Journal of Pharmaceutical and Biomedical Analysis 120 (2016) 342–351

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

An iterative approach for compound detection in an unknown pharmaceutical drug product: Application on Raman microscopy

Mathieu Boiret[a,*], Nathalie Gorretta[b], Yves-Michel Ginot[a], Jean-Michel Roger[b]

[a] Technologie Servier, Orléans, France
[b] IRSTEA, UMR ITAP 361, Montpellier, France

ARTICLE INFO

Article history:
Received 3 October 2015
Received in revised form
17 December 2015
Accepted 20 December 2015
Available online 28 December 2015

Keywords:
Raman microscopy
Orthogonal projection
Spectral angle mapper
Identification
Low dose compound

ABSTRACT

Raman chemical imaging provides both spectral and spatial information on a pharmaceutical drug product. Even if the main objective of chemical imaging is to obtain distribution maps of each formulation compound, identification of pure signals in a mixture dataset remains of huge interest. In this work, an iterative approach is proposed to identify the compounds in a pharmaceutical drug product, assuming that the chemical composition of the product is not known by the analyst and that a low dose compound can be present in the studied medicine. The proposed approach uses a spectral library, spectral distances and orthogonal projections to iteratively detect pure compounds of a tablet. Since the proposed method is not based on variance decomposition, it should be well adapted for a drug product which contains a low dose product, interpreted as a compound located in few pixels and with low spectral contributions. The method is tested on a tablet specifically manufactured for this study with one active pharmaceutical ingredient and five excipients. A spectral library, constituted of 24 pure pharmaceutical compounds, is used as a reference spectral database. Pure spectra of active and excipients, including a modification of the crystalline form and a low dose compound, are iteratively detected. Once the pure spectra are identified, multivariate curve resolution-alternating least squares process is performed on the data to provide distribution maps of each compound in the studied sample. Distributions of the two crystalline forms of active and the five excipients were in accordance with the theoretical formulation.

1. Introduction

Raman spectroscopy is becoming increasingly more accepted as a powerful tool in the pharmaceutical research and development environment since this technique has some major benefits [1,2]. By coupling a microscope with the usual Raman spectroscopy, hyperspectral images providing both spectral and spatial information can be acquired, containing a lot of information on the distribution of active pharmaceutical ingredients (API) or excipients in a product [3]. The development of these analytical methods is very useful to ensure and control the drug product quality during development and beyond post-marketing authorisation [4]. Even if Raman chemical imaging has been used to detect and quantify crystalline forms [5–7], to characterize particle size [8] or to assess blending effect on tablet quality [9], the main goal in the pharmaceutical industry remains the assessment of the product quality by determining the compound distributions within a tablet [10–12].

Because of the huge amount of data contained in hyperspectral images, a direct interpretation of the acquired images is not possible and several chemometric tools have previously been published to aid in this task. Hyperspectral data analysis can be divided in several parts [13] depending on the objectives, but most of the times it starts with a pre-processing step followed by a data analysis procedure. Pre-processing methods are usually applied to correct for external perturbations and undesired phenomena to focus on the targeted information. The next step consists of analysing the data by applying qualitative or quantitative chemometric tools such as principal component analysis (PCA) [14,15], independent component analysis (ICA) [16] or multivariate curve resolution-alternating least squares (MCR-ALS) [17]. These algorithms assume that the acquired spectra are the weighted sum of pure spectra of the formulation compounds. One challenging task during application of these chemometric tools on imaging techniques is how to effectively extract chemical information from the image [18] but, the quality of the extracted signals (related to each pure compound) is also a critical step of the multivariate data analysis.

A lot of algorithms have been previously studied to extract pure spectra (also named endmembers in the remote sensing

* Corresponding author.
E-mail address: mathieu.boiret@servier.com (M. Boiret).

http://dx.doi.org/10.1016/j.jpba.2015.12.038
0731-7085/

field) within a mixture dataset [19,20]. On the one hand, in the chemometrics community, SIMPLISMA (simple-to-use interactive self-modeling mixture analysis) [21], orthogonal projection approach (OPA) [22], independent component analysis (ICA) [23] or evolving factor analysis (EFA) [24] were used on spectroscopic data to identify these pure signals in a mixture dataset. On the other hand, in the remote sensing community, other approaches such as pixel purity index (PPI) [25], autonomous morphological endmember extraction (AMEE) [26], N-FINDR [27] or vertex component analysis (VCA) [28] appeared as powerful algorithms to extract pure features from hyperspectral images. However, all these algorithms are mainly based on either the hypothesis that each compound has a pure pixel in the image or that a signal contains a sufficient level of spectral contributions for a compound. Considering a low dose compound, it can be assumed that spatial and spectral information is scarce because only few pixels of the image contain the product of interest and because the associated spectral contributions are mixed with the other formulation compounds. Considering this specific case, there are no pure pixels in the studied dataset and identification of the low dose compound appeared as a real challenge [29,30]. In most chemometric methods, the targeted information is extracted by using the variability between the samples i.e. the differences between the acquired spectra or pixels. But, in the case of the low dose compound, these variations cannot be easily highlighted as the associated contributions are weak and spread into mixture spectra or noise contribution.

In most pharmaceutical applications, the studied formulations are known beforehand, but in some cases, analysts have limited information or do not have prior knowledge on the studied product. For instance, in forensic applications, illegal medicines can be analysed by vibrational spectroscopy to quickly detect counterfeit products [12,31,32]. Comparing with genuine drugs, counterfeit samples can be manufactured with different actives or excipients, and identification of product compounds without prior knowledge on the samples could be of interest for analysts. Moreover, during development of a pharmaceutical drug product, stability studies are performed to analyse the evolution of the product in time through different storage conditions (packaging, temperature and relative humidity). Because Raman chemical imaging combined with chemometric algorithms is useful to explore the inner structure of a pharmaceutical drug product [33], evolution of the active quality can be observed in terms of degradations or modifications of its crystalline forms [34]. Therefore, this analytical tool appeared as a very promising methodology to monitor these modifications.

In this article, the objective is the identification of pure compounds in a pharmaceutical tablet, assuming that analysts do not know the studied formulation beforehand and that a potential low dose product is present in the sample. A new methodology is proposed to provide the chemical composition of the tablet. By using a spectral library, compounds are iteratively detected by calculating spectral distances between images and reference spectra. Because each compound has its proper subspace containing chemical information and spectral variability, the associated spectral contributions can be iteratively removed by using orthogonal projections. This approach works exclusively in the signal space, describing the P-dimensional space (one axis per variable) in which the observations can be represented as vectors. Thus, it ensures the detection of a compound without requiring important variations between samples (or pixels). Therefore, by progressively identifying the formulation compounds, from the main product to a product with low contributions, this approach is particularly well adapted to detect all the compounds in a formulation. After spectral identification and in order to provide distribution maps of actives and excipients, MCR-ALS process is applied [35,36].

The remainder of the paper is organized as follows. Section 2 describes the experimental framework, including notations, samples and apparatus details. The proposed iterative approach will be described in this section. Section 3 presents the ability of the proposed approach to detect the pure compound in an unknown formulation and the MCR-ALS results. Finally, Section 4 presents our conclusions.

2. Materials and methods

2.1. Notations

Vectors are noted in bold lowercase, matrices in bold uppercase, and scalars in italic lowercase characters. Vectors are arranged in lines and one line represents one spectrum. The transposed forms of a vector $\mathbf{x}$ and a matrix $\mathbf{X}$ are noted $\mathbf{x}^T$ and $\mathbf{X}^T$, respectively. $\mathbf{I}$ is the identity matrix of dimensions $p \times p$, where p is the number of variables in a spectrum. $\mathbf{x}$ and $\mathbf{X}$ orthogonally projected to a vector basis $\mathbf{K}$ are noted $\boldsymbol{x}_{\perp}$ and $\boldsymbol{X}_{\perp}$. $\boldsymbol{\Sigma}$ is the Euclidian orthogonal projector to $\mathbf{K}$.

For a spectral matrix $\boldsymbol{X} \in \mathbb{R}$, the sample space $\mathbb{R}^n$ describes the N-dimensional space (one axis per observation) in which we can represent the variables (Raman shift) as vectors. The spectral space $\mathbb{R}^p$ describes the P-dimensional space (one axis per variable) in which we can represent the observations (sample spectra) as vectors.

2.2. Samples

A pharmaceutical tablet was especially manufactured by wet granulation for this study. The tablet was prepared by mixing and granulating one active pharmaceutical ingredient, Ivabradine (chronic heart failure treatment), commercialised by "Les Laboratoires Servier", and four excipients: metolose® (Shin Etsu, Tokyo, Japan), eudragit® (Evonik, Essen, Germany), microcrystalline cellulose, and maltodextrin. Dried and calibrated granulates were lubricated with magnesium stearate, which can be associated with a low dose compound as it represented only 0.5% (w/w) of the theoretical formulation. The lubricated granulates were compressed with a rotary press equipped with punches and dies allowing the production of tablets with the required shape. Film-coating and smoothing are carried out in rotative coating pans. The studied drug product contained 10% (w/w) of active in the tablet. The active is known to have several solid state forms (form 1 and form 2) but only the original active form 1 was used to manufacture the product. Submitted to high temperature conditions, the active is known to undergo a crystalline modification from form 1 to form 2. Before Raman chemical imaging analysis, the tablet was stored 3 months at 50 °C in a blister. In order to analyse the tablet core and to ensure a flat surface, the tablet was eroded with a Leica EM Rapid system (Leica, Wetzlar, Germany). A visual examination of the tablet did not provide any information concerning the distribution of the different compounds within the tablet.

2.3. Raman imaging system

The tablet image was collected using a RM300 PerkinElmer system (PerkinElmer, Waltham, MA) and the Spectrum Image version 6.1 software. A microscope equipped with an objective 100× magnification was coupled to the spectrometer and spectra were acquired through it with a spatial resolution of 10 μm in a Raman diffuse reflection mode. Wave number range was 3200–100 cm^{-1} with a resolution of 2 cm^{-1}. Spectra were acquired at a single point on the sample, then the sample was moved and another spectrum was taken. This process was repeated until spectra of points covering the region of interest were obtained. A 785 nm laser with a power of 400 mW was used. Four scans of three seconds were accumulated for each spectrum. An image of 30 pixels per 30 pixels