Journal of Pharmaceutical and Biomedical Analysis 51 (2010) 512–520

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Simultaneous quantitative analysis of mebendazole polymorphs A–C in powder mixtures by DRIFTS spectroscopy and ANN modeling

K. Kachrimanis*, M. Rontogianni, S. Malamataris

Department of Pharmaceutical Technology, School of Pharmacy, University of Thessaloniki, 54124 Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 13 May 2009
Received in revised form 24 August 2009
Accepted 2 September 2009
Available online 10 September 2009

Keywords:
FTIR spectroscopy
Mebendazole
Crystal polymorphism
Quantitative analysis
Artificial neural networks
Partial least squares regression

ABSTRACT

In the present study, a simple method, based on diffuse reflectance FTIR spectroscopy (DRIFTS) and artificial neural network (ANN) modeling is developed for the simultaneous quantitative analysis of mebendazole polymorphs A–C in powder mixtures. Spectral differences between the polymorphs are elucidated by computationally assisted band assignments on the basis of quantum chemical calculations, and subsequently, the spectra are preprocessed by calculation of 1st and 2nd derivatives. Then ANN models are fitted after PCA compression of the input space. Finally the predictive performance of the ANNs is compared with that of PLS regression. It was found that simultaneous quantitative analysis of forms A–C in powder mixtures is possible by fitting an ANN model to the 2nd derivative spectra even after PCA compression of the data (RMSEP of 1.75% for form A, 1.85% for B, and 1.65% for C), while PLS regression, applied for comparison purposes, results in acceptable predictions only within the 700–1750 cm^{-1} spectral range and after direct orthogonal signal correction (DOSC), with RMSEP values of 2.69%, 2.68%, and 3.40% for forms A, B, and C, respectively. Application of the ANN to commercial samples of raw material and formulation (suspension) proved its suitability for the prediction of polymorphic content.

1. Introduction

Mebendazole (5-benzoyl-2-benzimidazolecarbamic acid methyl ester) is a commonly used broad spectrum anthelminthic drug [1] with potential antitumor activity [2], administered orally as a tablet formulation and suspension. The European Pharmacopoeia [3] mentions its polymorphism without further information on the number of the existing polymorphs and the therapeutically preferred one.

It is known that mebendazole exists in three anhydrous crystal forms, designated A–C [4], with very low water solubility, in the order: B > C > A [4,5]. According to several reports, form A has no anthelminthic effect and renders the medication totally inactive when it exceeds 30% of mebendazole content in the formulation [6,7], while the use of form B has been associated with increased toxicity [5,8]. Form C is the preferred polymorph in pharmaceutical products, since it lacks the disadvantages of forms A and B, and it is thermodynamically stable at temperatures up to 180 °C [4], but presence of both polymorphs A and B has been reported in pharmaceutical raw materials [9], commercial oral suspensions [10] and tablet formulations as well [9,11]. Form A may occur either due to the followed process of synthesis or due to prolonged thermal exposure of raw pharmaceutical material or commercial product because it is the stable modification at higher temperatures [4,12]. Therefore, several vibrational spectroscopic methods have been proposed for the identification and quantitative analysis of forms C and A in powders [10,13,14], tablets [15,16] or suspensions [10], while the European Pharmacopoeia [3] recommends testing by FTIR spectroscopy. However, none of the proposed methods includes quantification of form B or deals with the simultaneous determination of all three known polymorphs. Considering the potential toxicity of formulations containing form B [4,8], and the reported presence in raw pharmaceutical materials or final formulations [9,11], a simple method for the routine simultaneous quantitative analysis of all three forms (A–C), in powder mixtures, could be important.

Generally, the simultaneous spectroscopic determination of polymorphs in multi-component powder mixtures is complicated by the non-additive effects due to light scattering or differences in path length and particle size or shape on the relation of spectral features and concentration [17]. Even when the crystal size of the mixture components is similar, the great differences in crystal morphology give rise to nonlinearities between the spectra and polymorphic content. Appropriate preprocessing and advanced multivariate calibration algorithm allow improved predictions when used in combination with spectroscopic techniques such as FTIR or FT-Raman [18,19]. Therefore, in the present study, in order to develop a simple method for the simultaneous quantitative analysis of mebendazole polymorphs in ternary powder

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

0731-7085/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2009.09.001

mixtures, diffuse reflectance FTIR spectroscopy (DRIFTS) and spectral data preprocessing is combined with artificial neural network (ANN) modeling. The DRIFTS spectroscopy was preferred over the conventional KBr disk technique because it requires minimal sample preparation that avoids vigorous grinding or compression of the powder that could induce a polymorphic transition, while in comparison to other spectroscopic techniques that require no sample preparation, such as FT-Raman and NIR, it has the additional advantage of being more widely available due to affordable instrumentation, and less sensitive to particle size effects, respectively [20]. For the PLS regression the spectral differences between polymorphs are discussed after computationally assisted band assignment on the basis of quantum chemical calculations. This band assignment, although not directly related to the quantitative analysis of the polymorphs, enhances understanding of spectral differences and rationalizes the selection of suitable spectral windows. Furthermore, the spectra are preprocessed prior to ANN model fitting, and the predictive performance of the ANNs is compared with that of PLS regression on the basis of the root mean squared error of prediction (RMSEP), as well as by linear regression of observed versus predicted concentrations of mixtures containing each polymorph as impurity in equimolar mixture of the other two. Finally, the developed method is applied in the determination of polymorphic purity of mebendazole for commercially available raw material and formulation (suspension).

2. Materials and methods

2.1. Materials

Mebendazole powder (raw material) was kindly donated by Janssen-Cilag (Athens, Greece). Distilled water and chloroform, formic acid and acetic acid of analytical purity purchased from Merck (Darmstadt, Germany) were used as crystallization solvents. Potassium Bromide (KBr) of spectroscopy grade, obtained from Pike (Pike Technologies, WI, USA) was used for the preparation of samples for DRIFTS spectroscopy.

Pure polymorphs A–C were experimentally prepared by using the commercial (Janssen-Cilag) mebendazole powder and applying suitable crystallization methods. They were characterized by X-ray powder diffraction methods and used as standards for the quantitative determination.

Two 30 ml vials of commercially available mebendazole suspension were purchased from a community pharmacy store (Vermox®, 600 mg/30 ml, Janssen-Cilag, vial 1 Lot No. 0709132 and 2 Lot No. 0902028, with expiry dates 09-2012 and 01-2014, respectively).

2.2. Methods

2.2.1. Crystallization of pure polymorphs A–C

Form A was obtained applying the solvent drop grinding method [21]. An amount of 0.5 g of commercial (Janssen-Cilag) mebendazole powder was manually ground using mortar and pestle, with the simultaneous addition of warm acetic acid droplets. The paste formed was dried in a conventional laboratory oven at 60 °C for 24 h before storage in air-tight jars.

Form B was crystallized from chloroform solutions by solvent evaporation. 100 mg of commercial mebendazole powder were dissolved in 200 ml of chloroform at 70 °C, and crystals of form B were produced by evaporating the solvent at low pressure, using a rotary evaporator (Rotavapor, Büchi, Switzerland). Five batches were combined in order to obtain a sufficient amount.

Form C was crystallized by the solvent exchange method, using formic acid as the good solvent, and water as the anti-solvent. A quantity of 1 g of commercial mebendazole powder was dissolved in 50 ml of formic acid at ambient temperature and 200 ml of water were added under stirring. The crystals were collected by vacuum filtration.

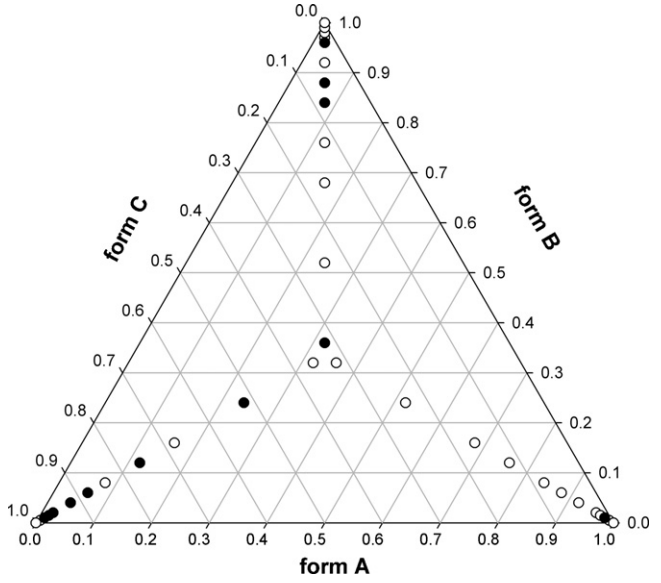

Fig. 1. Ternary plots showing the partitioning of input data into training (empty symbols) and test (full symbols) subsets.

2.2.2. Powder X-ray diffraction (PXRD)

PXRD patterns were recorded on a Siemens D-5000 diffractometer (Siemens AG, Karlsruhe, Germany) equipped with a theta/theta goniometer, a Goebel mirror (Bruker AXS, Karlsruhe, Germany), a 0.15° soller slit collimator and a scintillation counter. The patterns were recorded at a tube voltage of 40 kV and a current of 35 mA applying a scan rate of 0.005° 2θ/s in the angular range of 2–40° 2θ.

2.2.3. Preparation of polymorph mixtures

Each polymorph was successively treated as an impurity in equimolar mixtures of the other two. Three groups of 12 samples each (36 samples in total, Fig. 1) were prepared by geometrically mixing one polymorph with 50:50 (w/w) mixtures of the other two, using the following mass fraction range: 36:32:32, 52:24:24, 68:16:16, 76:12:12, 84:8:8, 88:6:6, 92:4:4, 96:2:2, 97:1.5:1.5, 98:1:1, 99:0.5:0.5, and 100:0:0 (w/w/w). This experimental setup does not cover equally the complete composition range but emphasizes in the low concentrations, since the "real world" problem that this method is expected to deal with would be the quantitation of form A and/or B impurities in form C bulk material.

2.2.4. Diffuse reflectance FTIR spectroscopy (DRIFTS)

DRIFTS spectra were recorded using a Pike diffuse reflectance accessory (Pike Technologies, WI, USA) mounted on a Perkin-Elmer Spectrum GX FTIR spectrometer (Perkin Elmer Inc., USA). The prepared mixtures were diluted to a 5% (w/w) concentration with dry KBr, and spectra were recorded in the range 400–4000 cm^{-1} for a total of 32 scans at a resolution of 4 cm^{-1}. The spectral data from each sample were exported in JCAMP format, and the Know-It-All Informatics System v.5.0 Academic Edition (Bio-Rad Laboratories, Inc.) was used to analyze the spectra.

2.2.5. Dataset partitioning – training and test subset selection

The raw spectral data were partitioned into uniform training and test subsets following the Kennard–Stone design or "uniform mapping algorithm" [22], which is considered to be one of the most efficient methods of optimal data partitioning [23,24].

2.2.6. Data preprocessing

In order to eliminate the effects of noise and nonlinearities of the spectra on the quantitative determination, the following two different preprocessing methods were applied to the spectra separately to the calibration and test subsets: direct orthogonal signal correction (DOSC) and calculation of first and second derivatives. The DOSC algorithm [25] removes variation of the spectral data that is orthogonal (uncorrelated) to the response variable(s). Derivatives eliminate baseline drifts and enhance spectral differences, and for this reason they are commonly used to enhance the very slight differences (small band shifts, shoulders or changes in the multiplicity of peaks) in the otherwise similar FTIR spectra of polymorphs. Second derivative spectra are usually preferred because they have inverse peaks at the same wavenumbers with the original peaks, and therefore they are directly comparable visually to the original spectra. Calculation of second derivatives was applied to the calibration and test subsets separately, after smoothing by the Savitzky–Golay algorithm [26], in order to minimize the noise increase induced by the derivative calculation.

2.2.7. ANN model fitting

In order to reduce the training time of the neural network, the dimensionality of the input space was reduced by Principal Component Analysis (PCA) before ANN model fitting. A feed forward back-propagation artificial neural network with five inputs (corresponding to the first five PCs), twenty hidden and three output units (each corresponding to an individual crystal form) was fitted to the training subset. In all cases, the logistic sigmoid transfer function was used for the input to hidden layer connections, and the linear function was used for the hidden to output layer connections. The networks were trained for 10,000 cycles using the scaled conjugate gradient descent method. The number of training cycles was selected by the "early stopping" method in order to avoid over-fitting, using the test set to monitor the prediction error. Training was repeated 5 times and the network's predictions were averaged, since in ANNs convergence is influenced by initial weight values. The Netlab toolbox for Matlab [27], freely available at http://www.ncrg.aston.ac.uk/netlab/index.php, was used for the application of the ANN models, and the Matlab v.7.0 R14 environment (Mathworks Inc.) was used for the calculations.

2.2.8. PLS model fitting

PLS regression was applied (after direct orthogonal signal correction, DOSC, preprocessing) as a standard method for comparison, since it is a well documented one with excellent predictive ability. The ChemoAC toolbox for Matlab (freely available at http://www.vub.ac.be/fabi/publiek/index.html) was used for PLS model fitting and the optimum number of PLS components was automatically determined by a leave-one-out cross-validation procedure.

2.2.9. Algorithm performance comparison

The accuracy of predictions was assessed on the basis of the root mean squared error of prediction (RMSEP), calculated by

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

as well as by linear regression of observed versus predicted concentration.

All calculations were performed on a PC running MS Windows XP, equipped with an Intel CoreDuo 1.66 GHz processor and 1 GB of RAM.

2.2.10. Application to commercial samples

In order to test the applicability of the ANN method to "real world" problems, content of polymorphs was predicted by fitting the ANN trained as described above to spectral data of the raw material supplied by Janssen-Cilag, as well as of mebendazole powder separated from the marketed suspensions.

For the separation of the powder samples from the suspensions the excipients (water-soluble taste-masking and flavoring agents, and water-insoluble cellulosic polymers used as thickening agents) were removed by centrifugation. After 90 min centrifugation, at 4500 rpm, three zones were formed: a clear supernatant zone, an intermediate opaque zone, and a sediment zone. The clear supernatant liquid zone was removed and disposed. The opaque zone was withdrawn and left to dry, and the sediment was repeatedly washed (three times) by suspending in 30 ml of distilled water and centrifuging at 4500 rpm for 15 min, resulting in the formation of opaque supernatant liquid and a sediment zone. The supernatant opaque liquids combined with the initial intermediate opaque zone, was dried and the resulting solid content was identified as a mixture of mebendazole with very high content of cellulosic gel-forming polymers. The washed sediment was also left to dry before mixing with KBr and used for DRIFTS spectroscopic determination.

3. Results and discussion

3.1. PXRD characterization

Fig. 2 shows the powder X-ray diffractograms of the commercial mebendazole powder and the prepared polymorphs (A–C). Table 1 lists the corresponding peak positions and relative intensities for the pure polymorphs. The diffractogram of commercial mebendazole powder (raw material, Fig. 2) shows the presence of a small amount of form A (indicated by a downward pointing arrow at 7.66° 2θ), underlining the problem of contamination with undesirable crystal forms [9,11]. Also, Fig. 2 shows that a characteristic feature of the diffractogram of form B is the low intensity of the peaks and the extensive line broadening, probably due to small crystallite size and/or micro-strain in the crystal lattice. This may be attributed to the low solubility of mebendazole in chloroform (<1 mg/ml) [28], and the crystallization of form B at high supersaturation, leading to high nucleation and growth rates, which may be responsible for the formation of many small crystals, and addition-

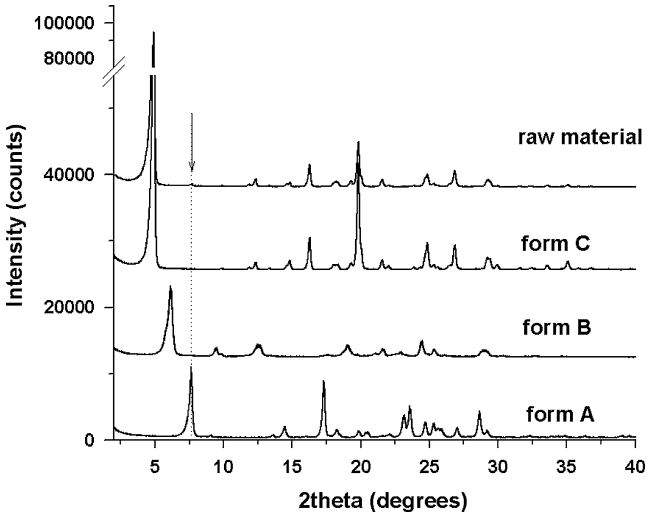

Fig. 2. Powder X-ray diffractograms of mebendazole polymorphs A–C, and commercial powder (raw material).

Table 1
PXRD interplanar distances (*d*-spacings) and relative intensities of the prepared polymorphs (A–C) together with those reported in the literature [12].

Form A				Form B				Form C			
Acetic acid		de Villiers et al.		Chloroform		de Villiers et al.		Formic acid-water		de Villiers et al.	
d (Å)	I/I_0	*d* (Å)	I/I_0	*d* (Å)	I/I_0	*d* (Å)	I/I_0	*d* (Å)	I/I_0	*d* (Å)	I/I_0
11.53	100.0	11.52	100	20.89	5.1	–	–	18.01	100.0	17.91	72
9.74	5.2	–	–	14.31	100.0	14.62	60	7.17	2.2	7.19	28
6.51	5.4	–	–	11.56	4.4	–	–	5.97	2.8	–	–
6.12	16.7	6.13	25	9.34	14.1	9.34	85	5.44	8.9	5.45	51
5.11	77.5	5.13	70	9.00	6.5	–	–	4.92	1.7	–	–
4.86	12.7	4.87	19	7.07	18.6	7.09	49	4.87	1.7	4.89	30
4.47	10.8	4.49	13	6.59	3.6	–	–	4.60	2.1	4.59	32
4.35	9.6	4.35	11	5.05	4.6	–	–	4.48	28.4	4.48	100
4.03	6.1	–	–	4.66	18.0	4.65	100	4.12	2.9	4.16	28
3.84	31.4	3.84	38	4.50	4.8	–	–	3.58	7.7	3.60	56
3.77	43.5	3.78	47	4.22	6.7	4.22	47	3.52	1.7	–	–
3.60	22.3	3.61	34	4.11	12.8	4.13	61	3.36	1.7	3.34	73
3.52	19.9	3.53	23	3.88	8.1	3.94	46	3.32	7.0	–	–
3.47	14.9	–	–	3.64	22.3	3.64	68	3.05	3.8	3.09	36
3.30	14.8	–	–	3.51	12.4	3.54	51	2.98	1.7	–	–
3.11	36.2	–	–	3.41	3.9	–	–	2.67	1.6	–	–
3.05	11.1	–	–	3.09	11.2	3.09	43	2.56	2.6	–	–
–	–	–	–	3.06	10.9	–	–	–	–	–	–
–	–	–	–	2.94	3.0	–	–	–	–	–	–

ally, for increased presence of defects. In general, the PXRD patterns of the three prepared forms (A–C) are in agreement with literature data [12], confirming the polymorphic purity of the standard materials.

3.2. DRIFTS spectroscopy

Fig. 3 illustrates DRIFTS spectra of the three prepared polymorphs of mebendazole (A–C), together with the theoretically calculated spectrum of mebendazole molecule in the gas phase (shown in absorbance mode). It is seen that the spectrum of form C has a more pronounced fine structure with well resolved, sharper peaks compared to the spectra of forms A and B. This is indicative of higher vibrational liberty, and most probably existence of mebendazole in more than a single molecule in the asymmetric unit. The spectra of the three polymorphs differ significantly in the fingerprint region (1500–400 cm^{-1}), but the most characteristic differences lie in the range of the N–H (3340–3400 cm^{-1}) and C=O (1700–1730 and 1635–1645 cm^{-1}) stretching vibrations, suggesting that those chemical groups form different hydrogen bonding interactions and are responsible for the different packing in the crystal lattice of the three polymorphs. Moreover, differences in the spectra of polymorphs A and C have been attributed to tautomerism [10].

For better understanding of the spectral differences between the three polymorphs, computationally assisted assignment of fundamental frequencies was performed by ab initio vibrational normal mode analysis for the lowest energy tautomer of mebendazole [10]. The starting geometry was represented in delocalized internal coordinate format [29], and was optimized by Density Functional Theory (DFT) calculations using the Becke, Lee, Young, and Parr (B3LYP) hybrid exchange-correlation functional [30] and the 6-31G* basis set. The Hessian matrix was numerically calculated at the same level of theory and normal modes of vibration were obtained within the harmonic approximation. Convergence to geometry of minimum energy was verified by the absence of imaginary vibrational modes (i.e. negative eigenvalues in the Hessian matrix). The calculated frequencies were scaled by a

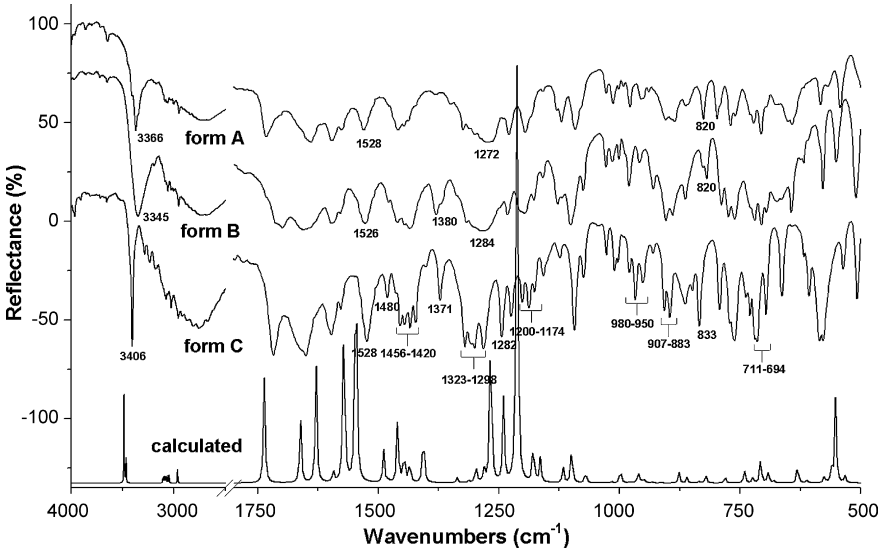

Fig. 3. DRIFTS spectra of mebendazole polymorphs A–C (most important polymorph-specific peaks are labeled), together with the theoretically calculated spectrum (shown in absorbance mode).

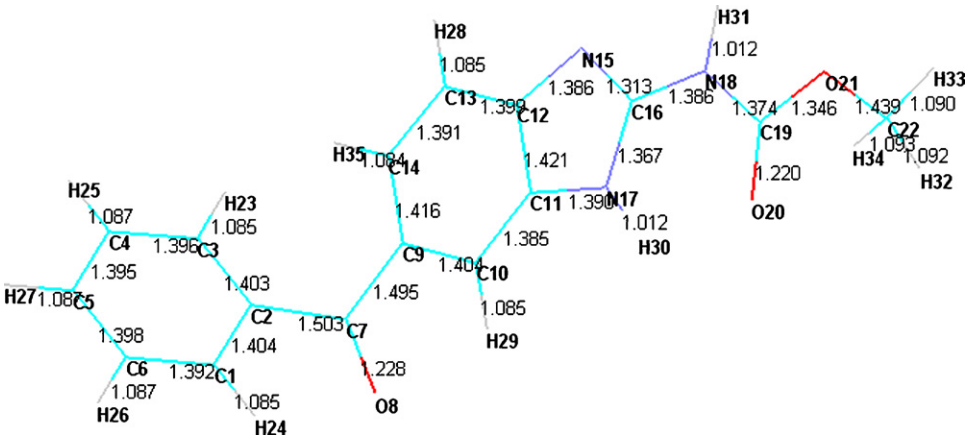

Fig. 4. Optimized molecular structure of mebendazole used for the normal coordinate analysis (atom numbering and bond lengths shown).

factor of 0.9614 [31]. Normal coordinate analysis [32] was performed in order to calculate the potential energy distribution (PED) and use it for the assignment of the vibrational modes. The PC GAMESS/Firefly [33] was used for the electronic structure calculations and subsequent vibrational analysis (normal coordinate analysis and calculation of the PED). All calculations were performed on a PC running MS Windows XP, equipped with an Intel CoreDuo T2300E at 1.66 GHz processor and 1 GB of RAM.

The optimized molecular structure is given in Fig. 4, and in Table 2 are listed the polymorph-specific experimental peaks, together with the computationally assisted band assignment of fundamental frequencies and the corresponding potential energy distribution (% PED). It should be noted that in large molecules like mebendazole the normal vibrational modes are highly complicated, consisting of the concerted motion of many different atoms. Therefore, following Strachan et al. [34], band assignments are proposed on the basis of percent PED and refer to portions of the molecule (benzimidazole, BZIMD, and benzoyl, Bz, rings, and methyl-carbamate, mCBM, moiety).

In Table 2, it is seen that the most profound polymorph-specific features in the FTIR spectra lie in the range of 3340–3400 cm^{-1}, which are attributed to the carbamate ν(N–H) stretching mode,

Table 2
Selected experimental FTIR frequencies of mebendazole polymorphs and corresponding frequency and band assignments of the DFT/6-31G calculated normal modes of the most stable tautomer, together with the potential energy density distribution (PED %).

FTIR frequency (cm^{-1})				Band assignment/PED (%)
Crystal polymorph			DFT/6-31G frequency	
A	B	C		
–	–	–	3485.5	BZIMD ν(N–H) 99.6
3366	3345	3402	3467.1	mCBM ν(N–H) 99.8
1730	1697	1715	1736.4	mCBM ν(C=O) 77.5
1634	1643	1643	1661.4	Bz ν(C=O) 81.6
1528	1526	1520	1546.6	BZIMD ring def. 70.5
–	–	1480	1546.6	BZIMD ring def. 70.5
1456	1456	1452	1460.1	mCBM δ(X–H) + torsion 27.0, BZIMD ring def. 26.7
–	–	–	1454.8	mCBM torsion 85.1
1447	1445	1445	1446.5	BZIMD ν(X–C) 20.3 + δ(C–H) 20.8, mCBM δ(C–H) 15.6
	1433	1431	1435.3	mCBM δ(C–H) 26.2 BZIMD ring def. 60.4
1422	–	1422	1434.5	Bz δ(C–H) 62.3
–	1379	1371	1406.6	BZIMD ν(N–C) 20.4, mCBM δ(X–H) 39.7 + ν(N–C) 10.4
1323	1317	1317	1313.0	Bz ν(C–C) 39.0 + δ(C–H) 50.4
1312	–	1308		–
1298	–	1296	1297.8	Bz ν(C–C) 32.7 + δ(C–H) 11.1, BZIMD δ(C–H) 4.0
1258	1269	1273	1267.1	BZIMD ν(C–N) 40.0 + ν(C–C) 33.9
1229	1232	1224	1212.9	mCBM δ(N–H) 27.2 + ν(C–N) 10.5 ν(C–O) 26.3
1192	1194	1200	1180.2	mCBM δ(C–H) 47.9 + τ(C–O–C–H) 10.0
1180	–	1186	1175.9	BZIMD δ(N–H) 30.6, Bz δ(C–H) 4.7
–	1177	1174	1164.2	Bz δ(C–H) 71.6 + ν(C–C) 15.0
978	978	980	965.1	Bz τ(C–C–C–H) 75.9
–	–	966	959.5	Bz δ(C–H) 10.9, BZIMD ring def. 32.7, mCBM ν(C–O) 7.7
955	957	951	949.9	mCBM ν(C–O) 45.9 + ν(C–N) 7.0, BZIMD ν(C–N) 5.3
903	903	907	913.6	Bz τ(C–C–C–H) 93.7
883	887	893	876.9	BZIMD τ(N–C–C–H) 70.0 + δ(C–C) 15.0
826	826	833	831.3	BZIMD δ(C–N) 43.5 + ν(C–C) 27.6
–	820	–	821.2	BZIMD τ(C–C–C–C) 39.2 + τ(C–C–C–H) 53.6
704	704	712	707.3	Bz τ(C–C–C–O) 8.5 + τ(C–C–C–H) 7.1 + τ(C–C–C–C) 6.5
–	–	–	693.4	BZIMD τ(C–N–C–N) 68.3
–	694	694	691.0	BZIMD τ(C–N–C–N) 43.1 + Bz τ(C–C–C–H) 20.2
584	585	585	560.9	BZIMD τ(C–C–N–H) 12.2 + τ(N–C–N–H) 73.8
568	578	578	552.9	BZIMD τ(C–C–N–H) 59.8 + τ(N–C–N–H) 19.8

ν: stretching; δ: bending; τ: torsion; def.: deformation; X = C, O; BZIMD: benzimidazole; Bz: benzoyl; mCBM: methyl carbamate.

and in the range 1630–1648 cm^{-1}, and 1700–1730 cm^{-1} that are attributed to the ν(C=O) stretching modes of the benzoyl and carbamate groups, respectively. Additionally, Table 2 shows that the frequency of the carbamate ν(N–H) stretching mode is lower than the theoretically expected (~3467 cm^{-1}) in all three polymorphs. It is remarkable that a single ν(N–H) peak, attributed to the carbamate N–H group, is present in the spectra of all three forms (at 3366, 3345, and 3402 cm^{-1} for forms A, B, and C, respectively). Based on the molecular structure of mebendazole, one more ν(N–H) mode is expected, arising from the benzimidazole N–H at or near 3480 cm^{-1}. The absence of such vibration in the solid state FTIR spectra indicates that the N–H group is involved in strong intermolecular interactions, possibly with the additional contribution of a weak intra-molecular hydrogen bond between O20···H30, shown in Fig. 4. Although the crystal structure of mebendazole is yet unknown and therefore the topology and strength of inter- and intra-molecular hydrogen bonds in the solid state (including this hydrogen bond) can only be speculated, a Cambridge Structural Database [35] search of benzimidazole carbamate fragment may be indicative of possible interactions. The search revealed four neutral benzimidazole carbamate derivatives (CSD Ref. Codes: BEWLOF, KEZJII, SAGQEW, VELDUL), and several salts (BMCBIB, QESBOG, RAVLAC, RAVLEG, NIZCIJ). In both neutral and salt structures, the C=O and N–H groups are involved in intermolecular hydrogen bonding interactions. Additionally, in the four neutral benzimidazol carbamate derivatives, the carbamate group acquires a trans-conformation, which facilitates the formation of the intra-molecular hydrogen bond with an O···H average distance ~2.18 Å. Therefore, it is expected that both inter- and intra-molecular hydrogen bonds exist in the structure of mebendazole, leading to the conclusion that the disappearance of the benzimidazole N–H vibration should be the result of strong intermolecular interactions with a small contribution of the O20···H30 intra-molecular hydrogen bond. This is further supported by the absence of the N–H stretching mode from the IR spectrum of crystalline benzimidazole [36], where only intermolecular hydrogen bonding is possible.

The lowered frequency of the ν(N–H) stretching mode compared to the theoretically expected value (at ~3467 cm^{-1}) in all three polymorphs indicates that the methyl-carbamate N–H group, is involved in hydrogen bonding interactions. The ν(N–H) frequencies follow the order: C > A > B, suggesting that in form C the carbamate N–H is involved in weaker intermolecular interactions, followed by forms A and B, with B showing significant peak broadening. Ayala et al. [37] have pointed out that the ν(N–H) and ν(C=O) frequencies of the three polymorphs follow a different order (A > C > B). The methyl-carbamate C=O group of form C seems to be involved in stronger hydrogen bonding interactions than that of form A, but weaker than those of form B, while the benzoyl C=O group of form A is the one that is involved in strongest hydrogen bonding interactions. This inconsistency between the ν(N–H) and ν(C=O) modes was attributed to the competition of intra- and intermolecular hydrogen bonds in which those groups are supposedly involved.

Another very informative region is that of the δ(N–H) in-plane bending modes (1528 and 1526 cm^{-1} in forms A and B, respectively, and 1520 cm^{-1} and 1480 cm^{-1} in form C). The lower δ(N–H) frequency of form C indicates a weaker hydrogen bonding interaction of this group.

Other vibrational modes with profound differences between polymorphs include:

(a) the methyl group bending modes, δ(C–H), in the range 1420–1456 cm^{-1}, where form C shows a quadruple peak instead of the triplet observed in the spectra of forms A and B (Fig. 3),
(b) the ν(C–N) stretch which appears at 1380 cm^{-1} in form B, at 1371 cm^{-1} in form C and is absent from the spectrum of form A,
(c) the in-plane bending of the benzoyl group aromatic hydrogens in the region 1298–1323 cm^{-1}, where C shows well defined, intense peaks instead of shoulders in the spectra of forms A and B,
(d) the in-plane bending of benzimidazole hydrogens with some contribution of the ν(C=N) stretch at 1272 cm^{-1} for form A, and 1282–1284 cm^{-1} (within spectral resolution limits) for forms B and C,
(e) the methyl δ(C–H) bending, the benzimidazole and methyl-carbamate δ(N–H) bending, and benzoyl aromatic hydrogen in-plane δ(C–H) bending modes in the range 1174–1200 cm^{-1}, where form C shows a well defined triplet instead of a broad peak with a small shoulder seen in the spectra of forms A and B,
(f) the benzoyl group aromatic hydrogen out-of-plane and benzimidazole hydrogen in-plane bending modes, δ(C–H), in the range 950–980 cm^{-1}, where form C shows a triplet instead of the double peak of forms A and B, as well as in the range 883–907 cm^{-1}, and 820–833 cm^{-1} (Fig. 3),
(g) and finally, some out-of-plane bending modes, δ(C–H) and δ(N–H), of the benzoyl and benzimidazole rings at 694–711 cm^{-1}.

From the above discussion the spectral zone for optimal PLS model may be indicated and furthermore it becomes clear that polymorph-specific features are widely distributed in the FTIR spectra and therefore a multivariate calibration approach should be followed for the quantitative analysis of polymorphic mixtures.

3.3. Simultaneous quantitative analysis

Preliminary tests have shown that calculation of 2nd derivative of the spectra and ranging to [0,1] led to optimal performance of an ANN model with 5 input, 20 hidden, and 3 output units, while orthogonal signal correction (DOSC) preprocessing using 2 OSC components and selection of the 700–1750 cm^{-1} spectral range was necessary to develop optimal PLS models that had 2 PLS components. The omitted spectral range (1750–4000 cm^{-1}) is dominated by the ν(C–H) modes (seen as a broad peak between 2000 and 3300 cm^{-1}) and the ν(N–H) modes (3200–3400 cm^{-1}). Improvement of the PLS model's goodness of fit when the ν(X–H) vibrations are omitted from the data indicates that these particular vibrations are mainly responsible for the noise and nonlinearity in the reflectance–concentration relationship.

In Table 3 is listed the observed polymorphic contents (%) of forms A–C, together with the values predicted by the ANN and the PLS model, and in Table 4 is listed the corresponding RMSEP (%), and linear regression parameters (coefficients of determination, slopes and intercepts) of observed vs predicted concentrations on the test subset selected by the Kennard–Stone algorithm, together with the limit of detection (LOD) calculated by the product 3× RMSEP, which is considered by Faber et al. [38] as a reasonable approximation of the LOD. From Table 3 it is seen that the ANN model performs considerably better than the PLS model in the low content range, which means that it is more suitable than the PLS model for the quantitation of impurities.

The superior performance of the ANN model in all three forms may be attributed to its training on the complete spectra and to the data preprocessing by a more generally available and simpler algorithm (2nd derivative) than the direct Orthogonal Signal Correction used for PLS. Even though the preprocessed data used for the ANN model have subsequently been compressed by PCA,

Table 3
Observed (experimental) contents (%) of forms A–C, together with the values predicted by the ANN model trained using the full spectral range, and by the PLS model fitted to the 700–1750 cm^{-1} range.

Mixture number	Form A			Form B			Form C		
	Obs. (%)	Pred. (%)		Obs. (%)	Pred. (%)		Obs. (%)	Pred. (%)	
		ANN[a]	PLS		ANN[a]	PLS		ANN[a]	PLS
1	98	97.9	96.7	1	1.0	1.7	1	1.1	1.5
2	32	28.4	29.2	36	39.6	38.9	32	31.9	32.0
3	8	9.0	5.8	84	82.8	85.0	8	8.3	9.2
4	6	5.1	10.3	88	89.3	82.2	6	5.6	7.5
5	2	0.0	4.7	96	99.2	95.0	2	0.8	0.4
6	24	21.4	24.9	24	23.4	26.1	52	55.1	49.0
7	12	12.9	15.1	12	12.2	10.7	76	74.9	74.3
8	6	6.6	7.4	6	5.4	7.4	88	88.1	85.2
9	4	2.4	2.1	4	2.5	6.1	92	95.1	91.8
10	2	1.9	6.4	2	0.5	4.4	96	97.7	89.2
11	1.5	1.7	2.3	1.5	−0.3	4.4	97	98.8	93.3
12	1	0.9	4.4	1	0.8	4.9	98	98.3	90.8

[a] Average of five replications.

Table 4
Root mean squared error of prediction (RMSEP %) together with the linear regression parameters of observed vs predicted concentrations (correlation coefficient, *R*, slope, and intercept), and limit of detection (LOD) for the ANN model and by the PLS model for the range 700–1750 cm^{-1}.

Polymorph/model	RMSEP (%)	*R*	Slope	Intercept	LOD (%)
Form A					
ANN[a]	1.7494 (0.2003)	0.9982 (0.0004)	1.0015 (0.0079)	0.6507 (0.3627)	5.25
PLS	2.6972	0.9961	1.0341	−1.6485	8.09
Form B					
ANN[a]	1.8514 (0.2087)	0.9991 (0.0003)	0.9736 (0.2775)	0.7051 (0.2775)	5.55
PLS	2.6841	0.9984	1.0432	−2.2533	8.05
Form C					
ANN[a]	1.6517 (0.1126)	0.9994 (0.0001)	0.9828 (0.0044)	0.2975 (0.2882)	4.96
PLS	3.4028	0.9988	1.0511	−0.6739	10.21

[a] Average of five replications and standard error in parentheses.

which is expected to inflict some loss of information and deterioration of the predictive performance, the ANN model's performance in all three crystal forms does not show deterioration or dependence on sample characteristics, in contrast to PLS. Therefore it is concluded that it is superior to PLS and should be preferred in the simultaneous quantitation of polymorphs, particularly in low concentrations.

Regarding computation times, training of the PLS model is extremely fast (<1 s). However, significant reduction of the input variables by the PCA compression (the first five principal components are used as inputs) reduces the computation cost greatly (<3 min), and the trade-off between prediction accuracy and computation speed is reasonable. A comparison with methods published in the literature in terms of predictive performance shows that all published methods consider only forms A and C, to levels ~10% (w/w), while the present method allows for the quantitation of all three forms down to <2% (w/w). Considering that form A is usually present in raw materials and finished products at concentrations well below 10% and additionally that form B has also been found as a common impurity in form C [9], the advantages of the developed ANN method are profound. Additionally, although it is not unexpected that the highly complex nonlinear ANN models will perform much better than conventional linear multivariate calibration models, the bottleneck in the application of an ANN model is the availability of training samples. Generally, ANNs require a reasonably large number of training samples, while the use of an external validation set – although not necessary – greatly improves the quality of the model. However, the number of calibration samples is usually kept at a minimum, not to mention that mixture compositions at the impurity level are avoided due to difficulties in sample preparation. Therefore, the superiority of the ANN quantitation method of this study underlines the importance of selecting appropriate ranges of mixture compositions for calibration, and that a small increase in the number of calibration samples can facilitate the use and greatly enhance the performance of a more advanced calibration model such as an ANN.

3.4. Application to commercial samples

Table 5 lists the percent concentration of forms A–C predicted by the trained ANN, and Fig. 5 illustrates the normalized reflectance difference of mebendazole raw powder material and form C plotted together with that of forms A and C, as well as the normalized reflectance difference of commercial suspension 1 and form C plotted together with the difference of forms B and C (suspension 2 is not shown because the result was almost identical to that of suspension 1). The ANN for the case of the commercial powder (raw material) predicts a form A content of 5.3% (Table 5), slightly above the detection limit, and a nearly 0% content of form B, which is in agreement with the powder X-ray diffractogram of Fig. 2, and

Table 5
Predicted concentration of forms A–C in the mebendazole powder (raw material) and two vials of commercially available suspensions (average of three determinations ± standard deviation).

Sample	Average concentration (%)		
	Form A	Form B	Form C
Raw material	5.3 ± 2.1	−0.6 ± 1.3	95.6 ± 0.9
Suspension 1	22.7 ± 2.7	8.4 ± 5.6	68.7 ± 3.3
Suspension 2	22.1 ± 2.5	10.7 ± 5.9	66.7 ± 4.8

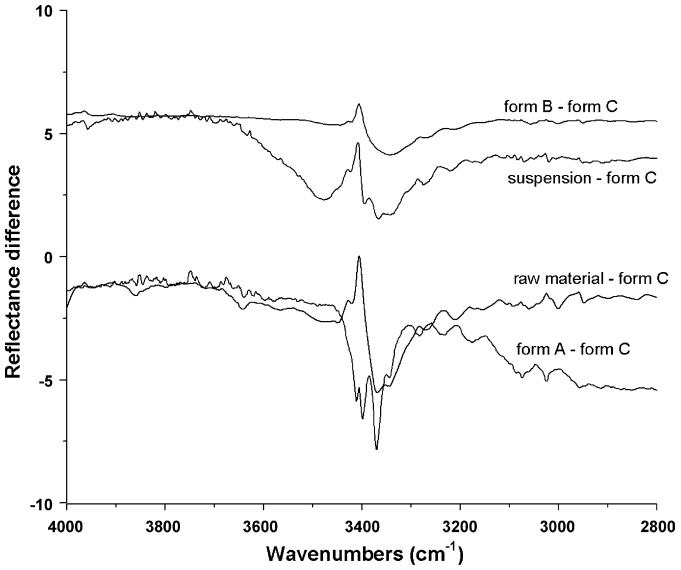

Fig. 5. Normalized reflectance difference of commercial mebendazole powder (raw material) and form C plotted together with that of forms A and C, as well as the normalized reflectance difference of commercial suspension 1 and form C plotted together with the difference of forms B and C.

the absence of similarity in the normalized reflectance differences (Fig. 5). In the case of the commercial suspensions, the ANN model predicts a rather high content of form A (~20%), and about 10% content of form B. This finding is in agreement with the similarity in the normalized reflectance differences (Fig. 5), where subtraction of the spectrum of form C from that of the powder separated from the suspension 1 leaves characteristic features of the spectra of forms A and B. Form A is the most common impurity and its presence is expected because it is the thermodynamically stable form. However, the presence of the least stable form B in the suspension is unexpected, and it could be the result of a solvent-induced transformation due to the prolonged contact with the solvent and the high supersaturation conditions. Although the presence of forms A and B in the commercial suspensions is supported by the spectral data, the predicted amounts should be treated with caution, because the effects of the very small (sub-micron) particle size of mebendazole and of possible trace amounts of cellulosic excipients require further elucidation.

4. Conclusions

Simultaneous quantitative analysis of mebendazole forms A–C, in powder mixtures, is possible by fitting an ANN model to the complete spectral range after 2nd derivative preprocessing and PCA compression of the data (RMSEP of 1.75% for form A, 1.85% for B, and 1.65% for C). The PLS regression results in acceptable predictions but only after direct orthogonal signal correction (DOSC) in the 700–1750 cm^{-1} range, i.e. omitting the range dominated by the ν(X–H) modes that is more susceptible to noise. Performance of the PLS model deteriorates in the case of form C, showing dependence on crystalline sample differences (RMSEP values of 2.69%, 2.68% for forms A, and B, but 3.40% for form C). The ANN model on the other hand possesses significant advantages over PLS, requiring simpler preprocessing (2nd derivative) and making use of the complete data, while being less sensitive to noise and sample effects, and having a more consistent predictive performance for all the mebendazole polymorphs (A–C). The ANN method is suitable for the prediction of the polymorphic content of commercial powdered raw materials and suspensions.

Acknowledgements

The authors thank Prof. Ulrich J. Griesser of the University of Innsbruck for the PXRD measurements.

References

[1] A. Raeymackers, J. Van Gelder, L. Roevens, P. Janssen, Synthesis and anthelminthic activity of alkyl-(5-acyl-1-benzimidazol-2-yl) carbamates, Arzneim.-Forsch./Drug Res. 28 (1978) 586–594.
[2] T. Mukhopadhyay, J. Sasaki, R. Ramesh, J. Roth, Mebendazole elicits a potent antitumor effect on human cancer cell lines both in vitro and in vivo, Clin. Cancer Res. 8 (2002) 2963–2969.
[3] European Pharmacopoeia, 5th edition, Monographs, Council of Europe, cop. Strasbourg, 2008, pp. 1981–1982.
[4] M. Himmelreich, B.J. Rawson, T.R. Watson, Polymorphic forms of mebendazole, Aust. J. Pharm. Sci. 6 (1977) 123–125.
[5] J. Costa, M. Fresno, L. Guzman, A. Igual, J. Oliva, P. Vidal, A. Perez, M. Pujol, Polymorphic forms of mebendazole: analytical aspects and toxicity, Circ. Farm. 49 (1991) 415–424.
[6] H. Ren, K.J. Luo, B.Z. Cheng, Effects of various anti-hookworm drugs on the cholinesterase activity of *Necator americanus*, *Ancylostoma duodenale*, *Ancylostoma caninum* and *Nippostrongylus braziliensis*, Chin. J. Parasitol. Parasit. Dis. 5 (1987) 13–17.
[7] P. Charoenlarp, J. Waikagul, C. Muennoo, S. Srinophakun, D. Kitayaporn, Efficacy of single-dose mebendazole, polymorphic forms A and C, in the treatment of hookworm and Trichuris infections, J. Trop. Med. Publ. Health 24 (1993) 712–716.
[8] F. Rodriguez-Caabeiro, A. Criado-Fornelio, A. Jimenez-Conzalez, L. Guzman, A. Igual, A. Perez, M. Pujol, Experimental chemotherapy and toxicity in mice of three mebendazole polymorphic forms, Chemotherapy 33 (1987) 266–271.
[9] W. Liebenberg, T.G. Dekker, A.P. Lötter, M.M. de Villiers, Identification of the mebendazole polymorphic form present in raw materials and tablets available in South Africa, Drug Dev. Ind. Pharm. 24 (1998) 485–488.
[10] S. Agatonovic-Kustrin, B. Glass, M. Mangan, J. Smithson, Analysing the crystal purity of mebendazole raw material and its stability in a suspension formulation, Int. J. Pharm. 361 (2008) 245–250.
[11] P. Froelic, F. Gasparotto, Mebendazole: identification of polymorphs of the drug in several different bulk substances and dosage forms (reference brand and generic) available in Brazil, Rev. Sci. Farm. Basica Aplicada 26 (2005) 205–210.
[12] M. de Villiers, R. Terblanche, W. Liebenberg, E. Swanepoel, T. Dekker, M. Song, Variable-temperature X-ray powder diffraction analysis of the crystal transformation of the pharmaceutically preferred polymorph C of mebendazole, J. Pharm. Biomed. Anal. 38 (2005) 435–441.
[13] Z. Sha, W. Sun, H. Gao, Infrared spectrometry determination of polymorphic form C in mebendazole, Yaoxue Xuebao Acta Pharm. Sinica 24 (1989) 932–936.
[14] B. Glass, S. Agatonovic-Kustrin, M. Brown, M. Mangan, E. Mopp, M. Aereboe, Analyzing the crystal purity of mebendazole raw material by DRIFT spectrometry and ANN spectral modelling, in: 7th International Conference/Workshop on Pharmacy and Applied Physical Chemistry Phandta 7 Innsbruck, Austria, September 07–11, 2003.
[15] A. Bunaciu, S. Fleschin, H.Y. Aboul-Enein, Analysis of mebendazole polymorphs by Fourier transform infrared spectrometry using chemometric methods, Spectrosc. Lett. 34 (2001) 527–536.
[16] H. Aboul-Enein, A. Bunaciu, S. Fleschin, Analysis of mebendazole polymorphs by Fourier transform IR spectrometry using chemometric methods, Biopolymers (Biospectroscopy) 67 (2002) 56–60.
[17] C. Pecharroman, J. Iglesias, Effect of particle shape on the IR reflectance spectra of pressed powders of anisotropic materials, Appl. Spectrosc. 54 (2000) 634–638.
[18] K. Kipouros, K. Kachrimanis, I. Nikolakakis, V. Tserki, S. Malamataris, Simultaneous quantification of carbamazepine crystal forms in ternary mixtures (I, III, and IV) by diffuse reflectance FTIR spectroscopy (DRIFTS) and multivariate calibration, J. Pharm. Sci. 95 (2006) 2419–2431.
[19] K. Kachrimanis, D.E. Braun, U.J. Griesser, Quantitative analysis of paracetamol polymorphs in powder mixtures by FT-Raman spectroscopy and PLS regression, J. Pharm. Biomed. Anal. 43 (2007) 407–412.
[20] B. Madari, J. Reeves III, P. Machado, C. Guimaraes, E. Torres, G. McCarthy, Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols, Geoderma 136 (2006) 245–259.
[21] A. Trask, W. Motherwell, W. Jones, Solvent-drop grinding: green polymorph control of cocrystallisation, Chem. Commun. 7 (2004) 890–891.
[22] R. Kennard, L. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148.
[23] W. Wu, B. Walczak, D.L. Massart, S. Heuerding, F. Erni, I. Last, K. Prebble, Artificial neural networks in classification of NIR spectral data: design of the training set, Chemom. Int. Lab. Syst. 33 (1996) 35–46.
[24] M. Daszykowski, B. Walczak, D.L. Massart, Representative subset selection, Anal. Chim. Acta 468 (2002) 91–103.
[25] J. Westerhuis, S. de Jong, A. Smilde, Direct orthogonal signal correction, Chemom. Intell. Lab. Syst. 56 (2001) 13–25.
[26] P. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky–Golay) method, Anal. Chem. 62 (1990) 570–573.

[27] I. Nabney, Netlab: Algorithms for Pattern Recognition. Advances in Pattern Recognition, Springer, 2001.
[28] S. Kumar, G. Chawla, M.E. Sobhia, A.K. Bansal, Characterization of solid-state forms of mebendazole, Pharmazie 63 (2008) 136–143.
[29] J. Baker, A. Kessi, B. Delley, The generation and use of delocalized internal coordinates, J. Chem. Phys. 105 (1996) 192–212.
[30] A. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.
[31] A. Scott, L. Radom, Harmonic vibrational frequencies: an evaluation of Hartree–Fock, Møller–Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors, J. Phys. Chem. 100 (1996) 16502–16513.
[32] J. Boatz, M. Gordon, Decomposition of normal-coordinate vibrational frequencies, J. Phys. Chem. 93 (1989) 1819–1826.
[33] A. Granovsky, PC GAMESS/Firefly Version 7.1.E, 2008, http://classic.chem.msu.su/gran/gamess/index.html.
[34] C. Strachan, S. Howell, Th. Rades, K. Gordon, A theoretical and spectroscopic study of carbamazepine polymorphs, J. Raman Spectrosc. 35 (2004) 401–408.
[35] F. Allen, The Cambridge Structural Database: a quarter of a million crystal structures and rising, Acta Cryst. B 58 (2002) 380–388.
[36] M. Morsy, A. Al-Khaldi, A. Suwaiyan, Normal vibrational mode analysis and assignment of benzimidazole by ab initio and density functional calculations and polarized infrared and Raman spectroscopy, J. Phys. Chem. A 106 (2002) 2196–9203.
[37] A. Ayala, H. Siesler, S. Cuffini, Polymorphism incidence in commercial tablets of mebendazole: a vibrational spectroscopy investigation, J. Raman Spectrosc. 39 (2008) 1150–1157.
[38] N. Faber, D. Duewer, S. Choquette, T. Green, N. Chesler, Characterizing the uncertainty in near-infrared spectroscopic prediction of mixed-oxygenate concentrations in gasoline: sample-specific prediction intervals, Anal. Chem. 70 (1998) 2972–2982.