Quantifying crystalline α-lactose monohydrate in amorphous lactose using terahertz time domain spectroscopy and near infrared spectroscopy

Warnecke, Solveig; Wu, Jian X.; Rinnan, Åsmund; Allesø, Morten; van den Berg, Frans W.J.; Jepsen, Peter Uhd; Engelsen, Søren Balling

Published in:
Vibrational Spectroscopy

Link to article, DOI:
10.1016/j.vibspec.2019.03.004

Publication date:
2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Quantifying crystalline α-lactose monohydrate in amorphous lactose using Terahertz Time Domain Spectroscopy and Near Infrared Spectroscopy

Solveig Warnecke, Jian X. Wu, Åsmund Rinnan*, Morten Allesø, Frans van den Berg, Peter Uld Jepsen, Søren Balling Engelsen

aDepartment of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
bDepartment of Pharmacy, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 København Ø, Denmark
cBiologics and Pharmaceutical Science, H. Lundbeck A/S, Ottiliavej 9, DK-2500 Valby, Denmark
dDTU Fotonik - Department of Photonics Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

*) corresponding author, aar@food.ku.dk, +45 35333542.

Abstract

Spray-dried lactose consists of an amorphous component (10-20%) as well as the crystalline monohydrate form [1]. It is commonly used as a diluent in direct compression, mainly because of its better flow characteristics compared to pure crystalline lactose. The amorphous form is metastable and can easily crystallize, which will affect the functionality of the pharmaceutical product. It is therefore of interest to establish methods for non-invasive and rapid assessment of the level of crystallinity in a pharmaceutical formulation. In this study, two spectroscopic methods, near infrared (NIR) spectroscopy and terahertz time domain spectroscopy (THz-TDS), are compared for their ability to determine low levels of crystalline lactose in a mixture. The aim was to
find the limit of detection and limit of quantification for the two techniques. Partial least squares (PLS) regression models were developed and the root-mean-square-error-of-cross-validation (RMSECV) for the models with full concentration range were found to be 2.91 % (w/w) and 0.87 % (w/w) for THz-TDS and NIR, respectively. Calibrations developed on samples containing 0-10 % (w/w) crystalline material resulted in RMSECVs of 0.30 % (w/w) and 0.20 % (w/w) for THz-TDS and NIR, respectively, while the limits of detection were 0.80 % (w/w) and 0.43 % (w/w), respectively. Both instrumental techniques are thus able to quantify the content of crystalline lactose in a mixture. To select one method over the other in an industrial quality assurance setting, further includes other aspects - such as sample handling, sample size, outlier information, instrument stability, etc. In all these aspects, NIR spectroscopy currently performs better than THz-TDS.

Abbreviations:

API: active pharmaceutical ingredient
iPLS: interval partial least squares regression
LOD: limit of detection
LOQ: limit of quantification
LV: latent variable
NIR: near-infrared spectroscopy
PC: principle component
PCA: principal component analysis
PE: polyethylene
PLS: partial least squares regression
QbD: quality by design
RMSECV: root-mean-square-error-of-cross-validation
SEM: scanning electron microscopy
SNV: standard normal variate
THz: terahertz
Keywords: Terahertz time domain spectroscopy (THz-TDS); NIR; lactose; crystalline; amorphous; limit of detection (LOD)

1. Introduction

It has been shown that about 80% of active pharmaceutical ingredients (APIs) have polymorphs [2]. The arrangement of the molecules in amorphous or different crystalline forms can impact the dissolution rate, solubility, bioavailability and manufacturability of pharmaceuticals [3]. An amorphous form may have short-range order, but unlike crystalline forms, the amorphous state has no long-range order of molecular packing [4]. Many APIs are poorly water soluble, which has increased the interest in formulating amorphous compounds and also amorphous binary mixtures [4, 5]. The amorphous form typically has a much higher apparent solubility and enhanced dissolution, due to the fact that it does not need to be released from a firm (typically) hydrogen-bonded crystal lattice before solubilization. Presence of smaller seeds of crystalline impurities and water can initiate and propagate a phase transition process, and over longer storage times change the form of the product which, in turn, will impact the stability, solubility, or other functionalities of the pharmaceutical product [6]. This highlights the need for analytical method development in order to detect and quantify the small amount of crystalline material.

Characterization of the amorphous and crystalline states can be accomplished via several different instrumental techniques that observe changes in physical properties such as density, viscosity, or heat capacity. The current gold standard to study long range molecular order is X-ray diffraction.
In X-ray powder diffraction (XRPD) lack of long-range order will be observed as a halo compared to the sharp crystalline reflections observed for crystalline forms [7]. The limit of quantification for crystalline structure in XRPD is approximately 5 % [7, 8]. Fix and Steffens [9] developed an improved XRPD method which is able to quantify amorphous and crystalline lactose in binary mixtures with a limit of detection (LOD) at the 0.5 % (w/w) level. Variations in particle shape and sample presentation is known to affect the XRPD pattern through preferred orientation. This effect is seen as random changes in diffraction intensities which impairs its use for quantification [10], an effect that can be partially eliminated by spinning the sample. Another well-established method for characterization of amorphous powders is differential scanning calorimetry (DSC) [7], which measures the presence of a glass transition temperature (T\textsubscript{g}). Fix and Steffens [9] found that DSC did not show linearity with the crystallinity of lactose in binary mixtures of amorphous and crystalline lactose, and concluded that DSC was not capable of detecting crystalline lactose below 10 %. Other studies have demonstrated that near infrared spectroscopy (NIR) and Raman spectroscopy are able to quantitatively differentiate between amorphous and crystalline pharmaceutical ingredients [8, 11, 12, 13, 14, 15, 16].

Terahertz spectroscopy [17] has, since it is measuring the crystal lattice vibrations [18], earlier shown good results in detecting crystalline materials [19, 20, 21, 22, 23, 24]. Zeitler et al. [24] found that terahertz spectroscopy could easily detect mixtures of the different sulfathiazole polymorphs, which was not possible with mid-IR or NIR spectroscopy. The electromagnetic radiation used in terahertz and NIR spectroscopy is orders of magnitude lower in energy and intensity as compared to Raman spectroscopy, which has demonstrated heating effects of the sample during measurement [25, 26]. As the photon energy of terahertz radiation is approximately a million times
smaller than that of X-ray radiation, and a thousand times smaller than that of UV light, no photochemical effects are expected in the terahertz region at the power levels relevant for terahertz spectroscopic investigations [23]. The corresponding numbers for NIR spectroscopy are approximately 100,000 times lower than X-ray radiation and 10-100 times lower than UV light. This makes it a promising technique for fast screening and characterization of API solid forms during pharmaceutical development and potentially during commercial manufacturing as a quality control tool.

Only a few terahertz spectroscopy studies have discussed the LOD and limit of quantification (LOQ) of different polymorphic API forms in mixtures. Strachan et al. [21] investigated carbamazepine form I in 0-10 % (w/w) mixtures with form III, and found an LOD of 1.2 % (w/w) and LOQ of 3.7 % (w/w). For enalapril acetate form I in a 0-10 % mixture with form II they found an LOD of 0.69 % (w/w) and LOQ of 2.1 % (w/w). For quantification of 0-100 % (w/w) binary mixtures of crystalline and amorphous indomethacin, the same authors [21] found an LOD of 1.1 % (w/w) and LOQ of 3.2 % (w/w). For crystalline fenoprofen calcium 0-100 % (w/w) the LOD was 2.7 % (w/w) and LOQ was 8.2 % (w/w).

In terahertz transmission spectroscopic experiments, it is often desirable to keep the amount of sample material to a minimum in order to maximize the spectral bandwidth of the instrument, which, without dilution, results in a sample thickness down to 0.2 mm [27]. This can cause multiple reflections from the pellet-air interfaces, effectively causing the signal reverberation in the sample – the so-called etalon or Fabry-Perot effect. In the time domain, this is seen as echoes of the main peak [27, 28, 29], and the effect manifests itself as oscillations in the measured transmission spectra. For this reason, sample dilution is often required in order not to exceed the limits of the dynamic range of the instrument [27, 30]. High-density polyethylene (PE) fine powder is frequently
added for this purpose in concentrations ranging from 25-96 % (w/w) of the total sample amount [19, 20, 21, 31, 32]. Ultrahigh molecular weight PE is a good binder and is virtually transparent and with a nearly frequency independent refraction index of 1.53 in the terahertz region [33]. When adding PE to the sample, the dynamic range of the spectrometer can be fully exploited, while thicker samples will result in a larger separation of the time domain of etalon echoes, which are then easy to remove by standard time window procedures [27, 28]. The drawback of adding PE is the need of sample preparation, making the analytical method less applicable for in-line process monitoring. Wu et al. [32] reported that when preparing a sample disk for analysis, drying time, composition, and particle size could influence the intensities and peak positions of the terahertz transmission spectra. Additionally, the authors found that the component concentration in the disks affected the terahertz spectra. For theophylline and lactose, higher concentrations gave more intense peaks, but also a higher spectral baseline shift, which was probably due to increased scattering of the THz radiation because of the finite size of the particles. The concentration levels of the disk were 4-12 % (w/w), and the measurements were done on a modified FT-IR interferometer with a range of 50-700 cm⁻¹ (1.5-21 THz), which is slightly higher than normally used in terahertz experiments and well into the infrared region.

In the present study, lactose monohydrate (the crystalline form of α-lactose) was selected as a model compound in binary amorphous-crystalline sample mixtures. Samples were purposely not diluted with PE in order to simulate the measurement performance in a production environment. The aim of this study was to estimate LOD and LOQ for terahertz time domain spectroscopy (THz-TDS) for crystalline lactose in a compact pellet. The omnipresent PAT spectroscopic technique NIR spectroscopy [34] is used for comparison.
2. Materials and methods

2.1 Sample preparation

\(\alpha\)-Lactose monohydrate (Pharmatose 450M) was purchased from DFE pharma (Goch, Germany). Lactose monohydrate was dissolved in Milli-Q water (0.20g·ml\(^{-1}\)). The solution was spray dried using a Büchi B-290 spray dryer (Büchi Labortechnik AG, Flawil, Switzerland) using the following process parameters: atomization gas flow and drying air flow rate 667 L·h\(^{-1}\), inlet temperature 175°C, outlet temperature 85-92 °C and a feed rate of 5 ml·min\(^{-1}\). In order to minimize the recrystallization of the amorphous lactose, the samples were stored at 5 °C immediately after spray drying. Prior to spectroscopic measurements the samples were weighed and gently mixed with mortar and pestle using the geometric dilution principle. A total of 21 binary samples – 2.5 g of each - with the following % (w/w) lactose monohydrate were produced: 0, 0.5, 1, 2, 5, 7, 10, 20, 30, 40, 50, 60, 70, 80, 90, 93, 95, 97, 99, 99.5 and 100 % (w/w). The samples from 0.5-7 % (w/w) and 93-99.5 % (w/w) were produced in duplicates, while the 0% and the 100% crystalline were all measured nine times.

2.3 Terahertz spectroscopy

From each of the powder mixtures a compact pellet with 0.25 g total mass was prepared by compressing with 2 tons of force in a hydraulic press (Atlas, Specac Inc., Swedesboro, New Jersey) between two disks of Teflon (20 mm diameter). The pellet and the Teflon blocks were transferred to a sample holder and analyzed in transmission mode using the Terahertz instrument T-Ray\textsuperscript{TM} 4000 THz-TDS System (Picometrix, Ann Arbor, Michigan, USA) interfaced with the T-REX software (ver. 4.5). Each measurement was acquired as an average of 10,000 scans (total recording
time of 100 seconds) and each sample was analyzed in triplicate. Reference waveforms were obtained with the sample holder including the Teflon blocks but no sample material. THz waveforms were smoothed before Fourier transformation. Subsequently, the ratio between the sample and the reference spectra was converted into absorbance units and used for further data analysis.

2.4 NIR spectroscopy

The powder mixtures were filled into a glass vial and NIR reflectance spectra were acquired using an ABB Bomem FT-NIR MB-160 (ABB Bomem Inc., Montreal, Canada, GRAMS Light software) with 8 cm\(^{-1}\) resolution. Each spectrum was recorded as the average of 64 scans in the range 4,000-10,000 cm\(^{-1}\) (1,000-2,500 nm, 121-303 THz). The spectra were measured in stationary reflectance mode through the bottom of the glass vials. The powder mixture inside the glass vials was gently mixed between measurements. Reference spectra were recorded using a standard material (Spectralon 99%, Lapsphere Inc., North Sutton, New Hampshire) and used to convert the tablet spectra into absorbance units.

2.5 X-Ray powder diffraction

X-ray powder diffraction measurements were carried out on a PANalytical X’Pert PRO MPD system (PW3040/60, Philips, The Netherlands) equipped with a PIXcel detector using Cu K\(\alpha\) radiation with \(\lambda = 1.542\text{"A}\) (45kV and 40mA) and automatic divergence slit. Samples were prepared on a zero background aluminum wafer and scanned from 5° to 40° 2θ with a step size of 0.026° and scanning speed of 0.067° per second.

2.6 Scanning electron microscopy
The particle morphology and size were analyzed by Scanning Electron Microscopy (SEM; XL30, FEI, Hillsboro, Oregon) and Helos Particle Size Analyzer (H0793, Sympatec, Clausthal, Germany) equipped with Aspiros and Vibri feeders and using Window 5 software.

2.7 Data treatment

All data processing was performed in Matlab 7.14.0 (Mathworks™, Natick, Massachusetts) using the PLS_Toolbox 6.7.1 (Eigenvector, Wenatchee, Washington). Both THz and NIR spectra were pre-processed with standard normal variate correction (SNV) [35] or when explicitly mentioned first or second derivatives (Savitzky-Golay) in order to minimize effects of scatter, minor differences in compact thickness and baseline drift [36].

Principal Component Analysis (PCA) [37] was applied to explore the main variations in both the NIR and terahertz spectral datasets. Partial Least Squares (PLS) regression [38] was performed on NIR and terahertz spectral datasets using the crystalline concentrations (% w/w) as reference. Variable selection through application of interval PLS (iPLS) [39] was used to identify the spectral ranges that contain the most information about the crystalline concentration. The complexity of each model was estimated by finding the minimum or breakpoint in the root-mean-square-error of cross validation (RMSECV) curve. For cross validation the samples were divided into three groups. The samples were ranked after concentration level and every third sample was taken out, so replicates were left out at the same time. RMSECV are calculated as follows:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$  \hspace{1cm} (1)

where $y_i$ is the crystallinity (here % w/w), $\hat{y}_i$ is the predicted value, and $N$ is the total number of samples.
The LOD and LOQ were calculated as described in the guideline Q2 from the International Conference on Harmonization (ICH) [40]:

\[
\text{LOD} = 3.3\sigma \\
\text{LOQ} = 10\sigma
\]

where \(\sigma\) is the standard deviation of the prediction of the 100 % amorphous samples.

3. Results and discussion

3.1 Particle size and morphology

Scanning electron micrographs of amorphous and crystalline lactose are presented in Figure 1, which illustrates that the amorphous lactose particle size is smaller compared to the crystalline. The average of three determinations of the median particle size, \(D_{50}\), ± standard deviation, was 5.7 ± 0.4 \(\mu m\) for amorphous lactose and 20.5 ± 0.3 \(\mu m\) for crystalline lactose. It can be observed that the spray drying created nearly spherical particles, and that some agglomerating of smaller particles is present in both samples.

![Figure 1: SEM of crystalline \(\alpha\)-lactose monohydrate (left) and amorphous lactose (right).](image-url)
3.2 Terahertz spectroscopy

Due to a large compact pellet thickness (approximately 1 mm) only the spectral range from 0.1-1.0 THz is useful for further analysis. Figure 2 presents the terahertz spectra of binary mixtures of amorphous and crystalline lactose with increasing level of crystallinity. By examining the concentration arranged spectra, the peak at 0.53 THz, due to an externally hindered rotational mode of the hydrogen-bonded crystalline lattice [31] becomes visible at approximately 10 % and increases with higher levels of crystallinity. Taday [41] found that crystalline lactose analyzed with terahertz spectroscopy has strong absorption lines at 0.58, 1.51, and 2.83 THz and McIntosh et al. [28] found the region between 0.3 and 0.8 THz to be the most informative for studying lactose due to larger scattering from the amorphous lactose above this frequency range. When the crystallization of amorphous lactose occurred, the peak at 0.53 THz increased in intensity and the elevated background at the higher frequencies decreased [28]. This matches well with the strong background seen at frequencies above 0.6 THz for low levels of crystallinity present in Figure 2 (low sample numbers).
Figure 2: Terahertz spectra 0.1-1.0 THz of lactose samples with increasing level of crystallinity.

Although a selective peak for lactose monohydrate is present in the current data set, particulate scattering and other baseline effects make it less likely for a multivariate data analytical approach to yield a more robust model than a univariate/single-peak method [42]. Hence, in order to understand the main variations in the underlying data set, an exploratory PCA analysis of the mean centered terahertz spectra colored according to level of crystallinity is presented in Figure 3a. As can be seen in Figure 3a, the two main principal components, PC1 and PC2, explain 98% of the total spectral variation and spreads out the samples from lowest to highest level of crystallinity. The corresponding loading plot for the PCA model is presented in Figure 3b. The loading plot of PC1 shows a positive baseline tilt (amorphous) and a negative peak (crystalline) which confirm that the two components are inversely correlated to each other in the calibration design. PC2 de-
scribe primarily the crystalline form and only to a small extent a little baseline tilt. A rapid oscillation can be seen at the baseline for both PC1 and PC2, but it is especially pronounced for the second loading. These oscillations are not inherent to THz-TDS and can be removed without affecting the shape of the significant features, e.g. using Spatially Variant Moving Average Filter [43]. However, it is doubtful that such a data treatment will change the calibration model as the most important feature of variation is the peak at 0.53 Hz.

Figure 3: a) Score plot from PCA of mean centered terahertz spectra, colored by level of crystallinity from amorphous (cyan) to 100 % (w/w) crystalline (pink). b) Corresponding loading plot for 0.1-1.0 THz for PC1 and PC2.

The predicted versus measured plot from the two-component PLS model of terahertz spectra (SNV plus mean centered) is shown in Figure 4a. The pattern of the predicted data points seems to follow a weak sigmoidal shaped curve which indicates that some non-linearity is present. However, the lower concentrations (up to around 30 % w/w) seem to fit one trendline and the highest concentrations (from 60-70 % w/w and up) seem to follow another trendline. The concentrations in the middle of the calibration range are not very well explained by this model. This can be explained by the fact that the THz signals for the amorphous and crystalline components are measured by
two different types of signals and that none of them are very sensitive to a low concentration. It is also possible that the distribution of the objects with many points at the extremes and only a sparse set in the middle of the concentration range, bias the model towards the extremes. The corresponding PLS loadings are shown in Figure 4b. From the figure it is obvious that the main information comes from the crystalline lactose peak at 0.53 THz and to a lesser extent from the 'elevated' background of the amorphous component in the range 0.6-1.0 THz.

As the main interest of this study is to estimate LOD and LOQ, regression models built on the reduced concentration range 0-10 % (w/w) crystalline lactose were calculated. The best result, using second derivative preprocessing, is shown in Figure 5. It was also investigated if an interval PLS model could improve the model performance. However, this was not found to be the case (results not shown).
Figure 5: PLS model of crystallinity 0-10 % (w/w) from Terahertz spectra treated with SNV, second derivative and mean centering. Three latent variables are used and the RMSECV is 0.30 % (w/w).

3.3 NIR spectroscopy

Figure 6 displays the SNV treated NIR spectra colored by crystallinity. A shift in the 1st overtone for O-H bands from 1460 nm to 1540 nm is observed when the crystallinity increase. In a similar experimental set-up Gombás et al. [8] found that the shift in O-H band occurred from 1480 nm to 1550 nm. Other spectral differences between amorphous and crystalline lactose observed in Figure 6 are the combination bands at 1900-1980 nm (O-H stretch and O-H deformation of H$_2$O) [16], the combination bands at 2040-2100 nm (O-H stretch and O-H deformation of ROH lactose) [8, 16], and the combination bands at 2220-2280 nm (O-H stretch and O-H deformation of lactose) [16].
Figure 6: Plot of SNV treated NIR spectra colored by crystallinity where amorphous is blue, 50 % (w/w) crystalline is green and 100 % (w/w) crystalline is red.

Figure 7: a) PCA score plot of NIR spectra after SNV and mean centering colored by crystallinity from amorphous (cyan) to 100 % (w/w) crystalline (pink). b) Corresponding loading plot for PC1 and PC2.

The PCA score plot based on the SNV and mean centered NIR spectra is presented in Figure 7a and the corresponding loading plot in Figure 7b. Most of the variation in the NIR spectra (97 %) is explained by the first principal component that clearly explains the degree of crystallinity.

When developing PLS models for the crystallinity, the best model was found using the combination of SNV preprocessing, mean centering and variable selection using the iPLS algorithm. The model using the full concentration range (0-100 % w/w) is shown in Figure 8, and the selection of
the spectral range found with iPLS is shown in Figure 9. The spectral range that gave the best calibration model for the full concentration range (0-100 % w/w), was 2015-2095 nm, which corresponds to the combination band of O-H for lactose (2040-2100 nm [16]).

![Figure 8: a) PLS calibration of 0-100 % (w/w) crystallinity for SNV, mean centered NIR spectra. The spectral area used is 2015-2095 nm. Two latent variables are used for the model and RMSECV is 0.87 % (w/w). b) Corresponding loading-weights plot from 2015-2095 nm.](image)
Figure 9: iPLS of crystallinity for SNV, mean centered NIR spectra. The red line is the RMSECV for global model using all wavelengths (1000-2500), where the individual bars are the RMSECV using each interval (based on 50 spectral variables). The white area (2015-2095 nm) is the one chosen for the full concentration range model. Local models are compared to the full-spectrum, global model by RMSECV. All intervals are tested individually and the one with the lowest error is chosen. Then the intervals are tried in combination with this selected interval, and intervals are added until the RMSECV increases by introduction of new intervals.

The best PLS models for NIR in the range of 0-10 % (w/w) crystalline lactose is shown in Figure 10a. The spectral areas that gave the best calibration, chosen by iPLS, are shown in the loading plot in Figure 10b. It is obvious that the main information used in the calibration is in the area 1870-2014 nm (O-H combination band, 1900-1980 nm [16]) and to a lesser extent 1537-1560 (O-H 1st overtone) and the area around 1200 nm, probably due to C-H 2nd overtone [8]. Obviously, the spectral areas chosen for the two concentration ranges are different. This is possibly due the fact that for 0-10 % (w/w) crystallinity, only minor differences in the spectra of the two species are present, and therefore more diverse spectral areas are required to support a reliable calibration.
Figure 10: a) PLS model of 0-10 % (w/w) crystallinity from SNV, first derivative, and mean centered NIR spectra. The spectral areas used are 1172-1185, 1200-1213, 1273-1304, 1537-1560, 1870-2014 nm, found by iPLS. There are used three latent variables for the model and RMSECV 0.20 % (w/w). b) Corresponding loading-weights plot.

Table 1: Overview of best achieved calibrations for the crystallinity by terahertz (THz) and NIR spectroscopy for 0-100 % (w/w) and 0-10 % (w/w) crystallinity.¹

<table>
<thead>
<tr>
<th>Pre-processing</th>
<th>THz 0-100 % (w/w)</th>
<th>NIR 0-100 % (w/w)</th>
<th>THz 0-10 % (w/w)</th>
<th>NIR 0-10 % (w/w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>0.1-1.0 THz</td>
<td>2015-2095 nm</td>
<td>0.1-1.0 THz</td>
<td>1172-1185, 1200-1213, 1273-1304, 1537-1560, 1870-2014 nm</td>
</tr>
<tr>
<td>Latent variables</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>RMSECV (% w/w)</td>
<td>2.91</td>
<td>0.87</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>LOD (% w/w)</td>
<td>3.21</td>
<td>3.76</td>
<td>0.80</td>
<td>0.43</td>
</tr>
<tr>
<td>LOQ (% w/w)</td>
<td>9.91</td>
<td>11.04</td>
<td>2.42</td>
<td>1.31</td>
</tr>
</tbody>
</table>

Table 1 gives a comparison of the models for crystallinity using terahertz or NIR spectroscopy for the full concentration range (0-100 % w/w), and for the low concentration range (0-10 % w/w).

¹ It should be noted that all the results and discussions are based on the cross-validated results. It would have been more optimal to have a test-set, but the results are still of such a clear manner, that we believe that our conclusions would hold even if the models had been presented for a test-set.
The table reveals that NIR spectroscopy in general performs significantly better than terahertz spectroscopy to predict lactose crystallinity. However, in the models for the full concentration range, the LOD and LOQ are higher for the NIR measurements than for terahertz measurements, and much higher than for the low concentration calibrations. For NIR calibration of the full concentration range, the chosen spectral area may describe phenomena that are mainly present in the high concentration of crystalline lactose, e.g. the presence of water bound to the lactose molecule. Therefore, this model performs poorly in explaining the amorphous lactose (LOD, LOQ), despite the lower RMSECV. In the low concentration calibration model for NIR, the spectral areas included in the calibrations are better at explaining the amorphous samples compared to the full concentration calibration. The preprocessing methods applied on the two calibrations are also different. In the low concentration models (0-10 % w/w crystallinity), both for NIR and terahertz data, derivatives are necessary to obtain good models. This makes sense for terahertz spectroscopy, where the amorphous lactose is seen as a broad background, and this is effectively removed by second derivatives [36]. Nørgaard et al. [16] quantified crystalline lactose in binary mixtures of amorphous and crystalline lactose (50-98 % w/w) with a RMSECV of 0.78 % (w/w), and mixtures of the same samples combined with whey permeate powder samples with a RMSECV of 0.63 %.

For binary mixtures of amorphous and crystalline lactose in the range of 10-90 % (w/w), Fix and Steffens [9] found a RMSECV of 1.79 % (w/w). Calibration with 0-10 % (w/w) crystalline lactose gave an RMSECV of 0.54 % (w/w), and LOD and LOQ of 0.75 % (w/w) and 2.5 % (w/w), respectively [9]. In this study LOD and LOQ for NIR (0-10 % w/w) are lower than the value reported by Fix and Steffens [9] in a similar experiment.
NIR spectroscopy, in general, performs better than terahertz spectroscopy for determining crystalline impurities or conversion to the crystalline state in amorphous lactose. However, NIR spectroscopy is a mature method that has been refined over many years, whereas currently terahertz spectroscopy is still developing. Terahertz spectroscopy still faces the instrumental challenges but has the potential to become a valid method in quantifying crystallinity in pharmaceuticals.

4. Conclusion

Despite the fact that NIR spectroscopy measure only indirectly intermolecular interactions and terahertz spectroscopy directly probe lattice vibrations, this study has shown that NIR spectroscopy in general performs better than terahertz spectroscopy for determining crystalline impurities or conversion to the crystalline state in amorphous lactose. PLS regression models were developed and the RMSECV for the models with full concentration range were found to be 2.91 % (w/w) and 0.87 % (w/w) for THz-TDS and NIR, respectively. Calibrations created on samples containing 0-10 % (w/w) crystalline material gave RMSECVs of 0.30 % (w/w) and 0.20 % (w/w) for THz-TDS and NIR, respectively, while the limits of detection were 0.80 % (w/w) and 0.43 % (w/w), respectively.

5. Acknowledgment

The authors wish to thank Mogens Frost, H. Lundbeck A/S for performing SEM and laser diffraction measurements, Tianwu Wang, DTU Fotonik, for assistance with terahertz spectroscopy, and Mingshi Yang for support with spray drying.
6. Funding

The authors wish to thank the Danish Ministry of Science, Technology and Innovation for financial support of a PhD stipendium to Solveig Warnecke through the QbD consortium (www.qbd.dk). Jian X. Wu acknowledges The Danish Council for Technology and Innovation via the Innovation Consortium NanoMorph (952320/2009) for financial support of his PhD stipendium.
7. Reference List