Journal of Pharmaceutical and Biomedical Analysis 128 (2016) 236–246

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Quantification and handling of nonlinearity in Raman micro-spectrometry of pharmaceuticals

Brigitta Nagy, Attila Farkas, Attila Balogh, Hajnalka Pataki, Balázs Vajna, Zsombor K. Nagy [1], György Marosi *

Department of Organic Chemistry and Technology, Budapest University of Technology and Economics, H-1111 Budapest, Budafoki út 8, Hungary

ARTICLE INFO

Article history:
Received 2 March 2016
Received in revised form 12 May 2016
Accepted 20 May 2016
Available online 21 May 2016

Keywords:
Hyperspectral imaging
Pharmaceutical analysis
Chemometrics
Multivariate regression
Variable selection
Nonlinear behavior

ABSTRACT

This work demonstrates how nonlinearity in Raman spectrometry of pharmaceuticals can be handled and accurate quantification can be achieved by applying certain chemometric methods including variable selection. Such approach proved to be successful even if the component spectra are very similar or spectral intensities of the constituents are strongly different. The relevant examples are: blends of two crystalline forms of carvedilol ("CRYST-PM" blend) and a three-component pharmaceutical model system ("PHARM-TM" blend). The widely used classical least squares regression (CLS) and partial least squares regression (PLS) quantification methods provided relatively poor root mean squared error of prediction (RMSEP) values: approximately 2–4% and 4–10% for CRYST-PM and PHARM-TM respectively. The residual plots of these models indicated the nonlinearity of the preprocessed data sets. More accurate quantitative results could be achieved with properly applied variable selection methods. It was observed that variable selection methods discarded the most intensive bands while less intensive ones were retained as the most informative spectral ranges. As a result not only the accuracy of concentration determination was enhanced, but the linearity of models was improved as well. This indicated that nonlinearity occurred especially at the intensive spectral bands. Other methods developed for handling nonlinearity were also capable of adapting to the spectral nature of both data sets. The RMSEP could be decreased this way to 1% in CRYST-PM and 3–6% in PHARM-TM. Raman maps with accurate real concentrations could be prepared this way. All quantitative models were compared by the non-parametric sum of ranking differences (SRD) method, which also proved that models based on variable selection or nonlinear methods provide better quantification.

1. Introduction

Chemical imaging, including Raman mapping, has been a particularly rapidly emerging technique in analysis and characterization of various substances in the last decade. This method has several advantages over non-imaging spectroscopic techniques. Besides qualitative and quantitative characterization of ingredients in bulk materials, it can provide further information about spatial distribution of major and even minor (sometimes trace) components [1]. Many diverse issues have been already solved using chemical imaging in the field of life sciences and diagnostics [2,3], forensic sciences and counterfeiting [4,5], food analysis [6,7], plastics [8] and artworks [9]. In recent years, the application of this approach has sparked explosively growing interest particularly in the pharmaceutical industry [10,11]. Raman (or NIR) chemical imaging is greatly applicable for performing detailed analysis in various steps of manufacturing processes [12,13]. Researchers pay more and more attention to quantification, further highlighting the relevance of this topic.

There are many issues in which Raman spectroscopy and hyperspectral imaging has helped tackle serious challenges, such as identifying unexpected chemical substances or new polymorphs [14,15], investigating blend homogeneity [16], or testing polymorphic stability [17,18]. These qualitative studies reveal various types of information about the samples, allowing better understanding of pharmaceutical processes. However, the interest, regarding pharmaceutics, is mostly focusing on the complex view of the structure and the real quantification of component. Since FDA approved the guidance on Process Analytical Technology (PAT) [19], the significance of spectroscopic techniques has extremely grown thanks

* Corresponding author.
E-mail addresses: zsknagy@oct.bme.hu (Z.K. Nagy), gmarosi@mail.bme.hu (G. Marosi).
[1] Additional author.

http://dx.doi.org/10.1016/j.jpba.2016.05.036
0731-7085/

to their superior adaptability into continuous manufacturing processes [20–23]. Ongoing quality control can be achieved through accurate spectral evaluation, to which, however, the use of chemometrics is essential.

Since a vibrational (such as Raman, IR or NIR) spectrum contains hundreds or thousands of wavenumbers of interest, it is a multivariate entity, which progresses to a further level of complexity when combined with hyperspectral imaging (which requires processing of even thousands of spectra within the same dataset). Numerous chemometric methods may serve the quantification efforts [10,24–26]. In the simplest cases univariate approaches using one selective wavenumber can provide useful results. However, in many cases a sufficiently selective band does not exist [10], in which case multivariate methods has to be used encompassing the whole spectral range or parts of it [27]. One of the most easily interpretable methods is the classical least squares (CLS). It can be adapted fast for simple spectrum characterization when the spectra of all components are known and spectrum of each compound can be generated from pure spectra using spectral contributions (as estimated concentrations). However, some interfering effects, such as spectral similarity of components or material interactions or nonlinear behavior, may occur making it necessary to use more sophisticated approaches. Some authors have reported successful application of widespread chemometric methods such as partial least squares regression (PLS) or principal component regression (PCR) [28–30]. These methods were successfully used in polymorphic studies, where the component spectra are only slightly different. PLS is especially preferred for quantification of selected polymorphs in a mixture.

A large part of the spectral range is often non-informative in the quantitative evaluation. In such cases variable selection methods, such as interval partial least squares (iPLS) [31,32] or genetic algorithms (GA) [33–35], are promising candidates for treating the spectra and retaining only the sufficiently descriptive variables. In some cases, a certain degree of nonlinearity appears in the data, caused by interaction between components or due to spectral preprocessing. This phenomenon can be handled by polynomial partial least squares (pPLS) [36,37], locally weighted regression (LWR) [38,39] or support vector machine for regression (SVR) [40,41]. pPLS works similarly to the conventional PLS regression, but it uses higher degree inner relations by determining polynomial functions between score values of dependent and independent variables called X-block scores and Y-block scores. LWR was applied based on PLS projection by fitting local PLS models to a specified range of adjacent observations. Applying SVR, the regression is carried out in a higher-dimensional feature space, in which the nonlinearity of the original input data can be handled properly. The construction of the suitable hyperplane is performed by a kernel function and regularized by several parameters (see Section 2.4.3)

Although the conventional data analysis methods (CLS, PCR, PLS) proved to be successful in many applications, advances in chemical imaging and in pharmaceutical process-monitoring calls for novel chemometric methods [25,42]. Nevertheless up to now only few pharmaceutical related studies have demonstrated the advantageous use of aforementioned chemometric ways for variable selection and handling of nonlinearity. Support vector machine, for instance, has been proposed as a promising candidate [42] and used in Raman [43] and UV [44] quantitative spectroscopic studies, iPLS was applied in determination of Vitamin B12 in pharmaceutical tablets [45] and in the quantification of ibuprofen-nicotinamide co-crystals [46], while LWR has been used in a NIR quantitative analysis [47]. However, detailed comparative study demonstrating the relevance of the mentioned methods for quantitative determination of polymorph ratio and tablet composition based on Raman mapping has not published yet. Thus, the aim of this study is to evaluate the applicability of these tools in analysis of two model systems of pharmaceutical importance.

2. Materials and methods

2.1. Materials

Two-component mixtures of crystalline polymorphs of carvedilol model drug (referred to as CRYST-PM) were studied. The commercial carvedilol product (EGIS Pharmaceuticals Plc., Budapest, Hungary) consisted of pure Form II polymorph. Form I was obtained by a solvent mediated polymorphic transition process. First, 25 g Form II was dissolved entirely in 120 mL ethyl-acetate (Merck, Germany) heating the solution until 77 °C. The solution then was cooled while 2.5 g of Form I polymorph was added as seed crystals. The recrystallization occurred at 50 °C in three hours [48]. Crystals were removed by filtration. After drying the product purity was verified by Raman mapping.

The three component pharmaceutical model system (PHARM-TM) contained imipramine (EGIS Pharmaceuticals Plc., Budapest, Hungary) as model drug and microcrystalline cellulose (FMC BioPolymer, Princeton, USA) and maize starch (Colorcon, West Point, USA) as excipients. Each component was sieved to ensure the same particle size range (50–100 μm), to avoid segregation.

2.2. Preparation of mixtures and tablets

Uniform binary mixtures were achieved by grinding and mixing carvedilol Form I and Form II in a mortar with pestle, creating nineteen blends with different mass ratios (see Table 1). The total weight of each mixture was 2.00 g and the measurements of the components were carried out on analytical balance (precision of 0.1 mg). As the precision of the weighted quantity of the components were within 5 mg, there was no significant difference between the prepared actual and nominal concentrations. The mixtures were prepared right before the Raman measurements. Sufficient homogeneity was obtained by ten minutes of thorough homogenization, which was checked by collecting three Raman maps per mixtures (See Section 2.3). As there were no differences in the spatial distri-

Table 1
CRYST-PM mass ratios and composition of calibration and validation sets.

Sample	Form I	Form II	type of set	Sample	Form I	Form II	type ofset
CAR_1	0%	100%	cal.	CAR_11	70%	30%	cal.
CAR_2	1%	99%	cal.	CAR_12	85%	15%	cal.
CAR_3	2%	98%	val.	CAR_13	89%	11%	val.
CAR_4	3%	97%	cal.	CAR_14	92%	8%	cal.
CAR_5	5%	95%	val.	CAR_15	95%	5%	val.
CAR_6	8%	92%	cal.	CAR_16	97%	3%	cal.
CAR_7	11%	89%	val.	CAR_17	98%	2%	val.
CAR_8	15%	85%	cal.	CAR_18	99%	1%	cal.
CAR_9	30%	70%	cal.	CAR_19	100%	0%	cal.
CAR_10	50%	50%	val.				