AAPS PharmSciTech, Vol. 12, No. 2, June 2011 (© 2011)
DOI: 10.1208/s12249-011-9618-x

Research Article

Multivariate Analysis of Phenol in Freeze-Dried and Spray-Dried Insulin Formulations by NIR and FTIR

Morten Jonas Maltesen,[1,5] Simon Bjerregaard,[2] Lars Hovgaard,[3] Svend Havelund,[4] Marco van de Weert,[1] and Holger Grohganz[1]

Received 2 November 2010; accepted 14 April 2011; published online 11 May 2011

Abstract Dehydration is a commonly used method to stabilise protein formulations. Upon dehydration, there is a significant risk the composition of the formulation will change especially if the protein formulation contains volatile compounds. Phenol is often used as excipient in insulin formulations, stabilising the insulin hexamer by changing the secondary structure. We have previously shown that it is possible to maintain this structural change after drying. The aim of this study was to evaluate the residual phenol content in spray-dried and freeze-dried insulin formulations by Fourier transform infrared (FTIR) spectroscopy and near infrared (NIR) spectroscopy using multivariate data analysis. A principal component analysis (PCA) and partial least squares (PLS) projections were used to analyse spectral data. After drying, there was a difference between the two drying methods in the phenol/insulin ratio and the water content of the dried samples. The spray-dried samples contained more water and less phenol compared with the freeze-dried samples. For the FTIR spectra, the best model used one PLS component to describe the phenol/insulin ratio in the powders, and was based on the second derivative pre-treated spectra in the 850–650 cm^{-1} region. The best PLS model based on the NIR spectra utilised three PLS components to describe the phenol/insulin ratio and was based on the standard normal variate transformed spectra in the 6,200–5,800 cm^{-1} region. The root mean square error of cross validation was 0.69% and 0.60% (*w*/*w*) for the models based on the FTIR and NIR spectra, respectively. In general, both methods were suitable for phenol quantification in dried phenol/insulin samples.

KEY WORDS: fourier transform infrared (FTIR) spectroscopy; insulin; multivariate analysis; near infrared (NIR) spectroscopy; volatiles.

INTRODUCTION

Strategies focusing on enhancing protein stability often utilise excipients and dehydration of the protein formulation (1). Upon dehydration, there is a significant risk that the composition of the formulation will change. This is especially an issue when the formulation contains volatile compounds which are frequently used as stabilisers and antimicrobials. The retention of volatiles after dehydration has, to some extent, been described by the microencapsulation theory introduced by Fink and the selective diffusion theory by Thijssen (2).

Insulin is a protein used in the treatment of diabetes, and it is often formulated with phenol and phenol derivatives acting as antimicrobials and stabilising agents (3). At concentrations used in liquid formulations, insulin molecules form dimers, and in the presence of divalent metal ions insulin dimers associate into discrete hexameric structures (4). The hexameric structure acts as an allosteric unit present in one of three global conformations named T_6, R_3T_3 and R_6 (5–10). The structural transition from the T to the R form involves a rearrangement of the first eight amino acid residues in the B chain, from an extended conformation in the T state to an α-helix in the R state. The transition is driven by the presence of phenol and phenol derivatives which act as ligands for the R state, shifting the equilibrium and thus the population probability towards the R state (6–11).

Recently, we have expanded the allosteric concepts of insulin to solid state formulations (12). We showed that it was possible to maintain the increase in the α-helix content of insulin in the presence of phenol after drying. The increase was monitored by Fourier transform infrared (FTIR) spectroscopy. An important part of that study was to correlate the structural changes of the insulin hexamer with the residual amount of phenol present in the formulation after drying. The residual phenol content in the dried samples was evaluated by reverse phase (RP) ultra performance liquid chromatography (UPLC) (12). The advantages of the UPLC methods are high resolution, high sensitivity and a high level of robustness, including reproducibility and accuracy (13). While these advantages in general make it a good analytical method both in research and in production of pharmaceutical products, several drawbacks are also present when utilising chromatographic methods. A compromise must often be made

[1] Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark.
[2] Department of Oral Protein Delivery, Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark.
[3] Department of Oral Formulation Development, Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark.
[4] Department of Diabetes Protein Engineering, Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark.
[5] To whom correspondence should be addressed. (e-mail: mjm@farma.ku.dk)

between resolution and analytical time, and compared to spectroscopic methods, sample preparation is often more labour intensive and time consuming (14). Moreover, they are carried out offline resulting in sample removal and transport. While this is no big concern in research facilities, it represents a severe problem in industrial production facilities. Among other issues, this has been addressed in the recent Process Analytical Technology (PAT) Initiative and the International Conference on Harmonisation Guidelines Q8–Q10 describing a preferred Quality by Design approach to product design and production (15). The initiatives and guidelines have been incorporated by the European Medicines Agency (16) and the US Food and Drug Administration (17) to ensure better products and a faster development time for each product. Basically, the framework aims at a better process understanding and process design by assessing critical quality attributes (CQA) and critical process parameters (CPP) important for the quality of the final product. The CQAs are preferably monitored online, in-line or at line making the analysis more efficient, resulting in a real time evaluation of the process instead of end batch analyses. This result in real time release of products, manage variability in real time and limit the number of rejections. The framework relies heavily on the design of experiments and multivariate analysis in order to understand highly complex processes (18,19). Near infrared (NIR) spectroscopy is a highly popular analytical method when incorporating the PAT initiative (20–23). The reasons for the success of NIR spectroscopy include no sample preparation, fast analysis time, and non-invasive and non-destructive sample analysis (24–26). NIR spectra originate from overtones and combinations bands of mid-IR vibrations which are analysed with FTIR spectroscopy.

The aim of this study is to evaluate the residual phenol content in spray-dried and freeze-dried insulin formulations by FTIR and NIR spectroscopy using multivariate data analysis. The establishment of a model that can be used for samples produced by different drying methods is desired. Moreover, online analysis of phenol content during dehydration would provide fast and real time monitoring of an important CQA throughout the dehydration process. The control method used is based on a reverse phase liquid chromatography method where the samples are reconstituted in order to measure the phenol content (12). The two spectroscopic methods evaluated, FTIR and NIR, yield similar type of information, but they have different advantages and drawbacks. FTIR spectroscopy is more sensitive, but the sample preparation is cumbersome and in-line measurements are not straight forward (24). On the other hand, NIR spectroscopy is fast, requires no sample preparation and can be performed online. While the method is less sensitive regarding chemical information, it is more sensitive to scattering from particles (25–27). Both spectroscopic methods are combined with multivariate analysis including principal component analysis (PCA) and partial least square (PLS) projections to latent structures. PCA is used to get an overview of the highly complex samples and PLS is used to correlate the spectra with the amount of phenol in the dried samples obtained from the reference method.

MATERIALS AND METHODS

Materials

Biosynthetic human insulin containing two Zn atoms per hexamer was kindly supplied by Novo Nordisk A/S, Denmark. All other chemicals were commercially available chemicals of analytical grade. Deionized water was filtered using a Millipore system (Millipore, Billerica, MA, USA) and used for all samples.

Sample Preparation

Insulin solutions were prepared by dissolving the insulin in a minimal volume of 0.2 M ice-cold hydrochloric acid resulting in a pH of approximately 2.5 which is below the isoelectric point of the insulin monomer (5.3) and hexamer (6.4) (28). After dissolution of the insulin, the pH was adjusted to 7.5 with 0.2 M ice-cold sodium hydroxide. The insulin concentration was determined from absorbance at 276 nm using a molar extinction coefficient of 6,200 M^{-1} cm^{-1}. The insulin concentration was adjusted to 15 mg/mL (2.5 mM) by adding phenol and deionised water resulting in samples with a phenol/insulin molar ratio from 0 to 40 (corresponding to 0–100 mM phenol). Ten samples were prepared with a phenol concentration between 0 and 100 mM phenol. If necessary, the pH was adjusted to pH 7.5 after addition of phenol. The amount of phenol is given as the molar ratio between phenol and insulin monomer.

Spray Drying

Spray drying was performed with a Büchi B-290 spray dryer (Büchi Labortechnik AG, Postfach, Switzerland). The humidity of the inlet drying air was controlled and kept at a relative humidity of 20%. A two-fluid nozzle design with nitrogen as atomising gas was used in a co-current mode. Process parameters used were nozzle gas flow rate 11.1 L/min (40 mm), feed flow rate 3.5 mL/min (10%) and aspirator capacity 100%. Three drying air temperatures were used (100°C, 150°C and 200°C) resulting in outlet air temperatures of 40°C, 70°C and 100°C. Spray-dried particles were separated from the drying air by a standard cyclone. The spray-dried powders were stored in closed vials placed in a desiccator over silica gel at 5°C. A total of 30 samples were spray-dried, ten samples for each drying air temperature. All samples were analysed separately with RP-UPLC, FTIR and NIR.

Freeze Drying

Freeze drying was performed with a Freeze Dryer Epsilon 2-4 LSC (Martin Christ GmbH, Osterode am Hartz, Germany). Three-millilitre samples in 10-mL vials were equilibrated to 5°C prior to freeze drying. The samples were placed on a shelf in the freeze dryer and were frozen by cooling them at 1°C/min down to −40°C and held at −40°C for 2 h. Primary drying was performed at 1 mbar and −40°C for 24 h. A secondary drying was performed at 0.01 mbar and −10°C for 15 h followed by an increase in temperature to 10°C. Finally, the pressure was kept at 0.01 mbar and 10°C for 1 h. The freeze-dried powders were stored in closed vials placed in a desiccator over silica gel at 5°C. A total of ten samples were freeze-dried containing ten different phenol/insulin ratios. All samples were analysed separately with RP-UPLC, FTIR and NIR.

Reversed Phase Ultra Performance Liquid Chromatography

The reference method used to analyse the phenol concentration of dehydrated samples was RP-UPLC on a Waters Acquity system (Waters Corporation, Milford, MA, USA) (12). The spray-dried or freeze-dried samples were reconstituted in water to a total solid content of approximately 2.5 mg/mL and 2 μL was injected on the C18 column (Acquity UPLC BEH C18, 1.7 μm, 2.1×50 mm) kept at 30°C. The mobile phases were A: sodium sulphate (200 mM), sodium phosphate (40 mM), acetonitrile (10%, *v*/*v*) (pH 7.2) and B: acetonitrile in water (65.5%, *v*/*v*). An isocratic elution with 75% (*v*/*v*) of eluent A in 2 min was followed by a linear gradient from 75% (*v*/*v*) eluent A to 80% (*v*/*v*) eluent B in 1.5 min. A linear gradient returned the column to 75% (*v*/*v*) eluent A in 0.5 min. The total run time per sample was 5 min at a flow rate of 500 μL/min. The samples were kept at 4°C. A detection was performed at 276 nm. The method yielded linear responses at least up to 1 mM phenol and 0.6 mM insulin. These concentrations were used as standards, and six injections were performed for each standard. The coefficient of variation was below 2% for both phenol and insulin in the tested range.

Fourier Transform Infrared Spectroscopy

The FTIR spectra of the dried samples were collected on a Bomem IR spectrometer (Bomem, Quebec, Canada) in single beam transmittance mode. A KBr disc was prepared for each sample with approximately 1 mg protein and a total disc weight of approximately 300 mg. The instrument and sample chamber were continually purged with dry air to remove water vapour. Each spectrum consisted of 256 scans collected in a single beam with a resolution of 4 cm^{-1} at ambient temperature. A total of 30 spray-dried samples and ten freeze-dried samples were analysed with FTIR. No internal standard is used for quantification, since the FTIR spectra are correlated with the relative phenol/insulin ratio.

Near Infrared Spectroscopy

The NIR spectra of the dried samples were collected on a Bomem FTLA 2000 series FT-NIR spectrometer (Bomem, Quebec, Canada) in reflectance mode. The NIR spectra were recorded through the bottom of the glass vials for each dried sample. The samples were rotated and measured again to decrease possible effects of uneven powder distribution. The samples were measured in the range of 8,000 to 4,000 cm^{-1} with a resolution of 8 cm^{-1} at ambient temperature. Each spectrum consisted of 32 scans.

Multivariate Data Analysis

PCA and PLS were performed using a Simca-P v. 11.5 (Umetrics AB, Umeaa, Sweden). In PCA, the data set consists of *J* variables and *I* objects and is arranged in one matrix (*I*×*J*). The PCA projects objects and variables to low dimensional spaces, and it can be used to find patterns, relations and differences between objects and/or variables. Principal components (PC) are computed vectors which span the largest variation in the data set, and form the axes in the new, reduced coordinate system.

In PLS, the data set consists of two matrices, the **X** matrix consisting of *J* variables and *I* objects (*I*×*J*), and the **Y** matrix consisting of *I* objects and *M* reference values (*I*×*M*). PLS uses components that describe **X** to predict **Y**. Thus, PLS models can be used to correlate the spectral data sets with the amount of phenol measured by the reference method. Spectral pre-processing is often applied prior to multivariate data analysis to reduce variation caused by unwanted variation and noise. Derivates are often applied to improve resolution of overlapping bands (25). In the PCA analysis of the FTIR data, the spectra were pre-processed by taking the second derivatives using the Savitzky–Golay algorithm. Pre-processing using second derivatives is performed to reduce noise and enhance spectral features. However, to avoid spectral distortions, it is important to limit the smoothing. Therefore a different number of points in the algorithm were tested, but 11 points yielded the best results based on visual inspection of noise and spectral features. In the PCA analysis of the NIR spectra, a standard normal variate (SNV) transformation was used to reduce the variation caused by the physical characteristics (29). Based on previous experience, the PLS regression was performed on untreated spectra, standard normal variate spectra and second derivate spectra (30). In all cases, the spectra were centred prior to analysis regardless of the pre-treatment method, while no other scaling (such as univariate scaling) was applied. A cross validation was performed on all PLS models. All multivariate models included both freeze-dried and spray-dried samples in order to investigate both methods simultaneously. For comparisons root mean square error of cross validation (RMSECV) was used together with the goodness of fit (R^2) and predictability (Q^2).

RESULTS AND DISCUSSION

As reference, the amount of phenol in each freeze-dried and spray-dried sample was analysed by RP-UPLC. Phenol elutes earlier than insulin (0.7 min *versus* 2.7 min) due to its more hydrophilic nature and baseline separation is obtained for the elution peaks. The concentration injected is correlated with the elution peak area for both phenol and insulin in the tested concentration range. The residual amount of phenol in the dried samples depends strongly on the drying method. The freeze-dried samples contain significantly more residual phenol after drying compared with a similar spray-dried sample (Fig. 1). At initial phenol/insulin ratios below 10, the amount of retained phenol is above 90% in freeze-dried samples and below 40% in spray-dried samples. Furthermore, the residual amount of phenol for both freeze-dried and spray-dried samples is not linearly correlated with the initial amount of phenol prior to drying. At higher initial concentrations, an increase in phenol concentration only slightly increases the residual amount of phenol after drying, and the relative retention of phenol decreases significantly for both drying methods (Fig. 1). The retention of phenol is unaffected by the drying temperature in the spray drier (data not shown) indicating that film formation at the surface of the droplet is independent of the temperature in the tested temperature range (2). Previous studies have showed that for insulin

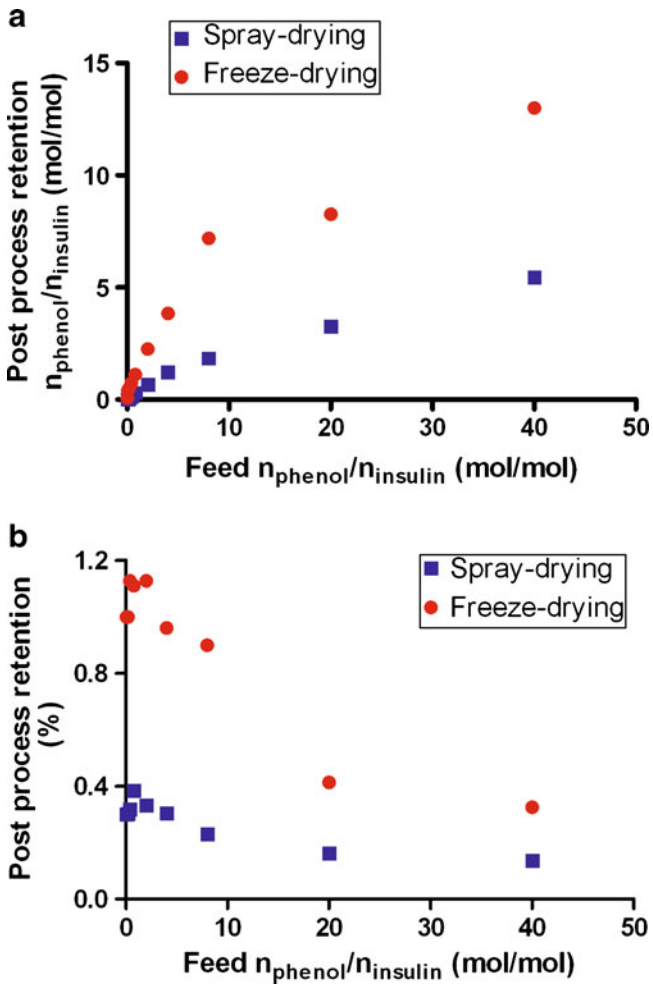

Fig. 1. The phenol/insulin ratio after freeze drying (*circles*) and spray drying (*squares*) plotted as a function of the phenol/insulin ratio before drying (**a**). The phenol/insulin ratio is determined by RP-UPLC as mole per mole. The percentage of the phenol/insulin ratio retained after freeze drying (*circles*) and spray drying (*squares*) as a function of the initial phenol/insulin ratio (**b**)

concentrations below 15 mg/mL film formation is slow resulting in highly wrinkled particles (31). This result in a lower retention of phenol in the spray-dried samples compared with samples where the film is formed faster. For freeze-dried samples, film formation is not critical, and higher amounts of phenol are retained compared with spray-dried samples.

Fourier Transformed Infrared Spectroscopy

In order to establish characteristic spectral bands of solid phenol in the mid-infrared range, an amorphous precipitate of phenol on a KBr pellet was prepared, and an infrared spectrum was recorded (Fig. 2). The infrared spectrum is similar to the recorded spectrum of crystal phenol indicating a rapid crystallisation of the amorphous phenol precipitate (data not shown).

Several distinct bands are present in the infrared spectrum of phenol which can be used for quantification. The three bands at 1,593, 1,496 and 1,475 cm^{-1} are assigned to the stretching of the carbon atoms in the phenyl ring, υ(CC). The alcohol group in the phenol molecule gives rise to two quite strong bands at 1,365 and 1,224 cm^{-1}, respectively. The first band originates from the in-plane O–H bending, δ(OH), and the second band is associated with the C–O stretching, υ(CO). The two bands at 813 and 750 cm^{-1} are due to the out-of-plane bending vibrations of the C–H bond in the phenyl ring, π(CH). The band at 690 cm^{-1} originates from the out-of-plane bend of the carbon atoms in the phenyl ring, π(CC), and is quite strong in the spectrum (32,33).

The mid-infrared spectrum of freeze-dried insulin without phenol is given in Fig. 2. The spectrum is dominated by the amide I, amide II and amide III bands. The amide I band from 1,700 to 1,600 cm^{-1} has been assigned to the stretching of carbonyl groups in the protein backbone, υ(C=O) and depends on the bond angles and hydrogen bonding of the carbonyl groups, correlating with the secondary structure of the protein. The amide II band (1,580–1,500 cm^{-1}) is more complex and originates from the combination of δ(NH) and υ(CN) modes. The broad band from 1,350 cm^{-1} to 1,200 cm^{-1} forms the amide III band and is associated with the combination of υ(CN) and δ(NH), but vibrations of the amino acid side chains in the proteins also contribute significantly to the band (34). A comparison of the phenol and dried insulin spectra indicates that the region below 1,100 cm^{-1} is best suited

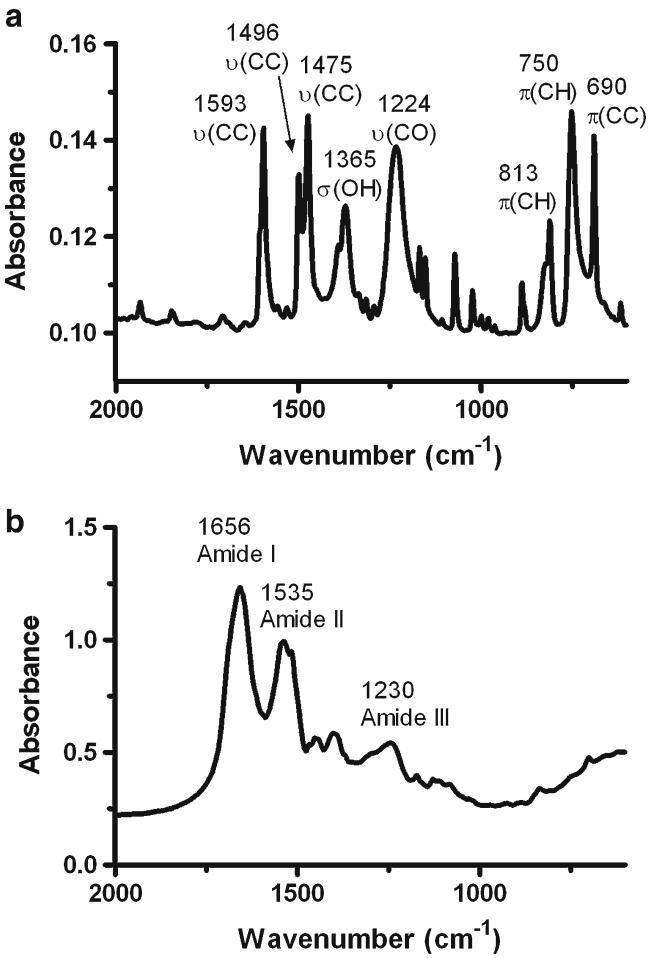

Fig. 2. FTIR spectra of dried phenol (**a**) and dried insulin (**b**). Band assignments are depicted in the figure

for phenol quantification in dried insulin formulations, as this region contains strong bands originating from phenol and contains no significant insulin-related bands.

Multivariate Analyses of FTIR Spectra

To further investigate the ability to differentiate insulin powders dried with different amounts of phenol, a PCA was performed on the second derivative spectra in the range of 2,000 cm^{-1} to 600 cm^{-1} (Fig. 3). In Fig. 3, the samples are clustered according to the phenol/insulin ratio. Three principal components explain 98.7% of the variation in the spectral data set. The first principal component (52.2%) explains part of the phenol/insulin ratio in the samples as seen by peaks at 813, 755 and 695 cm^{-1} in the loading plot corresponding to the π(CC) and π(CH) vibrations of phenol (Fig. 3). In addition, the α-helix content in the samples is depicted in the first principal component with a maximum at 1,656 cm^{-1}, and the α-helix content increased with a decreasing score (Fig. 3). The second derivative transformation of the spectra reverses the sign of the spectra and a peak in the loading corresponds to a decrease in the original spectra. Part of the increase in α-helix content is correlated with the phenol/insulin ratio as an increase in ratio results in an increase in the α-helix content of the insulin hexamer (12). However, some of the variations cannot be explained by the insulin/phenol ratio and is credited to unknown variations in the dried powders. The second principal component (43.1%) explains more of the variation in the phenol/insulin ratio as seen from the loading plot which was dominated by the bands observed in the phenol spectrum. The peaks observed in the loading plot at 1,593, 1,498, 1,473, 813, 755 and 695 cm^{-1} corresponding to υ(CC), π(CH) and π(CC) vibrations in the phenol spectrum (Fig. 3). Only minor variations are observed for the insulin spectrum in the second principal component. The third principal component (3.4%) shows little variation in the phenol/insulin ratio, but a significant variation in the insulin spectrum (data not shown).

The influence of the phenol/insulin ratio on the FTIR spectra was shown via PCA, and subsequently the correlation between RP-UPLC and FTIR spectra was investigated using PLS. An overview of the results from several models is given in Table I. For the untreated spectra, five PLS components were needed, all showing features from phenol and insulin spectra. The first PLS component of the model explains 91.8% of the variation in the spectra, but explains only 7.8% of the phenol/insulin ratio. The second and third PLS components account for 5.7% and 1.6% of the variation in the spectra, and 47.1% and 34.0% of the variation in the phenol/insulin ratio. The fourth and fifth PLS components use only 0.7% and 0.1% of the variation in the spectra to explain 5.0% and 4.0% of the variation in the phenol/insulin ratio (data not shown). Thus, most of the variation in the phenol/insulin ratio in the samples is described by a small amount of variation in the spectra. By reducing the spectral range of the data set, better models were obtained where more of the variation in the spectra was used to describe the majority of the variation in the phenol/insulin ratio. For the spectral range of 850–650 cm^{-1}, which contains the strong bands for π(CH) and π(CC) vibrations, the RMSECV decreases to a

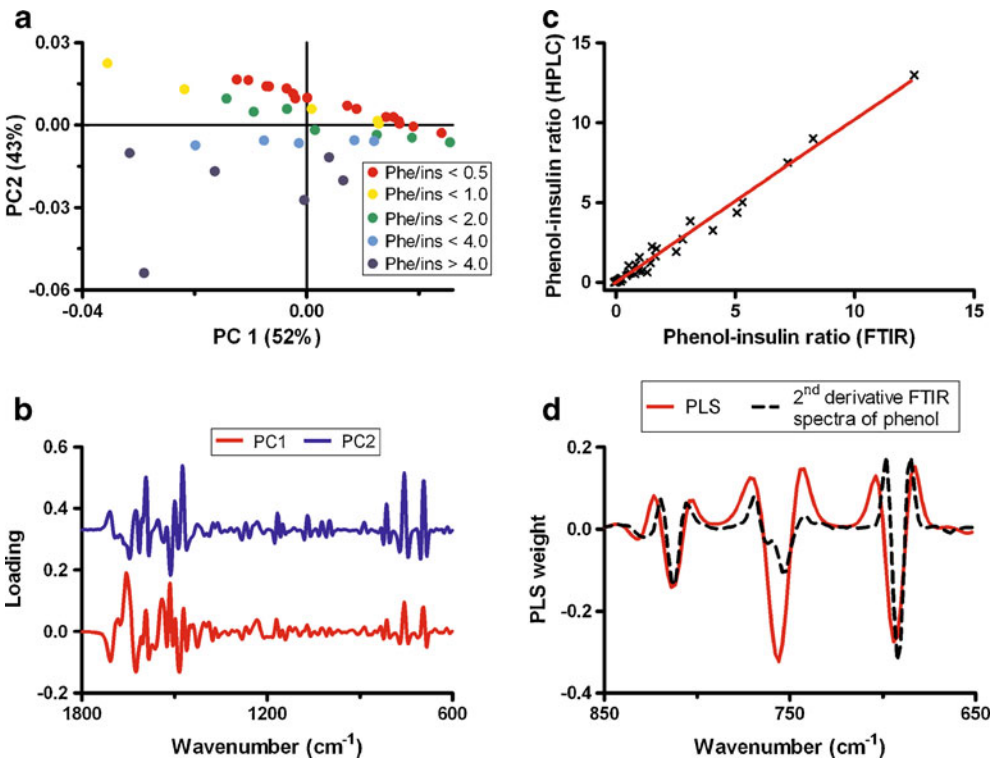

Fig. 3. Multivariate analysis of second derivative FTIR spectra. Score plot of PC1 against PC2 from PCA, coloured according to phenol/insulin ratio (**a**). Loading plots of PC1 and PC2 from PCA (**b**). Correlation between phenol/insulin ratio obtained by the PLS model based on the spectral range 850–650 cm^{-1} and measured by RP-UPLC (**c**). Loading plot of the PLS component and the second derivative FTIR spectra of dried phenol (**d**)

Table I. PLS Model Overview for the Correlation Between UPLC and IR in the Determination of Phenol/Insulin Ratio Utilising Various Pre-Treatments and Wavenumber Ranges

Spectroscopic method	Pre-treatment	Wavenumber range (cm^{-1})	Number of PCs	RMSECV (ratio)
FTIR	Untreated	4,000–600[a]	5	0.49
FTIR	Untreated	1,500–600[b]	2	0.57
FTIR	Untreated	850–650[c]	2	0.48
FTIR	Untreated	1,300–1,000[d]	2	1.06
FTIR	SNV	4,000–600[a]	4	0.77
FTIR	SNV	1,500–600[b]	2	0.80
FTIR	SNV	850–650[c]	2	0.87
FTIR	SNV	1,300–1,000[d]	2	0.86
FTIR	Second derivative	4,000–600[a]	1	0.77
FTIR	Second derivative	1,500–600[b]	1	0.55
FTIR	Second derivative	850–650[c]	1	0.43
FTIR	Second derivative	1,300–1,000[d]	1	0.55
NIR	SNV	7,400–4,000	4	0.59
NIR	SNV	6,200–4,500	4	0.63
NIR	SNV	6,200–5,800	3	0.37

[a] The spectral range includes the π(CC), π(CH), υ(CO), δ(OH) and υ(CC) vibrations of the phenol molecule
[b] The spectral range includes the π(CC), π(CH), υ(CO), δ(OH) and υ(CC) vibrations of the phenol molecule
[c] The spectral range includes the π(CC) and π(CH) vibrations of the phenol molecule
[d] The spectral range includes the υ(CO) and δ(OH) vibrations of the phenol molecule

phenol/insulin ratio of 0.48, based on only two PLS components. The first component still describes the main variation in the spectra (95.6%) and only a small part of the variation in the phenol/insulin ratio (18.0%), whereas the second PLS component primarily describes the variation in the phenol/insulin ratio (80.7%) based on 4.1% of the spectral variation.

For all models based on the untreated data, the first PLS component only describes a small part of the variation in the phenol/insulin ratio and a large part of the variation in the spectra corresponding to the large variation in baseline offset. To overcome this variation, different transformations of the spectral data set were evaluated. A SNV transformation has been used in NIR spectroscopy to reduce the variation in the baseline offset (25,30). However, for the FTIR spectra, the obtained models contain the same number of principal components and did not explain more of the data compared with models based on raw spectral data. The best models were obtained with the second derivative spectra where only one PLS component is sufficient to describe both the variation in the spectra and the phenol/insulin ratio, yielding the lowest RMSECV value (0.43). Both R^2 (0.977) and Q^2 (0.964) show a high degree of correlation between the actual phenol/insulin ratio measured by UPLC and the phenol/insulin ratio predicted from the PLS model (Fig. 3). The PLS component is dominated by the π(CH) and π(CC) bands at 813, 750 and 690 cm^{-1} as observed in the phenol spectrum (Fig. 3). PLS models based on only freeze-dried or spray-dried samples were not superior to the models based on all dehydrated samples.

Near Infrared Spectroscopy

NIR spectra are sensitive to baseline offsets, and the NIR spectra in Fig. 4 were therefore SNV-corrected prior to analysis. An NIR spectrum of crystal phenol was used as reference since amorphous phenol was found to recrystallise rapidly. The NIR spectrum of phenol is characterised by fewer and broader bands, which have a lower intensity compared with the FTIR spectrum of phenol (Fig. 4). This is expected due to the lower molar absorptivity of NIR bands which originate from overtones and combination bands. However, several distinct bands are present in the NIR spectrum of phenol which can be used for quantification. The spectral region 6,600–5,600 cm^{-1} is assigned to overtones, and a distinct band is found in this region at 5,995–5,955 cm^{-1}, corresponding to the first overtone of C–H stretching, 2υ(CH). The bands found in the region 5,000–4,000 cm^{-1} are all combination bands mostly assigned to C–H stretching υ(CH) + υ(CH). For phenol, the bands at 4,648, 4,550, 4,300 and 4,046 cm^{-1} are C–H combination bands (35,36). No apparent O–H stretching band is observed in the solid phenol samples.

The NIR spectrum of freeze-dried insulin without phenol is given in Fig. 4. The bands in the region 5,900–5,700 cm^{-1} are, like the phenol spectrum, the first overtones of the C–H stretching. The spectrum is dominated by combination bands in the region 5,200–4,000 cm^{-1}, similar to bands found for other proteins (37). Most of these bands have been tentatively assigned to combinations of amide I, amide II, amide III, amide A, amide B and C–H stretching bands (38,39). The band at 5,150 cm^{-1} originates from the combination band of water and can be used for the quantification of water in the dried samples (40,41). Comparing the spectra of dried phenol and dried insulin, several bands of insulin and phenol are overlapping, especially in the area below 5,000 cm^{-1} which usually results in the strongest absorption bands. However, the bands at 6,105–6,045 cm^{-1} and 5,995–5,955 cm^{-1} in the phenol spectrum are located in a spectral region that is rather featureless in the insulin spectrum. Therefore, these may be suitable for the quantification purposes of the phenol/insulin ratio in the samples.

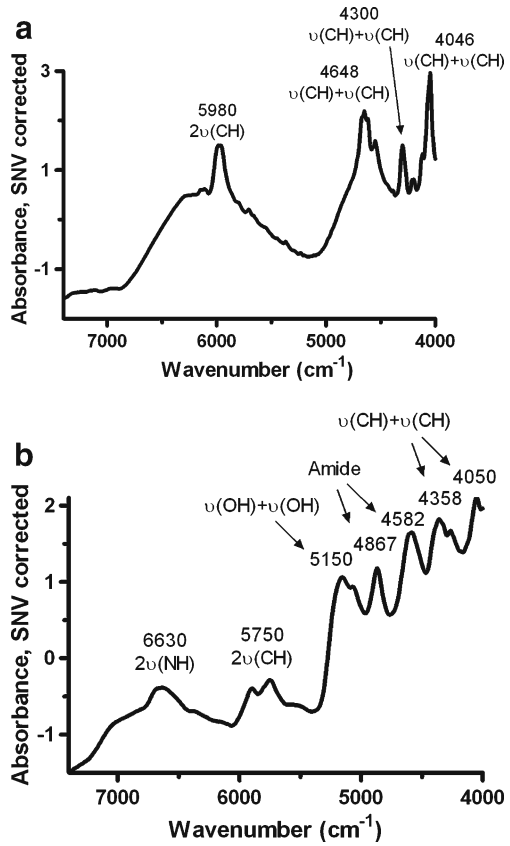

Fig. 4. SNV-corrected NIR spectra of dried phenol (**a**) and dried insulin (**b**). Band assignments are depicted in the figure

Multivariate Analyses of NIR Spectra

PCA was performed on the SNV-corrected NIR spectra in order to further evaluate the possibility to differentiate insulin powders dried with different amounts of phenol (Fig. 5). Three PCs were used to describe 95.4% of the variation in the NIR spectra. The first principal component describes 73.5% of the variation in the data set, and the loading of PC1 is dominated by a main peak at 5,150 cm^{-1} and a second broader peak at 6,900 cm^{-1} (Fig. 5). Both peaks originate from the absorption bands of water, where the first corresponds to the combination band and the second to the first overtone of water. Only minor contributions from the phenol spectrum are observed. Thus, the separation found along PC1 is dominated by the difference in the water content of the samples. The main feature of PC1 is the separation between the freeze-dried and spray-dried samples (Fig. 5). In general, the freeze-dried samples have a negative score on PC1. In contrast, most spray-dried samples have a positive score on PC1, correlating with a positive peak in the loading plot for the two water absorption bands. Thus, the spray-dried samples contain more residual water than the freeze-dried samples. The second principal component describes 15.2% of the total variation in the spectral data set. The loading plot indicates that the variation is mainly due to the variation in the spectral region 5,000–4,000 cm^{-1} corresponding to the combination bands (Fig. 5). Most of the combination bands can be credited to the insulin spectra, and only minor contributions from phenol are observed. In summary, most of the variation is found in the spectral region 5,000–4,000 cm^{-1}, where differentiation between phenol and insulin is difficult. The water content is inversely correlated with phenol/insulin ratio as seen from opposing peaks in the loading plots (Fig. 5).

PLS models were constructed by correlating the SNV-corrected spectra with the phenol/insulin ratio measured by RP-UPLC. An overview of the obtained models is given in Table I. The first model is based on the spectral range 7,400–4,000 cm^{-1} and yields a RMSECV of 0.59 with four PLS components. The first PLS component describes 71.7% of the variation in the spectra and 41.8% of the variation in the phenol/insulin ratio. The next three PLS components describe 15.3%, 7.1% and 2.9% of the variation in the spectra and 28.7%, 16.1% and 11.8% of the variation in the phenol/insulin ratio (data not shown). All the PLS components show distinct peaks originating from the phenol spectrum. In addition, the first PLS component contains information about the water content. In order to obtain simpler and more robust models, the spectral ranges 6,200–4,500 cm^{-1} and 6,200–5,800 cm^{-1} were tested. The spectral range 6,200–4,500 cm^{-1} contains the most prominent bands in the phenol spectrum as well as the water band, whereas 6,200–5,800 cm^{-1} contains the broad phenol band between 5,995 and 5,955 cm^{-1}. The spectral range 6,200–4,500 cm^{-1} results in a model with a higher RMSECV (0.63) using four PLS components, and the model is therefore inferior to the first model. However, in the range 6,200–5,800 cm^{-1}, only three PLS components are needed to achieve an RMSECV of 0.37, resulting in a better model compared with the model based on the whole spectral range (Fig. 5). The first PLS component corresponds to the large absorption band in the phenol spectrum (5,995–5,955 cm^{-1}). It explains 93.8% of the spectral variation and 96.3% of the variation in the phenol/insulin ratio. The next two PLS components describe only minor variations in the spectral data and phenol/insulin ratio (Fig. 5). The third component shows a band between 6,140 and 6,040 cm^{-1}, corresponding to the small band between 6,145 and 6,100 cm^{-1} that can be seen in the phenol spectrum. It is therefore reasonable to include the third component into the model, while further components would lead to over-fitting of the model.

Comparison of Methods

PLS models based on the FTIR and NIR spectra are both able to predict the phenol/insulin ratio. The best FTIR PLS model is based on the second derivative spectra in the spectral range of 850–650 cm^{-1} (Tables I and II). The best NIR PLS model is constructed from the SNV-corrected spectra in the range of 6,200–5,800 cm^{-1}. The R^2 and Q^2 are above 0.96 for all models (Table II), indicating a strong predictive strength. The biggest difference between the basic models and the optimised models is the number of PLS components. For models containing a higher number of PLS components the first PLS components did not explain the variation in the phenol/insulin ratio. For models based on large spectral regions, the first PLS components are not related to the phenol/insulin ratio but to some other variation in the samples making the models less robust. This variation might be important in the overall understanding of the

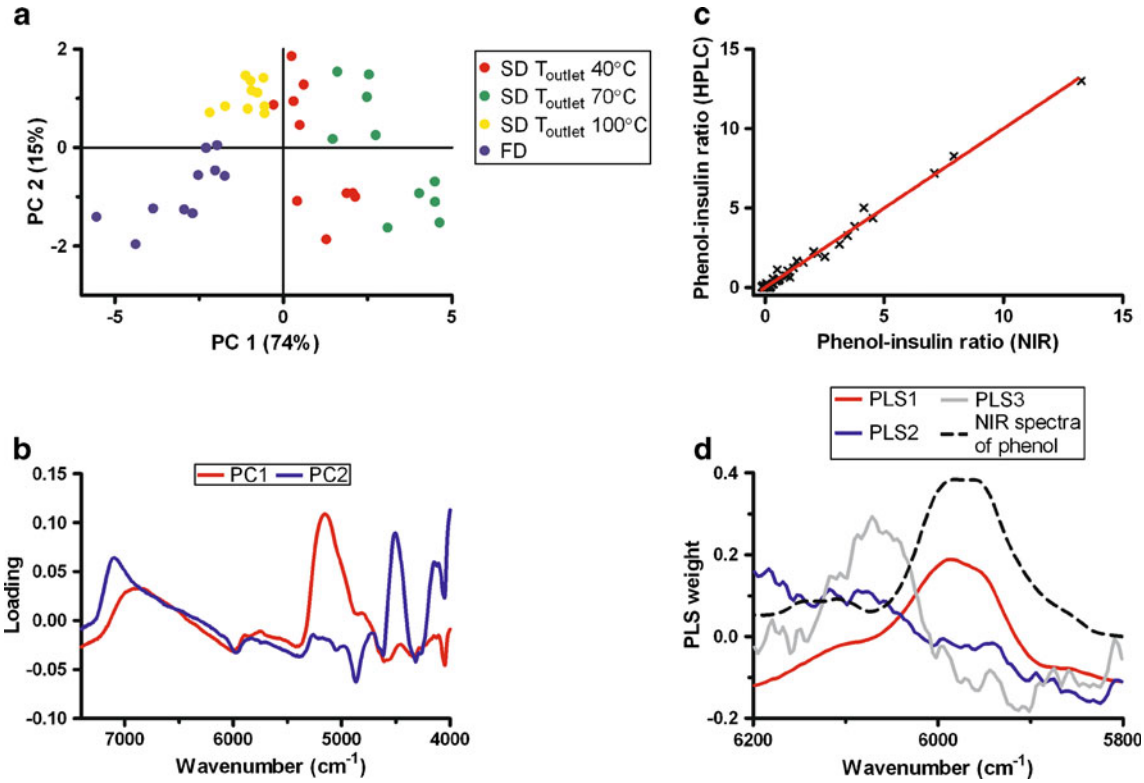

Fig. 5. Multivariate analysis of SNV-corrected NIR spectra. Score plot of PC1 against PC2 from PCA, coloured according to drying method (**a**). Loading plots of PC1 and PC2 from PCA (**b**). Correlation between phenol/insulin ratio obtained by the PLS model based on the spectral range 6,200–5,800 cm^{-1} and measured by RP-UPLC (**c**). Loading plots of the three PLS components and the SNV-corrected NIR spectra of dried phenol (**d**)

sample set and in the identification of CPPs, but not for the quantification of the phenol/insulin ratio. The RMSECV values are similar for all the models. The best model based on the FTIR spectra yielded a RMSECV of 0.43 phenol/insulin ratio, while the best model based on NIR spectra yielded a RMSECV of 0.37 phenol/insulin ratio. These molar ratios correspond to mass percentages of 0.69% and 0.60% (*w*/*w*), respectively.

In general, both methods are suitable for phenol quantification in dried phenol/insulin solids, as both FTIR and NIR spectroscopy yield similar chemical information about the sample. The spectra of the solid samples dried by freeze drying and spray drying, which are sufficiently similar to be incorporated in the same calibration model. Thus, when comparing the two methods, additional advantages and disadvantages need to be considered for the methods. FTIR spectroscopy has the advantage of a higher degree of chemical information about the sample, which includes information on the secondary structure of the protein. However, if the purpose for the measurements is quantification of the phenol/insulin ratio, this information is not necessary and FTIR spectroscopy should not be chosen based on this. In addition, FTIR experiments require more elaborate sample preparation, and in the case of KBr pellets, the preparation is destructive. A sample preparation can be avoided using attenuated total reflectance FTIR spectroscopy, but this method is highly dependent on the homogeneity of the sample due to the short penetration of the light in the sample. On the other hand, NIR spectroscopy has the advantages of extremely fast measuring time combined with no sample preparation. Furthermore, it can be conducted online, which is important for the PAT initiative. While the structural information obtained is not as good as with FTIR spectroscopy, studies have correlated changes in the NIR

Table II. Comparison of Cross-Validated PLS Models Based on FTIR and NIR

Model	1	2	3	4
Spectroscopic method	FTIR	FTIR	NIR	NIR
Pre-treatment	Untreated	Second derivative	SNV	SNV
Wavenumber (cm^{-1})	4,000–600	850–650	7,400–4,000	6,200–5,800
RMSECV (ratio)	0.49	0.43	0.59	0.37
R^2	0.979	0.977	0.984	0.990
Q^2	0.973	0.964	0.975	0.980

spectra with changes in the amide I band in the FTIR region (38). In addition, NIR spectroscopy can yield information about particle size and water content of the powder (41,42).

CONCLUSION

After dehydration of the insulin samples containing phenol, the residual amount of phenol in the samples was analysed by RP-UPLC. The FTIR and NIR spectra were correlated with the phenol/insulin ratio using a PLS analysis. There was a difference in the residual phenol/insulin ratio between the two drying methods applied. In addition, the spray-dried samples contained more water and less phenol, whereas the comparable freeze-dried sample contained less water and more phenol. The best FTIR model contained one PLS component and was based on the second derivative spectra in the 850–650 cm^{-1} region. The best NIR model contained three PLS components and was based on the SNV pre-treated spectra in the 6,200–5,800 cm^{-1} region. The best FTIR model yielded an RMSECV of 0.43% or 0.69% (*w*/*w*), and the NIR model yielded an RMSECV of 0.37% or 0.60% (*w*/*w*). In conclusion, both methods are suitable for the quantification of the phenol in the dried powders.

ACKNOWLEDGEMENTS

The authors would like to acknowledge funding by Apotekerfonden af 1991 for the Bomem FTIR spectrometer

REFERENCES

1. Maltesen MJ, van de Weert M. Drying methods for protein pharmaceuticals. Drug Discov Today Technol. 2008;5(2–3):e81–8.
2. Coumans WJ, Kerkhof PJAM, Bruin S. Theoretical and practical aspects of aroma retention in spray drying and freeze drying. Drying Technol. 1994;12(1–2):99–149.
3. Dunn MF. Zinc-ligand interactions modulate assembly and stability of the insulin hexamer—a review. Biometals. 2005;18 (4):295–303.
4. Brange J, Langkjaer L. Insulin structure and stability. Pharm Biotechnol. 1993;5:315–50.
5. Baker EN, Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, *et al*. The structure of 2Zn pig insulin crystals at 1.5 A resolution. Philos Trans R Soc Lond B Biol Sci. 1988;319 (1195):369–456.
6. Chang X, Jorgensen AM, Bardrum P, Led JJ. Solution structures of the R6 human insulin hexamer. Biochemistry. 1997;36 (31):9409–22.
7. Derewenda U, Derewenda Z, Dodson EJ, Dodson GG, Reynolds CD, Smith GD, *et al*. Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. Nature. 1989;338(6216):594–6.
8. Kaarsholm NC, Ko HC, Dunn MF. Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin. Biochemistry. 1989;28(10):4427–35.
9. Roy M, Brader ML, Lee RW, Kaarsholm NC, Hansen JF, Dunn MF. Spectroscopic signatures of the T to R conformational transition in the insulin hexamer. J Biol Chem. 1989;264 (32):19081–5.
10. Smith GD, Dodson GG. The structure of a rhombohedral R6 insulin hexamer that binds phenol. Biopolymers. 1992;32(4):441–5.
11. Choi WE, Borchardt D, Kaarsholm NC, Brzovic PS, Dunn MF. Spectroscopic evidence for preexisting T- and R-state insulin hexamer conformations. Proteins. 1996;26(4):377–90.
12. Maltesen MJ, Bjerregaard S, Hovgaard L, Havelund S, van de Weert M. Analysis of insulin allostery in solution and solid state with FTIR. J Pharm Sci. 2009;98(9):3265–77.
13. Wu N, Clausen AM. Fundamental and practical aspects of ultrahigh pressure liquid chromatography for fast separations. J Sep Sci. 2007;30(8):1167–82.
14. Novakova L, Vlckova H. A review of current trends and advances in modern bio-analytical methods: chromatography and sample preparation. Anal Chim Acta. 2009;656(1–2):8–35.
15. International Conference on Harmonisation (ICH). Quality guidelines: Q8-10. Geneva, Switzerland, ICH. 2005. http://www.ich.org/home.html. Accessed 3 May2011.
16. European Medicines Agency (EMA). Note for guidance on pharmaceutical development (EMEA/CHMP/167068/2004), London, United Kingdom, EMA. 2006. http://www.ema.europa.eu. Accessed 3 May 2011.
17. Food and Drug Administration (FDA). Guidance for industry. PAT—a framework for innovative pharmaceutical development, manufacturing and quality assurance. Rockville, United States, FDA. 2004. http://www.fda.gov. Accessed 3 May 2011.
18. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. Nat Biotechnol. 2009;27(1):26–34.
19. Rathore AS. Roadmap for implementation of quality by design (QbD) for biotechnology products. Trends Biotechnol. 2009;27 (9):546–53.
20. Christensen D, Alleso M, Rosenkrands I, Rantanen J, Foged C, Agger EM, *et al*. NIR transmission spectroscopy for rapid determination of lipid and lyoprotector content in liposomal vaccine adjuvant system CAF01. Eur J Pharm Biopharm. 2008;70(3):914–20.
21. De Beer TR, Vercruysse P, Burggraeve A, Quinten T, Ouyang J, Zhang X, *et al*. In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. J Pharm Sci. 2009;98(9):3430–46.
22. Wu H, Khan MA. Quality-by-Design (QbD): an integrated process analytical technology (PAT) approach for real-time monitoring and mapping the state of a pharmaceutical coprecipitation process. J Pharm Sci. 2010;99(3):1516–34.
23. Wu H, Tawakkul M, White M, Khan MA. Quality-by-Design (QbD): an integrated multivariate approach for the component quantification in powder blends. Int J Pharm. 2009;372(1–2):39–48.
24. Teixeira AP, Oliveira R, Alves PM, Carrondo MJ. Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative. Biotechnol Adv. 2009;27(6):726–32.
25. Reich G. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. Adv Drug Deliv Rev. 2005;57(8):1109–43.
26. Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharm Biomed Anal. 2007;44(3):683–700.
27. Rasanen E, Sandler N. Near infrared spectroscopy in the development of solid dosage forms. J Pharm Pharmacol. 2007;59(2):147–59.
28. Kaarsholm NC, Havelund S, Hougaard P. Ionization behavior of native and mutant insulins: pK perturbation of B13-Glu in aggregated species. Arch Biochem Biophys. 1990;283(2):496–502.
29. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trading of near-infrared diffuse reflectance spectra. Appl Spectrocopy. 1989;43(5):772–7.
30. Grohganz H, Fonteyne M, Skibsted E, Falck T, Palmqvist B, Rantanen J. Role of excipients in the quantification of water in lyophilised mixtures using NIR spectroscopy. J Pharm Biomed Anal. 2009;49(4):901–7.
31. Maltesen MJ, Bjerregaard S, Hovgaard L, Havelund S, van de Weert M. Quality by design—spray drying of insulin intended for inhalation. Eur J Pharm Biopharm. 2008;70(3):828–38.
32. Jakobsen RJ, Brasch JW. Far infrared studies of hydrogen bond of phenols. Spectrochim Acta. 1965;21(10):1753–63.
33. Kubinyi M, Billes F, Grofcsik A, Keresztury G. Vibrational-spectra and normal coordinate analysis of phenol and hydroquinone. J Mol Struct. 1992;266:339–44.
34. Barth A. Infrared spectroscopy of proteins. Biochim Biophys Acta. 2007;1767(9):1073–101.
35. Ishiuchi S, Fujii M, Robinson TW, Miller BJ, Kjaergaard HG. Vibrational overtone spectroscopy of phenol and its deuterated isotopomers. J Phys Chem A. 2006;110(23):7345–54.

36. Rospenk M, Czarnik-Matusewicz B, Zeegers-Huyskens T. Near infrared spectra (4000–10 500 cm(−1)) of phenol-OH and phenol-OD in carbon tetrachloride. Spectrochim Acta A Mol Biomol Spectrosc. 2001;57(1):185–95.
37. Izutsu KI, Fujimaki Y, Kuwabara A, Hiyama Y, Yomota C, Aoyagi N. Near-infrared analysis of protein secondary structure in aqueous solutions and freeze-dried solids. J Pharm Sci. 2006;95(4):781–9.
38. Bai SJ, Nayar R, Carpenter JF, Manning MC. Noninvasive determination of protein conformation in the solid state using near infrared (NIR) spectroscopy. J Pharm Sci. 2005;94(9):2030–8.
39. Bruun SW, Sondergaard I, Jacobsen S. Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. Gluten powder. J Agric Food Chem. 2007;55(18):7234–43.
40. Cao W, Mao C, Chen W, Lin H, Krishnan S, Cauchon N. Differentiation and quantitative determination of surface and hydrate water in lyophilized mannitol using NIR spectroscopy. J Pharm Sci. 2006;95(9):2077–86.
41. Zhou GX, Ge Z, Dorwart J, Izzo B, Kukura J, Bicker G, *et al.* Determination and differentiation of surface and bound water in drug substances by near infrared spectroscopy. J Pharm Sci. 2003;92(5):1058–65.
42. Nieuwmeyer FJ, Damen M, Gerich A, Rusmini F, van der Voort Maarschalk K, Vromans H. Granule characterization during fluid bed drying by development of a near infrared method to determine water content and median granule size. Pharm Res. 2007;24(10):1854–61.