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

the rest of the mixtures (97.75, 95.5, 91, 32, 16, 12, 6, 2, 1.5 and 1%, w/w) were assigned to the test set.

All calculations were performed on a PC running MS Windows XP, equipped with an Intel Pentium 3 GHz processor and 768 MB RAM. The Simca-P v.9 (Umetrics AB) was used for the OSC, SNV and MSC transformations, as well as fitting of the PLS models, while the Kennard-Stone routine of the ChemoAC toolbox for the Matlab (Matlab 6.5, Mathworks Inc.) was used for the separation of the data into uniform training and test subsets. The predictive performance was assessed on the basis of the root mean squared error of cross-validation, of calibration and of prediction, calculated by:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n}} \quad (1)$$

3. Results and discussion

In Fig. 1 PXRD patterns of the starting materials (form I and II) used in this study are shown, and in Table 1 the *h k l* planes, peak positions and relative intensities of paracetamol forms I and II reported in the literature [4], together with corresponding values for the starting materials used in this study are listed. It is seen that there is good agreement with the previously published experimental data, confirming the polymorphic purity of the samples used to prepare the mixtures. Moreover, in the PXRD pattern of form II, Fig. 1, it is seen that the intensity of the 24.03° 2θ reflection is unusually high, a strong indication of preferred orientation of the crystalline particles. This is also evident from the relative intensities listed in Table 1, where it is seen that all relative intensities are much lower in the melt-crystallized sample compared to those of the solution-crystallized orthorhombic form reported in the literature [4]. This reflection corresponds to the 0 0 2 Miller plane, which is a well-defined slip plane, existing in the structure of form II. This slip plane is responsible for the reduced elasticity of this polymorph and it is possible that grinding of the melt-grown crystalline material during sample preparation, results in extensive fracture along the 0 0 2 plane, and in the subsequent increase of the corresponding reflection intensity in the diffractograms. Another interesting difference can also be seen between the two monoclinic forms, in the relative intensities of the 0 2 2 and the 1 1 1 *h k l* plane reflections. The 0 2 2 plane shows the highest intensity reflection in the data reported by Nichols and Frampton [4], while the 1 1 1 plane shows the highest intensity reflection in the commercial product used in this study. This observation is also consistent with strong preferred orientation effects, and is one of the reasons why spectroscopic methods for the quantitative analysis of paracetamol polymorphs are advantageous over PXRD based methods, since they are not sensitive to particle orientation.

In Fig. 2, FT-Raman spectra of pure polymorphs I and II are presented together with the difference spectrum (form II–form I), and in Fig. 3 representative FT-Raman spectra of pure form II are shown, after preprocessing by the OSC, SNV and MSC algorithms. From Fig. 2 it is seen that the most distinct difference lies in the lattice vibration region, where the orthorhombic polymorph has a very intense band at 122 cm^{-1}, which is absent from the spectrum of form I. Other differences can be found in the region between 1540 and 1680 cm^{-1}, where the observed Raman bands are attributed to the amide carbonyl group vibrations and the aromatic hydrogens, and the region around 1200 cm^{-1}, where the stretching vibrations of the aromatic C–O and C–N

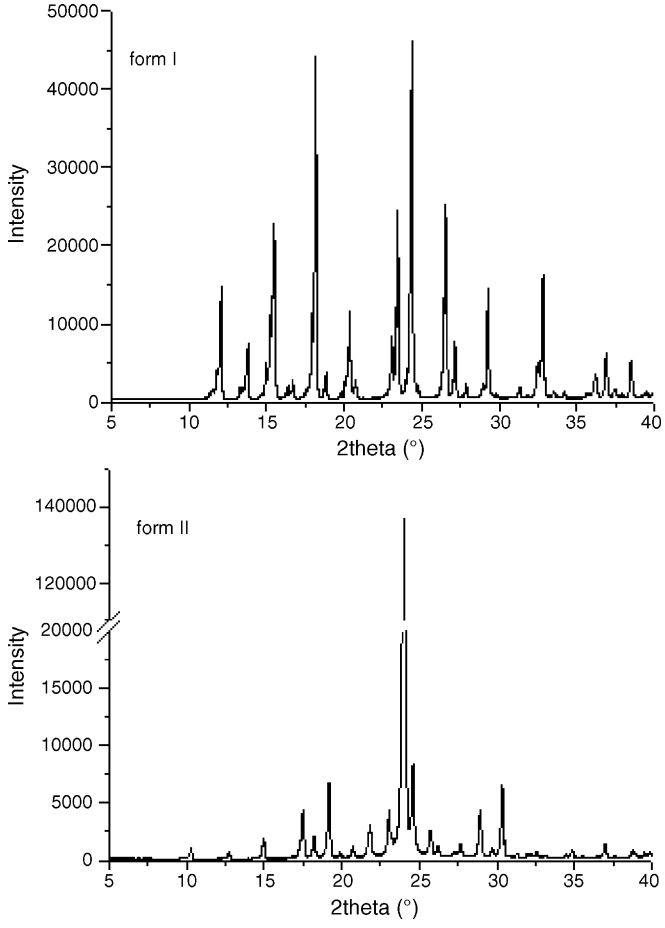

Fig. 1. Powder X-ray diffractograms of paracetamol forms I and II used in this study.

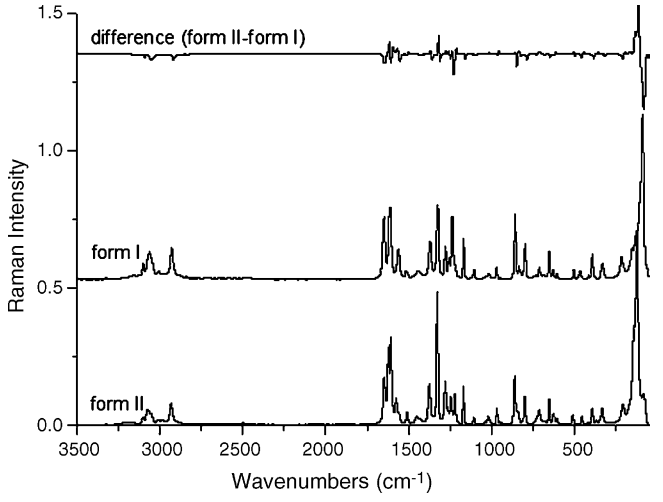

Fig. 2. FT-Raman spectra of paracetamol forms I and II used in this study, together with the corresponding difference spectrum (form II–form I).

Table 1
PXRD *h k l* planes, corresponding peak positions and relative intensities for the monoclinic (form I) and orthorhombic (form II) polymorphs of paracetamol reported in the literature [4] and the starting materials used in this study (commercial, form I and melt-grown, form II)

Monoclinic form (I)					Orthorhombic form (II)				
Nichols and Frampton [4]			Commercial		Nichols and Frampton [4]			Melt-grown	
h k l	2θ (°)	I/I_{max} (%)	2θ (°)	I/I_{max} (%)	*h k l*	2θ (°)	I/I_{max} (%)	2θ (°)	I/I_{max} (%)
0 1 1	12.11	26	12.08	26	200	10.32	4	10.29	1
1 0 −1	13.83	18	13.78	12	210	12.76	5	12.73	1
0 0 2	15.24	3	14.99	9	020	14.99	14	14.96	1
1 0 1	15.51	72	15.51	47	211	17.51	26	17.49	4
1 1 0	15.70	4	16.40	3	220	18.23	22	18.21	1
1 1 −1	16.73	11	16.71	5	021	19.20	49	19.19	6
1 1 1	18.18	68	18.17	100	121	19.88	3	–	–
0 2 0	18.91	13	18.84	6	400	20.72	9	20.70	1
0 2 1	20.38	39	20.35	21	221	21.84	22	21.82	2
1 1 −2	20.76	7	20.73	5	002	24.04	100	24.03	100
1 1 2	23.09	9	23.09	16	102	24.60	14	24.58	8
1 2−1	23.48	62	23.46	47	230	24.86	9	–	–
0 2 2	24.37	100	24.35	93	420/112	25.70	13	25.71	2
1 0 −3	24.74	5	–	–	131/202	26.18	5	26.18	1
1 2 −2	26.55	62	26.53	51	–	–	–	27.29	1
2 1 −1	27.17	11	27.16	14	231	27.66	12	27.65	1
2 0 −2	27.85	4	27.86	3	022	28.41	3	–	–
1 2 2	28.40	2	–	–	122	28.93	21	28.90	4
2 1 1	29.01	7	29.00	4	312	29.70	5	29.70	1
1 1 3	29.27	6	29.28	27	222	30.33	27	30.31	4
0 2 3	29.89	3	–	–	600	31.28	5	31.31	1
1 2 −3	31.27	5	31.31	3	240	32.03	7	32.08	1
2 2 −1	31.88	3	–	–	322	32.57	6	32.55	1
0 3 2/1 3 1	32.58	17	32.48	10	141/431	33.08	4	–	–
–	–	–	32.789	32	132	33.64	2	–	–
1 3 −2	34.17	2	34.19	2	611	34.46	5	–	–
1 3 2	35.71	3	–	–	232	34.84	7	34.84	1
1 1 4	36.19	10	36.19	7	422	35.45	3	–	–
0 3 3	36.90	18	36.90	11	512	36.49	3	–	–
2 0 −4/2 2 −3	37.47	7	37.45	2	440/621	36.96	17	36.95	1
2 1 3	37.90	4	–	–	630	38.89	7	38.73	1
–	–	–	38.48	9	–	–	–	39.54	1

appear. Finally, there is a difference in the 450–470 cm^{-1} region, which has been exploited for the development of a univariate calibration method [8]. Regarding the effect of the preprocessing on the spectra, from Fig. 3 it is seen that the OSC preprocessing algorithm distorts the spectra, while the SNV and MSC algorithms produce similar results, showing less pronounced differences to the original spectrum (Fig. 2), which are mostly related to the scale and position of the spectrum. The similarity between the SNV and MSC algorithms has been pointed out by Dhanoa et al., [18] and Helland et al., [19].

In Table 2, observed (experimental) versus predicted form I concentrations of the test subset selected according to the Kennard-Stone design are listed, and in Table 3 are summarized the root mean squared error of cross-validation in the complete data set, of calibration in the training and of prediction in the test subset, for the PLS models fitted to the OSC, SNV and MSC preprocessed spectra. Additionally, plots of observed versus predicted form I concentrations (not shown) proved that the PLS model fitted by the leave-one-out cross-validation method on the complete data preprocessed by the OSC algorithm is performing better than the rest in the whole concentration range, although this model contains a single latent variable. In both the cases of SNV and MSC preprocessing, the PLS models failed, particularly in the low form I concentration range, although they contain three PLS components, in agreement with the cor-

Table 2
Predicted concentration values of form I by PLS regression for the test subset

Concentration of paracetamol form I (%, w/w)			
Observed	Predicted by PLS after preprocessing by algorithm		
	OSC (1)	SNV (3)	MSC (3)
1	0.85	0.07	5.55
1.5	1.47	3.37	2.31
2	1.85	1.59	3.35
6	5.95	6.91	4.71
12	11.74	7.60	10.47
16	15.50	12.26	10.11
32	31.82	29.20	23.68
91	90.47	90.21	87.74
95.5	95.61	79.29	86.11
97.75	96.26	83.42	95.69

In parentheses, the numbers of latent variables used in each model are given.

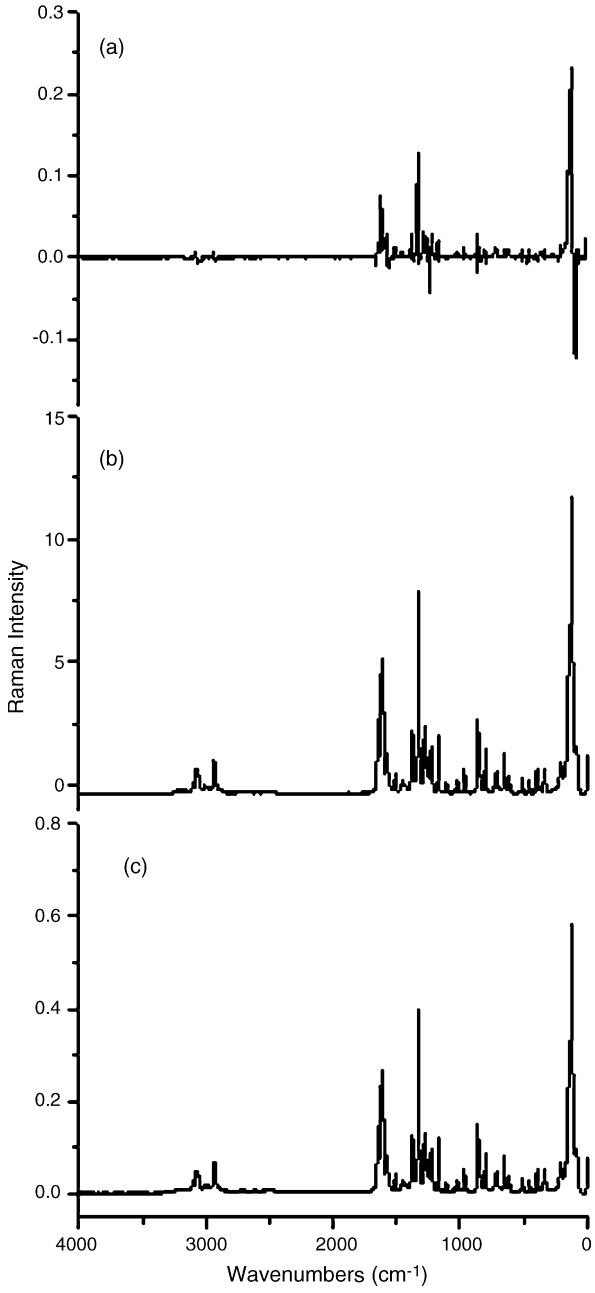

Fig. 3. Representative FT-Raman spectra of the orthorhombic form (II), after preprocessing by the: OSC (a), SNV (b) and MSC (c) algorithm.

responding results for the separate test subset (Table 2). The RMSECV, Table 3, of the model trained on OSC preprocessed data (0.500%), is significantly lower than that of the models trained on the SNV (2.394%) and MSC (2.764%) preprocessed data. Although leave-one-out cross-validation is suitable for small datasets such as the one used in the present study, it is considered optimistic and asymptotically incorrect [20], therefore, the performance of the PLS models was validated using a separate test subset consisting of equal number of data points to the training subset, in order to minimize the risk of over-fitting. From the data listed in Table 3 it is seen that, again, the PLS model fitted to the OSC preprocessed data is performing much better, showing lower RMSEC (0.842%) and RMSEP (0.538%)

Table 3
Root mean squared error of cross-validation (RMSECV) for the complete dataset, of calibration (RMSEC) for the training and prediction (RMSEP) for the test subset, together with linear regression parameters (correlation coefficient, *R*, slope and intercept) of observed versus predicted concentrations of form I in the test subset

	Preprocessing algorithm		
	OSC (1)	SNV (3)	MSC (3)
Root mean squared error			
RMSECV (%)	0.500	2.398	2.764
RMSEC (%)	0.842	0.911	1.572
RMSEP (%)	0.538	7.177	4.838
Regression coefficients			
R	0.9999	0.9940	0.9957
Slope ± S.E.	1.0053 ± 0.0033	1.1122 ± 0.0432	1.0501 ± 0.0345
Intercept ± S.E.	0.1332 ± 0.1776	0.5610 ± 2.0518	0.8511 ± 1.7298

In parentheses, the numbers of latent variables used in each model are given.

and linear regression parameters of observed versus predicted concentration of form I close to the ideal values (correlation coefficient 0.9999, slope 1.005 and intercept 0.133). Interestingly, the RMSEP is of the same magnitude with the RMSECV, and slightly smaller than the RMSEC, which is an indication that the model is not over-fitted. Finally, it is not surprising that the SNV and MSC preprocessing leads to models of comparable predictive performance and with three latent variables, since it has been demonstrated that these preprocessing methods lead to fairly equivalent results [18,19].

4. Conclusions

A fast and precise FT-Raman method for the quantitative analysis of paracetamol polymorphs I and II in powder mixtures is developed applying PLS regression to the complete spectra after appropriate preprocessing. It was found that preprocessing of the spectra by the OSC algorithm resulted in the development of simpler PLS models with excellent predictive performance, while preprocessing by the SNV and MSC algorithms resulted in PLS models with worse predictive performance despite the larger number of latent variables. Although the mixing method emphasizes in the lower range of form I concentrations, it showed excellent predictive performance also in the high concentration range of form I.

References

[1] M. Haisa, S. Kashino, H. Maeda, Acta Cryst. B 30 (1974) 2510–2512.
[2] P. Di Martino, A.M. Guyot Hermann, P. Conflant, M. Drache, J.C. Guyot, Int. J. Pharm. 128 (1996) 1–8.
[3] N. Al Zoubi, I. Nikolakakis, S. Malamataris, J. Pharm. Pharmacol. 54 (2002) 325–333.
[4] G. Nichols, C.S. Frampton, J. Pharm. Sci. 87 (1998) 684–693.
[5] N. Al Zoubi, K. Kachrimanis, S. Malamataris, Eur. J. Pharm. Sci. 17 (2002) 13–21.
[6] M.E. Auer, U.J. Griesser, J. Sawatzki, J. Mol. Struct. 661–662 (2003) 307–317.
[7] S.N. Campbell Roberts, A.C. Williams, I.M. Grimsey, S.W. Booth, J. Pharm. Biomed. Anal. 28 (2002) 1135–1147.

[8] N. Al-Zoubi, J.E. Koundourellis, S. Malamataris, J. Pharm. Biomed. Anal. 29 (2002) 459–467.
[9] D. Pratiwi, J.P. Fawcett, K.C. Gordon, T. Rades, Eur. J. Pharm. Biopharm. 54 (2002) 337–341.
[10] C. Strachan, D. Pratiwi, K. Gordon, T. Rades, J. Raman Spectrosc. 35 (2004) 347–352.
[11] T. Ueno, K. Urakami, A. Higashi, K. Umemoto, M. Godo, K. Kitamura, Yakugaku Zasshi 125 (2005) 807–814.
[12] B. Ivanova, J. Mol. Struct. 738 (2005) 233–238.
[13] K. Kipouros, K. Kachrimanis, I. Nikolakakis, S. Malamataris, Anal. Chim. Acta 550 (2005) 191–198.
[14] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137–148.
[15] S. Wold, H. Antti, F. Lindgren, J. Öhman, Chemometr. Intell. Lab. Syst. 44 (1998) 175–185.
[16] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Appl. Spectrosc. 43 (1989) 772–777.
[17] P. Geladi, D. MacDougall, H. Martens, Appl. Spectrosc. 39 (1985) 491–500.
[18] M.S. Dhanoa, S.J. Lister, R. Sanderson, R.J. Barnes, J. Near Infrared Spectrosc. 2 (1994) 43–47.
[19] I.S. Helland, T. Naes, T. Isaksson, Chemometr. Intell. Lab. Syst. 29 (1995) 233–241.
[20] M. Stone, J. R. Stat. Soc. B38 (1977) 44–47.