Thermochimica Acta 526 (2011) 200–204

Contents lists available at SciVerse ScienceDirect

Thermochimica Acta

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

Thermogravimetric determination of L-ascorbic acid in non-effervescent formulations using multiple linear regression with temperature selection by the successive projections algorithm

Rosa Lucia Simencio Otero[a,d,1], Roberto Kawakami Harrop Galvão[b], Mário César Ugulino Araújo[c], Éder Tadeu Gomes Cavalheiro[d,*]

[a] CCDM-UFSCar, Via Washington Luiz, km 235 – Caixa Postal 60 13560-971, São Carlos/SP, Brazil
[b] Instituto Tecnológico de Aeronáutica–Divisão de Engenharia Eletrônica–CEP 12228-900, São José dos Campos/SP, Brazil
[c] Departamento de Química–Universidade Federal da Paraíba–Campus I–CEP 58051-970, João Pessoa/PB, Brazil
[d] Instituto de Química de São Carlos – Universidade de São Paulo – Av. do Trabalhador São-carlense, 400, CEP 13.566-590, São Carlos/SP, Brazil

ARTICLE INFO

Article history:
Received 9 June 2011
Received in revised form
11 September 2011
Accepted 21 September 2011
Available online 29 September 2011

Keywords:
Thermogravimetry
Multivariate calibration
Temperature selection
Successive projections algorithm
Analysis of pharmaceutical formulations

ABSTRACT

Changing experimental thermogravimetric conditions, such as heating rate and sample mass, may not be sufficient to eliminate the effect of interactions in pharmaceutical analysis. This motivates the investigation of a chemometric approach to determine active drugs in pharmaceutical formulations. The use of multiple linear regression (MLR) with temperatures selected by the successive projections algorithm (SPA) for determination of L-ascorbic acid (AA) in simulated non-effervescent formulations using microcrystalline cellulose as excipient is evaluated. For comparison, two other multivariate calibration methods, MLR with temperatures selected by genetic algorithm (GA) and partial least squares (PLS) using the entire range of temperatures were chosen. MLR–SPA provided the best predictions, in agreement with the expected AA concentration (correlation of 0.991 and root-mean-square error of 0.8%, m/m in the range 61.3–74.9%, m/m). No significant differences were found between the MLR–SPA values and those from iodimetric titration, according to *t*-test at 95% confidence level.

1. Introduction

Thermogravimetric (TG) curves present an essentially quantitative character which can be exploited for the determination of components in several kinds of samples provided that a change in mass occurs [1]. The main advantages of using TG data for the analysis of pharmaceutical formulations are the possibility of simultaneous determination of more than one component in the sample and the reduction in the volume of solvents and waste when compared to chromatographic procedures.

Despite the essentially quantitative character of thermogravimetry, its application for the determination of components present in pharmaceutical formulations may be hampered by physical–chemical interactions between components.

Radecki and Wesolowski [2,3] first investigated the interactions of different components of pharmaceutical formulations on the decomposition of pure pharmaceuticals and concluded that TG curves can be used for qualitative identification and quantification of the analyte. They also suggested that the thermal processes that allow these determinations are dehydration, decarboxylation and weight losses relative to the reaction of effervescent component mixtures, decomposition and formation of intermediates, or physical events of a component in the pharmaceutical formulation.

However interactions between the components of a sample can lead to an overlay of thermal events that cannot be distinguished by changing the experimental parameters such as heating rate, sample mass, nature and flow of the purge gas, etc. Such interactions are the result of chemical and/or physical processes that arise when drugs and excipients are mixed together, and are frequently observed in thermal analysis of pharmaceutical formulations [4]. Under these circumstances, the quantitative determination of the compounds involved in these thermal events cannot be achieved with the precision expected from a thermogravimetric measurement. Therefore, despite the number of investigations concerning the determination of pharmaceuticals by thermal methods [5–10], the interactions described above are still a limiting factor for many applications.

---

* Corresponding author. Tel.: +55 16 33738054; fax: +55 16 3373 9958.
E-mail address: cavalheiro@iqsc.usp.br (É.T.G. Cavalheiro).

[1] Escola de Engenharia de São Carlos – Departamento de Engenharia de Materiais, Aeronáutica e Automobilística – SMM – Universidade de São Paulo – Av. do Trabalhador São-carlense, 400, CEP 13.560–970, São Carlos/SP, Brazil.

0040-6031/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.tca.2011.09.020

Within this context, the use of chemometrics may be an interesting alternative for the treatment of TG data when the direct determination of the desired component is not possible due to interactions between the components of the formulation [11]. In the present work, a chemometric approach is proposed to improve the analysis of the drug in a mixture where interactions and overlapping thermal events occur.

For this purpose, multiple linear regression (MLR) is employed to obtain a multivariate calibration model with temperatures selected by the successive projections algorithm (SPA). SPA is a variable selection algorithm specifically designed to remove collinearity from the data set to improve numerical conditioning and reduce noise propagation in MLR models [12,13]. SPA has been successfully employed in several chemometric problems including: wavelength selection in ICP-OES [12], UV–vis [13], FTIR [14] and NIR spectrometry, as well as potential selection in voltammetry [15] and molecular descriptors in QSAR/QSPR studies [16].

SPA can be regarded as a forward selection method, in which the number of variables in the model is progressively increased from one up to a maximum value established by the analyst. In MLR with mean-centered data, the maximum number of variables cannot exceed the number of calibration samples minus one [17]. In SPA, each variable to be included in the model is chosen in order to display the least collinearity with respect to the previous ones. The initial variable and the overall number of variables are optimized in order to maximize the agreement between the MLR model predictions and the expected values over a given validation set [13].

The proposed approach was applied to the determination of L-ascorbic acid (AA) in ternary systems water/AA/microcrystalline cellulose (MC). Such TG determination is a challenging task in view of interactions and overlapping thermal events which are known to occur [8]. The performance of the MLR–SPA model is evaluated in terms of the prediction error with respect to the expected AA concentration values. For comparison, two other multivariate calibration techniques are also employed, namely MLR with temperatures selected by a genetic algorithm (GA) [18,19] and partial least squares (PLS) using the entire range of temperatures in the TG curve.

In order to assess the potential of the proposed approach as an alternative to classical methods for AA analysis, the MLR–SPA predictions are also compared with those obtained with the official iodimetric method [20], which is recommended in the Brazilian Pharmacopoea [21].

2. Experimental

2.1. Experimental design

The calibration mixtures were designed according to the XVERT algorithm [22] with the following ranges (%, m/m) for the component concentrations: water: 0.5–4.5; microcrystalline cellulose (MC): 19.5–42.5; L-ascorbic acid (AA): 55.0–79.5. These ranges were chosen to encompass the expected composition of typical commercial formulations. A two-level, three-component, mixture design including a central point yielded five calibration mixtures.

A set of four validation mixtures was designed for use in the temperature selection algorithms (MLR–GA and MLR–SPA) and in the determination of the number of factors in PLS. The composition of such mixtures was established by taking intermediate points between the extreme vertices and the central point of the XVERT design. Finally, four prediction mixtures were specified with random concentration values inside the calibration range. These mixtures were used to compare the models with respect to their predictive ability. For this reason, they were not employed in any stage of the model-building process.

Table 1
Composition of the calibration (C1–C5), validation (V1–V4), and prediction (P1–P4) mixtures (%, m/m).

	Water	MC	AA
C1	0.5	42.5	57.0
C2	4.5	19.5	76.0
C3	0.5	20.0	79.5
C4	4.5	40.5	55
C5	2.5	31.0	66.5
V1	1.5	25.0	73.5
V2	1.5	37.0	61.5
V3	3.5	25.0	71.5
V4	3.5	37.0	59.5
P1	1.2	31.3	67.5
P2	3.9	24.8	71.3
P3	3.0	35.7	61.3
P4	1.0	24.1	74.9

Table 1 presents the composition of the calibration, validation, and prediction mixtures. These mixtures were prepared by weighing the desired amount of each component with a precision of 0.1 mg and homogenizing in a mortar for at least 10 min. The mixtures were kept in sealed vials.

2.2. Iodimetric titration

A 1.50×10^{-2} mol L^{-1} iodide solution was prepared by dissolution of I_2 and potassium iodide in water and standardized with sodium thiosulfate. The L-ascorbic acid titrations were performed using 0.1000 g of each prediction mixture dissolved in 100.0 mL of de-ionized water in the presence of the starch indicator [20].

2.3. Thermal measurements

The TG curves were recorded using a TGA-2950 thermogravimetric module coupled to a TA-2000 thermal analyzer, both from TA-Instruments. Samples of about 10 mg (±1 μg) of each mixture were placed in an open alumina crucible and heated at 10.0 °C min^{-1} under a dynamic oxygen atmosphere (gas flow 100 ml min^{-1}) at atmospheric pressure.

2.4. Data pre-processing

The TG curves were smoothed by taking the average value of the mass measurements within each 1 °C temperature interval from 32 °C to 590 °C. Measurements were normalized by the initial mass value of each curve. The resulting dimensions of the TG data matrices (**X**) were 5 × 558 (calibration), 4 × 558 (validation), and 5 × 558 (prediction).

Each column of the **X** (TG data) and **Y** (concentration data) matrices was centered on the mean of the calibration set prior to the modelling procedures.

2.5. Modelling procedures for AA determination

The number of selected temperatures for MLR modelling was allowed to vary between one and four. This upper limit is given by the number of calibration mixtures (five) minus one (because one degree of freedom is lost due to the mean-centering operation) [17].

Both SPA and GA were implemented to minimize the root-mean-square error obtained by applying the resulting MLR model to the validation set (RMSEV) [17]. The optimization process in SPA is restricted to subsets of variables with small collinearity, as described elsewhere [13]. The GA employed standard binary chromosomes with length equal to the number of temperatures in the TG curve (a "1" gene indicates a selected temperature) [18,19]. One-point crossover and mutation operators were employed with