{bmatrix} \quad Y_{upd} = \begin{bmatrix} Y \\ Y_{new} \end{bmatrix}$$

The number of newly added samples can be small so that these samples may not have enough weight compared to the initial calibration set. To solve this problem a proposed method by giving them higher weights, this approach also called the weighting scheme method [18–21].

2.5. Statistical Data Analysis

Chemometric models performance and prediction ability was evaluated using RMSECV and RMSEP which are expression of the average error in the analysis of components in calibration and validation sets, with optimal number of latent variables. As shown in equations below:

$$RMSECV/P = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(c_i - c_{i,pred}\right)^2}$$

Another commonly used parameter for evaluation of prediction ability of multivariate calibration model is Relative Standard Error of Prediction (RSEP), calculated by the following equation:

$$RSEP\% = \sqrt{\frac{\sum_{i=1}^{n}\left(c_i - c_{i,pred}\right)^2}{\sum_{i=1}^{n} c_i^2}} \times 100$$

where c_i is the reference value and $c_{i,\ pred}$ is the predicted value of the analyte in i sample and n equal to number of samples used.

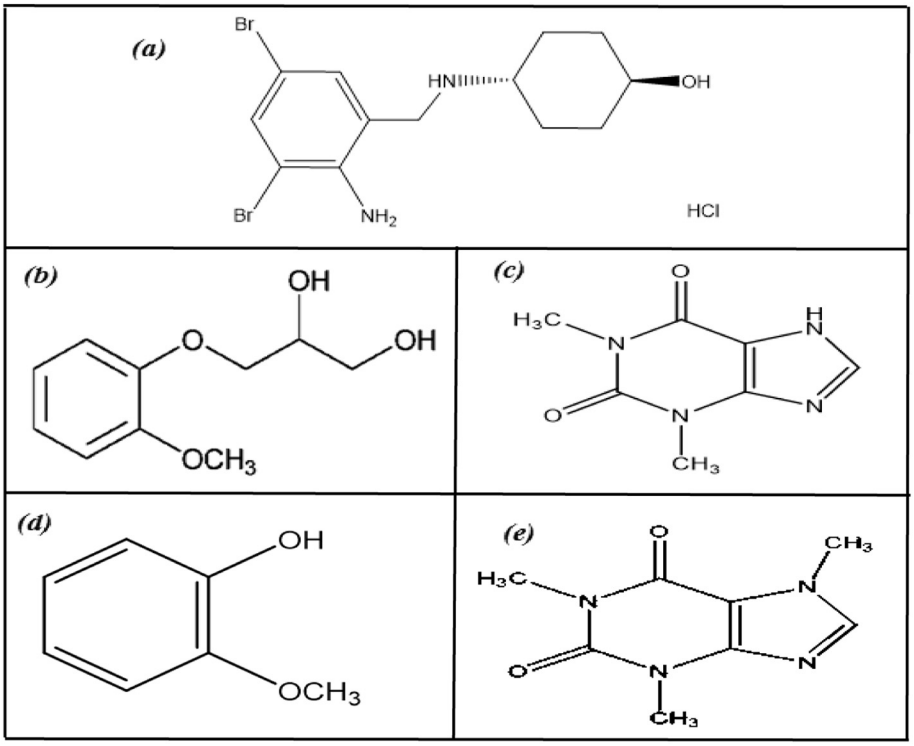

Fig. 1. Structures of (a) ambroxol HCl, (b) guaifenesin, (c) theophylline, (d) guaiacol and (e) caffeine.