Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 43 (2007) 407–412
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Quantitative analysis of paracetamol polymorphs in powder mixtures by FT-Raman spectroscopy and PLS regression

Kyriakos Kachrimanis [a,*], Doris E. Braun [b], Ulrich J. Griesser [b]

[a] *Department of Pharmaceutical Technology, School of Pharmacy, University of Thessaloniki, 54124 Thessaloniki, Greece*
[b] *Institute of Pharmacy, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria*

Received 3 March 2006; received in revised form 15 July 2006; accepted 18 July 2006
Available online 28 August 2006

Abstract

A fast and simple method for the quantitative analysis of monoclinic (form I) and orthorhombic (form II) paracetamol was developed, based on FT-Raman spectroscopy and PLS regression. Three different preprocessing algorithms, namely orthogonal signal correction (OSC), standard normal variate transformation (SNV) and multiplicative scatter correction (MSC), were applied in order to eliminate effects caused by sample preparation and sample inhomogeneities. Subsequently, PLS regression models were fitted and their predictive performance was evaluated on the basis of the root mean squared error of cross-validation (RMSECV) over the complete data set. Furthermore, the data were split into two equal-sized training and test subsets by the Kennard-Stone design and the errors of calibration (RMSEC) and prediction (RMSEP) were calculated. It was found that the OSC preprocessing contributes to a significant increase in the predictive performance of the PLS regression model (RMSECV = 0.500%, RMSEC = 0.842% and RMSEP = 0.538%) in the overall concentration range of form I, compared to the SNV (RMSECV = 2.398%, RMSEC = 0.911% and RMSEP = 7.177%) and MSC (RMSECV = 2.7648%, RMSEC = 1.572% and RMSEP = 4.838%). In addition, the model developed on OSC preprocessed data is more parsimonious, requiring a single latent variable, compared to three latent variables required by the models fitted to the SNV and MSC preprocessed data. The proposed multivariate calibration presents a significant improvement over existing methods for the quantitation of paracetamol polymorphs.

Keywords: FT-Raman; Paracetamol; Crystal polymorphism; Multivariate calibration; PLS regression; Orthogonal signal correction; Standard normal variate transformation; Multiplicative scatter correction

1. Introduction

The metastable orthorhombic polymorph of paracetamol, form II, [1] is of particular interest in tablet manufacturing, because it has well-defined slip planes in its crystal lattice and is suitable for direct compression [2]. Reproducible crystallization of large amounts of form II from ethanol solutions has been achieved by the optimization and scaling-up [3] of a seeding technique proposed by Nichols and Frampton [4]. However, solution-grown form II is usually contaminated with crystals of the monoclinic polymorph (form I) depending on harvesting time and drying conditions [5]. Therefore, a fast and simple method for the quantitative analysis of forms I and II in crystalline powders is important for the evaluation of polymorphic purity, as well as for the prediction of polymorphic stability upon prolonged storage. Powder X-ray diffraction (PXRD) remains the most powerful technique for the characterization of crystal structures, however its use in quantitative determination is limited by the dependence of peak intensities on preferred orientation effects, while particle size and strain caused by crystal defects (e.g. dislocations) can be a source of peak broadening. Vibrational spectroscopic (FTIR and FT-Raman) techniques represent an attractive alternative, as they are faster and the required instrumentation is in general less costly and more widely available. FT-Raman spectroscopy in particular, is well-suited for the identification and quantitative analysis of crystal polymorphs [6], since it requires very little sample preparation, thus minimizing the risk of solid-state transformations of metastable polymorphs, while the particle size and shape of the sample have little effect [7]. The potential of Raman techniques in the quantitative analysis of polymorphs has been demonstrated in the case of powder mixtures [7–10], as well as drug products [6,11].

* Corresponding author. Tel.: +30 2310 997666; fax: +30 2310 997652.
E-mail address: kgk@pharm.auth.gr (K. Kachrimanis).

0731-7085/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2006.07.032

A spectroscopic FTIR as well as an FT-Raman method for the quantitative analysis of paracetamol polymorphs I and II in powder mixtures has been developed by Al Zoubi et al. [8] by the use of univariate classical least squares regression. The spectroscopic methods performed equally well with PXRD, with a reported limit of detection (LOD) of 1.2% (w/w) for form I. A similar result was reported by Ivanova [12] for a linear dichroic FTIR method. However, the aforementioned FTIR methods consider form I contents higher than 10% (w/w), and an additional drawback is that they require KBr pellet preparation, which involves a risk of polymorphic transition during grinding. The FT-Raman method of Al Zoubi et al. [8] utilizes the 454–465 cm^{-1} pair of bands, which are partially overlapping, and it requires peak deconvolution [8], which is a rather elaborate procedure, for routine analysis tasks. Furthermore, univariate methods lack the advantages of modern multivariate calibration algorithms, which have been proven highly efficient in the quantitation of polymorphs in powder mixtures [10,13]. Therefore, in the present study, it was considered important to develop a simple FT-Raman method for the quantitative analysis of binary mixtures of paracetamol crystal forms I and II, taking full advantage of the high efficiency of modern multivariate calibration methods. In particular, PLS regression is applied utilizing the complete spectral range, and the predictive performance of the developed calibration models is assessed over the complete data set on the basis of the root mean squared error of cross-validation (RMSECV), as well as of calibration (RMSEC) and prediction (RMSEP), after splitting the data set into two equal-sized training and a test subsets, respectively, applying the Kennard-Stone design [14]. The orthogonal signal correction (OSC) [15], the standard normal variate transformation (SNV) [16] and the multiplicative scatter correction (MSC) [17] preprocessing algorithms are applied in order to eliminate non-additive effects caused by sample preparation and sample inhomogeneities, for example. Their efficiency in improving the performance of the PLS models is compared and discussed.

2. Materials and methods

2.1. Materials

Commercial paracetamol (form I) was obtained from Apoka (Apoka Pharma Produktions und Handelgesellschaft m.b.H., Austria) and was used as received. Pure orthorhombic paracetamol (form II) was prepared by melt crystallization. The crystal form and purity of both polymorphs (I and II) used as starting materials was verified by powder X-ray diffraction.

2.2. Methods

2.2.1. Powder X-ray diffraction

The powder X-ray diffraction experiments were conducted using a Siemens D-5000 diffractometer (Siemens AG, Karlsruhe, Germany), equipped with a theta/theta goniometer, a Cu Kα radiation source, a Goebel mirror (Bruker AXS, Karlsruhe, Germany), a 0.15° soller slit collimator and a scintillation counter. Powder samples were scanned in the angular range of 2–40°, at a scan rate of 0.005° 2θ/s at a tube voltage of 40 kV and a tube current of 35 mA.

2.2.2. FT-Raman spectroscopy

FT-Raman spectra were recorded on a Bruker RFS 100 FT-Raman spectrometer, equipped with a diode pumped Nd:YAG laser (1064 nm) as the excitation source, and a liquid nitrogen cooled, high sensitivity Ge detector (Bruker Optik GmbH, Ettlingen, Germany). A few milligrams of the sample were placed into a small aluminum sample cup and lightly packed. For each spectrum 64 scans were performed at a resolution of 4 cm^{-1} over the range 0–4000 cm^{-1}. A Blackman-Harris B4 term was used as apodization function. The spectral data from each sample were exported in electronic format and the Know-It-All Informatics System v.5.0 (Bio-Rad Laboratories, Inc.) was used for analysis, and finding peak attributions.

2.2.3. Preparation of mixtures

Eighteen binary mixtures were prepared by geometrically mixing pure polymorphs II and I. The mixing procedure emphasized in the lower concentrations of form I, although it covers sufficiently the entire concentration range. Finally, 20 samples (including the starting materials) were used. The concentration of form I in the samples was: 100, 97.75, 95.5, 91, 82, 73, 64, 48, 32, 24, 16, 12, 8, 6, 4, 3, 2, 1.5, 1 and 0% (w/w).

2.2.4. Multivariate calibration

The complete spectrum range of 0–4000 cm^{-1} was used after application of OSC, SNV and MSC preprocessing algorithms, in order to eliminate sources of non-linearity or remove features uncorrelated with the concentration of the analyte. In particular, the OSC algorithm [15] calculates parts of the spectrum that are orthogonal (uncorrelated) to the concentration of the analyte and removes them. The OSC preprocessed spectra are subsequently subjected to mean-centering (subtraction of the average from each spectrum). The SNV transformation [16] centers each spectrum separately by subtracting its mean and then scales it by its own standard deviation. Finally, MSC [17] eliminates light scattering or change in path length effects for each sample relative to the average of the calibration set by shifting and rotating each spectrum so that it fits closely to the average spectrum of the dataset. The algorithm operates on a segment representative of the baseline of the spectra and the fitting to the average spectrum is performed by least squares.

Cross-validation by the leave-one-out method was initially applied in order to evaluate the models, and subsequently, the data were split into homogeneous training and test subsets, each consisting of 10 samples, applying the Kennard-Stone design [14]. This algorithm selects data points sequentially, starting from a point closest to the average spectrum of the data set and adding subsequent points on the basis of the maximum squared distance to all of the already selected points, guaranteeing that the selected data points are uniformly distributed within the original data set. The training set selected this way, contained the 100, 82, 73, 64, 48, 24, 8, 4, 3 and 0% (w/w) mixtures, while