

Research Article

Low-Content Quantitation in Entecavir Tablets Using 1064 nm Raman Spectroscopy

Yanlei Kang,¹ Yushan Zhou,² Qiaoyu Wu,¹ Ning Wang,¹ and Jianguang Zhou ¹

¹Research Center for Analytical Instrumentation, Institute of Cyber-Systems and Control, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

²School of Chemical Engineering, Zhejiang University of Technology, Hangzhou, China

Correspondence should be addressed to Jianguang Zhou; jgzhou@zju.edu.cn

Yanlei Kang and Yushan Zhou contributed equally to this work.

Received 27 March 2020; Revised 10 June 2020; Accepted 12 June 2020; Published 28 June 2020

Academic Editor: Wagner A. Alves

Copyright © 2020 Yanlei Kang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The nondestructive and high sensitive analysis of a low content of an active pharmaceutical ingredient (API) was a difficult problem, especially in a complex system of pharmaceutical formulations. In this paper, a rapid and no sample preparation method was developed, which used a 1064 nm Raman spectrometer to detect entecavir monohydrate (ETV-H) in Baraclude tablets. Entecavir was a drug approved by FDA for the treatment of chronic hepatitis B and became the first choice in the market. The wavelength selection results displayed that the signal-to-background ratio of the Raman spectrum with 1064 nm excitation wavelength was 14 times that of the commonly used 785 nm wavelength. The partial least squares (PLS) method was used to calibrate concentration models containing 0.1% to 1.0% w/w% ETV-H in calibration set samples. Different preprocessing methods were used to eliminate the background interference and extract more spectral information. Calibration samples were used to choose the best performing model. Then, all the calibration samples combined with the best performing models' parameters successfully predicted the content of ETV-H in Baraclude tablets. Combining baseline processing and standard normal variate (SNV) with PLS, the model showed a good result with an R^2 of 0.973, RMSEC of 0.05%, and RMSEP of 0.03% on the spectral region of 1350–1700 cm^{-1} . The limit of detection of this model was 0.17%. These results showed that 1064 nm Raman spectroscopy technology could be an alternative analytical procedure to quantify low-content API in intact tablets.

1. Introduction

In 2005, Baraclude (entecavir) tablets were approved by the FDA for the treatment of chronic hepatitis B, which had infected more than 400 million people in the world [1]. In the studies of FDA, people treated with Baraclude showed significant improvement in the liver inflammation and liver scarring caused by HBV. So, entecavir (ETV) had been the first choice for hepatitis B treatment in the market [2]. ETV was a carbocyclic 2'-deoxyguanosine analogue, which could phosphorylate into the form of triphosphate that can inhibit HBV in active cells. The chemical name for ETV was 2-amino-1, 9-dihydro-9-((1S, 3R, 4S)-4-hydroxy-3-(hydroxymethyl)-2-methylenecyclopentyl)-6H-purin-6-one. ETV

monohydrate (ETV-H) was the API of commercial ETV pharmaceutical tablets.

In the pharmaceutical industries, pharmaceutical quality control (QC) and quality assurance (QA) were necessary at all stages of product processing from a raw material to a packaged product. In the pharmaceutical production process, the final drug product might contain an insufficient or excessive content of API [2, 3]. So, QC of the active pharmaceutical ingredients (APIs) content in pharmaceutical tablets was necessary to ensure the safety and efficacy of the drug products [4–6]. With respect to low content API formulations, FDA described that the API in the formulations was less than 1% [7]. The possible problem for low-content API formulations was the low relative potency of the

final product due to its vulnerability to loss and contamination in the production process [8]. So, it was more significant for the QC of low-content API formulations. Traditional QC methods such as High-Performance Liquid Chromatography (HPLC), solid state NMR spectroscopy, Mass Spectroscopy (MS), and X-ray powder diffraction were time consuming, destructive, expensive, and required lengthy sample preparation [9–13]. Motivated by the process analytical technology (PAT) initiative, QC techniques were being developed to provide real-time process control for better understanding of the chemical and physical processes during the production process. Vibrational spectroscopic techniques such as Near Infrared spectroscopy (NIRS) and Raman spectroscopy emerged as valuable tools for pharmaceutical quality analysis because of their fast analysis speed, no need for sample preparation, and no damage to the sample [14–17]. Due to the higher resolution than NIR, Raman spectroscopy became the more promising QC technique in the pharmaceutical industry [18–20]. Griffen et al. reported quantification of the low level polymorph content (0.62–1.32%) in tablets by transmission Raman spectroscopy [21]. Li et al. had demonstrated low-content (<0.1%) quantification in powders using 785 nm Raman spectroscopy [22]. Assi et al. studied the application of a handheld Raman spectrometer for the quantification of ciprofloxacin in proprietary ciproxin tablets and generic ciprofloxacin tablets [23]. Daniela et al. studied the potential of Raman spectroscopy as process analytical technology (PAT) for the in-line and real-time monitoring of the powder blending process and proved that Raman spectroscopy was effective in the determination of API in the tablets. They also found that Raman performed better than the traditional HPLC analysis [24]. Liljana et al. compared three technologies including midinfrared, near-infrared, and Raman spectroscopy for the quantitative analysis of low API in blending powders and found that Raman presented the most favorable statistical indicators in this comparative study [25].

Since the Raman signal was proportional to the concentration of the analyte in the sample, the application of univariate regression analysis (peak height/peak area) was theoretically feasible [26, 27]. However, in the quantitative analysis of most tablets, useful single peaks were difficult to find due to the interference of other components in the tablet. So, the application of more complex multivariate methods such as PLS (partial least-squares) regression and PCA (principal components analysis) was required. The multivariate methods could handle thousands of variables to improve the accuracy of the methods, since not only the intensity or area of the selected bands but also the variation over the entire spectral range were considered. Therefore, Raman spectroscopy combined with complex multivariate algorithms provided a promising basis for the development of successful quantitative models, even for fairly complex mixtures or low content of the tested components. Farias et al. applied Raman spectroscopy combined with PLS to determine and quantify crystalline forms of the API in final products as obtained in the production lines [28]. Maxwell et al. reported the development of chemometric models (PLS and PCR) of Raman spectroscopy to determine the polymorphic changes of theophylline in pharmaceutical products [29].

Before building quantitative models of APIs in tablets, it was necessary to pretreat the data to remove the interference information in Raman spectra. Baseline correction was commonly used in Raman spectroscopy to eliminate fluorescent contaminants and instrumental factors. Three commonly used preprocessing methods were multiplicative scatter correction (MSC), standard normal variate (SNV), and Savitzky–Golay derivatives (SG derivatives). The purpose of MSC was to remove scattering artifacts and that of SNV was to remove the scattering variations between measurements [30, 31]. SG derivatives were used for the purpose of removing noise and background variances [32].

In the present study, Raman spectroscopy with an excitation wavelength of 1064 nm was used to quantify the low-content ETV-H in tablets for the first time. Calibration samples with different concentrations of ETV-H in the dosage forms were prepared and determined by Raman spectroscopy. Pretreated methods (baseline correction, MSC, SNV, and Savitzky–Golay first and second derivatives) were performed for the data. Then, the data were used to build the quantitative model to detect the low-content ETV-H in Baraclude tablets in the market.

2. Materials and Methods

2.1. Materials. ETV-H was obtained from Zhejiang Ausun Pharmaceutical Co., LTD (purity > 99.9%). Lactose monohydrate (Lac-H), Microcrystalline cellulose (MCC), Crospovidone (PVPP), Polyvinyl pyrrolidone K30 (PVP K30), and Magnesium stearate (MgSt) were purchased from Aladdin Ltd. (Shanghai, China). Baraclude® (Entecavir, Bristol Myers Squibb Company) was purchased from the local pharmacy.

2.2. Preparation of Calibration and Test Set Samples. As the commercially available Baraclude tablets only had an API content of 0.25% w/w%, the calibration set was developed as with an ETV-H mass concentration of 0.1%–1.0% w/w%. The content of each component in the calibration set samples is shown in Table 1. All the components were sieved through a 400-mesh sieve (with an aperture of 38 μm) for homogeneity and weighed according to the quantities listed in the table. In order to mix absolutely, the total weight of each concentration sample was increased to 2000 mg. Tablets were prepared by mixing the API and all excipients using an MX-S vortex oscillator. All concentration samples were divided into ten tablets, and each tablet (200 mg) was pressed by using a YP-15 manual powder compactor (Josvok technology co. LTD, Tianjin, China) with a 10 mm die set. The compression force was 25 MPa with a dwell time of 1 min. Each concentration was tested by Raman spectroscopy using three tablets.

The test set contained two different specifications of Baraclude tablets (0.5 mg and 1.0 mg ETV-H in tablet, the concentration of ETV-H both was 0.25% w/w%) and homemade tablets with an ETV-H mass concentration of 0.5% w/w%. Each test sample was determined by Raman spectroscopy using two tablets.

TABLE 1: Contents of each tablet with different API concentrations in calibration samples and test samples.

API mass concentration	ETV-H (mg)	Lac-H (mg)	MCC (mg)	PVPP (mg)	PVP K30 (mg)	MgSt (mg)
Calibration samples						
0.1%	0.2	120.7	65.1	8.0	5.0	1.0
0.2%	0.4	120.6	65.0	8.0	5.0	1.0
0.3%	0.6	120.4	65.0	8.0	5.0	1.0
0.4%	0.8	120.3	64.9	8.0	5.0	1.0
0.6%	1.2	120.0	64.8	8.0	5.0	1.0
0.8%	1.6	119.8	64.6	8.0	5.0	1.0
1.0%	2.0	119.6	64.5	7.9	5.0	1.0
Test samples						
Baraclude 0.25%	0.5	120.5	65.0	8.0	5.0	1.0
Baraclude 0.25%	1.0	240.0	130.0	16.0	10.0	2.0
0.5%	1.0	120.2	64.8	8.0	5.0	1.0

2.3. Instrumentation and Software. The Raman spectra were collected by using a Rigaku Progeny handheld Raman spectrometer (Rigaku Co., Tokyo, Japan) with a 1064 nm high-power excitation laser. The instrument could give a maximum laser power of 490 mW at the source. The actual laser power reaching the sample was 142 mW. The focused spot diameter was 25 μm . Raman spectra were recorded in the wave number range 200–2500 cm^{-1} at a resolution of 8–11 cm^{-1} with transmission volume phase gratings.

A LabRAM HR Evolution Raman spectrometer (Horiba Jobin Yvon Inc.) with 532 nm, 633 nm, and 785 nm excitation wavelengths was also used to obtain the Raman spectra of ETV-H. The laser output power was set to 10 mW (maximum output power), and the integration time was set to 10s over the Raman shift range of 200 cm^{-1} to 4000 cm^{-1} .

Multivariate data analysis was carried out using Matlab R2017b (Mathworks Inc., MA, USA).

2.4. Spectral Pretreatments and Chemometrics. In order to establish better analysis and prediction methods, it was necessary to select the wave number range that was favorable for predicting analytes and eliminating noise before quantitative analysis. This study chose the correlation coefficient method to select the wave number range. The correlation coefficient method was realized by calculating the correlation between the ETV-H concentration in tablets and wave numbers. The correlation coefficient is calculated by

$$r_j = \frac{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

where j represented the j th wave number and i represented the i th sample, $x_{i,j}$ was the Raman intensity of the j th wave number of the i th sample, \bar{x}_j was the mean of the j th wave number, \bar{y} was the mean reference value for all samples, and y_i was the reference value of the i th sample.

After selecting the wave number range, Raman data should be pretreated by preprocessing methods.

Baseline correction: this method selected the polynomial fitting method for the pretreatment of the spectra. Savitzky–Golay first and second derivatives, standard normal variant (SNV), and multiplicative scattering correction (MSC) were performed before quantitative analysis. The first

and second Savitzky–Golay derivatives were performed with a window size of 10 points and a second-order polynomial. The calibration set was validated by dividing the samples randomly into two sets (one set was the calibration set and another set was the validation set).

Partial least squares (PLS) were used for quantitative modeling. To avoid the overfitting of the model, the most suitable PLS model should have a low number of PLS latent variables (factors) and low values of three parameters, which were the root-mean-square error of cross validation (RMSECV), root-mean-square error of calibration (RMSEC), and root-mean-square error of prediction (RMSEP). RMSE was defined as the following equation:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}, \quad (2)$$

where x_i is the measurement value, y_i is the prediction value, and n is the number of samples [33].

Limit of detection (LOD) of the calibration model was defined as follows [21]:

$$\text{LOD} = \frac{3.3\sigma}{S}, \quad (3)$$

where σ is the standard deviation of the regression fit and S is the slope of the calibration curve.

3. Results and Discussion

3.1. The Determination and Selection of Different Raman Wavelengths. The ETV-H solid powder was measured by different excitation wavelengths of Raman. As shown in Figure 1, the Raman spectrum obtained from 1064 nm excitation wavelength got higher signal-to-background (SBR) and lower background fluorescence interference than the 785 nm, 633 nm, and 532 nm Raman spectra. For the 532 nm excitation wavelength spectrum, the SBR ratio (Signal/Background, S/B) at 1487 cm^{-1} (the highest peak in the spectrum) was 0.1. Also, the S/B was 1.0 for 633 nm, 1.9 for 785 nm, and 27.3 for 1064 nm. Obviously, the S/B of the spectrum with an excitation wavelength of 1064 nm was significantly higher than several other wavelengths. S/B of the Raman spectrum with 1064 nm excitation wavelength was 14 times that of the commonly used 785 nm wavelength.

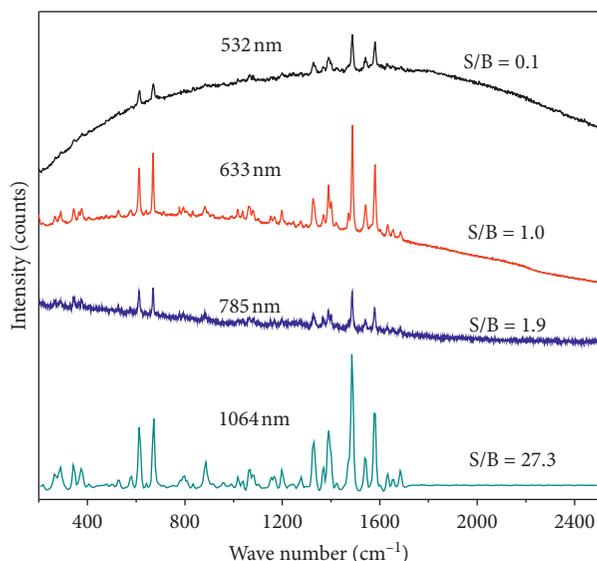


FIGURE 1: Raman spectra of ETV-H obtained by different excitation wavelengths.

Therefore, this work used 1064 nm excitation wavelength to quantify ETV-H in tablets.

3.2. Chemometric Models for Quantitation of Baraclude Tablets. Calibration set samples (21 samples) with different API mass concentrations (w/w%) were measured by Raman spectroscopy at 1064 nm excitation wavelength. The coating layer of Baraclude tablets (the ingredient was Opadry®) was carefully scraped off with a special blade. After the coating layer was removed, Baraclude tablets (0.5 mg and 1.0 mg ETV-H in tablet, the concentration of ETV-H was both 0.25% w/w%) were measured to obtain Raman spectra of the test set. All Raman spectra of the samples are shown in Figure 2. With the increase of the concentration of ETV-H in tablets, Raman intensity changed slightly at the peak of 1580 cm^{-1} (characteristic Raman peak of ETV-H). So, it was necessary to select the wave number range instead of the single peak.

Raman spectra of ETV-H and all excipients are presented in Figure 3. The assignments of all Raman characteristic peaks of ETV-H and all excipients are listed in Table 2. The main features of ETV-H Raman spectra were the strong bands at 1487 cm^{-1} and 1580 cm^{-1} , which were assigned to the stretching vibration of N-C and bending vibration of H-N-H. The strong band at 614 cm^{-1} was assigned to the bending vibration of C-C-C, and the other strong band at 672 cm^{-1} was assigned to the torsion vibration of N-C-N-C, N-C-N-H and out-of-plane deformation vibration of N-N-N-C, O-C-N-C. The bending vibration of OCH and torsion vibration of H-O-C-H gave rise to a strong band at 1329 cm^{-1} . The stretching vibration of N-C, C-C, O-C and bending vibration of N-C-H, H-C-H gave rise to Raman bands of 1065 cm^{-1} , 1196 cm^{-1} , and 1684 cm^{-1} .

The Raman characteristic peaks of lactose monohydrate (Lac-H) were 355 cm^{-1} , 376 cm^{-1} , 398 cm^{-1} , 473 cm^{-1} ,

852 cm^{-1} , 879 cm^{-1} , 915 cm^{-1} , and 954 cm^{-1} . The strong bands in the region of $852\text{--}915\text{ cm}^{-1}$ were ascribed to the stretching vibration of O-C and C-C and out-of-plane deformation vibration of O-C-C-C. The strong bands at 355 cm^{-1} , 376 cm^{-1} , 398 cm^{-1} , and 473 cm^{-1} were assigned to the bending vibration of O-C-C. The $954\text{--}1138\text{ cm}^{-1}$ region could be assigned to the stretching vibration of O-C. The strong bands in the region of $1220\text{--}1472\text{ cm}^{-1}$ were ascribed to the bending vibration of H-C-O and H-C-H.

Raman characteristic peaks for microcrystalline cellulose (MCC) were 379 cm^{-1} , 439 cm^{-1} , 455 cm^{-1} , 899 cm^{-1} , 972 cm^{-1} , 998 cm^{-1} , 1036 cm^{-1} , 1063 cm^{-1} , 1094 cm^{-1} , 1123 cm^{-1} , 1154 cm^{-1} , and 1380 cm^{-1} . The band region of $439\text{--}899\text{ cm}^{-1}$ could be assigned to the out-of-plane deformation vibration of O-C-C-C. Also, the stretching vibration of O-C appeared as a series of bands in the region of $972\text{--}1154\text{ cm}^{-1}$. A strong band at 379 cm^{-1} was ascribed to the bending vibration of H-O-H, O-C-C. In addition, the other strong band at 1380 cm^{-1} was assigned to the bending vibration of H-C-O.

It could be concluded that ETV-H differed from all the excipients in the spectra region of $1350\text{--}1700\text{ cm}^{-1}$. This meant that API could be distinguished in this region. The correlation coefficient method was also used to select the characteristic wavelength, and the results showed that the wave number range of $1350\text{--}1700\text{ cm}^{-1}$ could be used as the feature range. After selecting the characteristic wavelength range, the number of wave numbers used for the quantitative model could be reduced from 512 to 72.

The data used to support the findings of this study are available from the corresponding author upon request.

The blue dotted frame represents the selected spectral range of $1350\text{--}1700\text{ cm}^{-1}$.

Partial least squares (PLS) method was used to calibrate concentration models in 21 calibration set samples containing 0.1%, 0.2%, 0.3%, 0.4%, 0.6%, 0.8%, and 1.0% w/w ETV-H. The optimum number of PLS latent variables (LVs) was chosen by comparing the root mean squared error of cross validation (RMSECV, leave-one-out validation) to ensure that most of the variations were included [34]. As shown in Figure 4(a), three PLS components were selected based on the minimal RMSECV values obtained by combining baseline correction and SNV for the pretreatment of the samples within the wave number range of $1350\text{--}1700\text{ cm}^{-1}$. This methodology avoided overfitting of the model when excess potential variables were selected.

A latent variables loadings plot of the model (three PLS components, baseline correction, and SNV for the pretreatment of the samples within the wave number range of $1350\text{--}1700\text{ cm}^{-1}$) is shown in Figure 4(b). The loadings plot revealed that ETV-H was positively correlated with latent variable 1 (LV1) and latent variable 2 (LV2), such as the characteristic peaks of ETV-H at 1487 , 1540 , and 1580 cm^{-1} .

The positively correlated peaks in LV1 were 288 cm^{-1} , 355 cm^{-1} , 398 cm^{-1} , 473 cm^{-1} , 850 cm^{-1} , 874 cm^{-1} , 1083 cm^{-1} , 1138 cm^{-1} , 1196 cm^{-1} , 1260 cm^{-1} , 1325 cm^{-1} , 1467 cm^{-1} , 1487 cm^{-1} , 1540 cm^{-1} , 1578 cm^{-1} , and 1686 cm^{-1} . Referred to the data in Table 2, the correlated peaks at 288 cm^{-1} , 1196 cm^{-1} , 1325 cm^{-1} , 1487 cm^{-1} , 1540 cm^{-1} , 1578 cm^{-1} ,

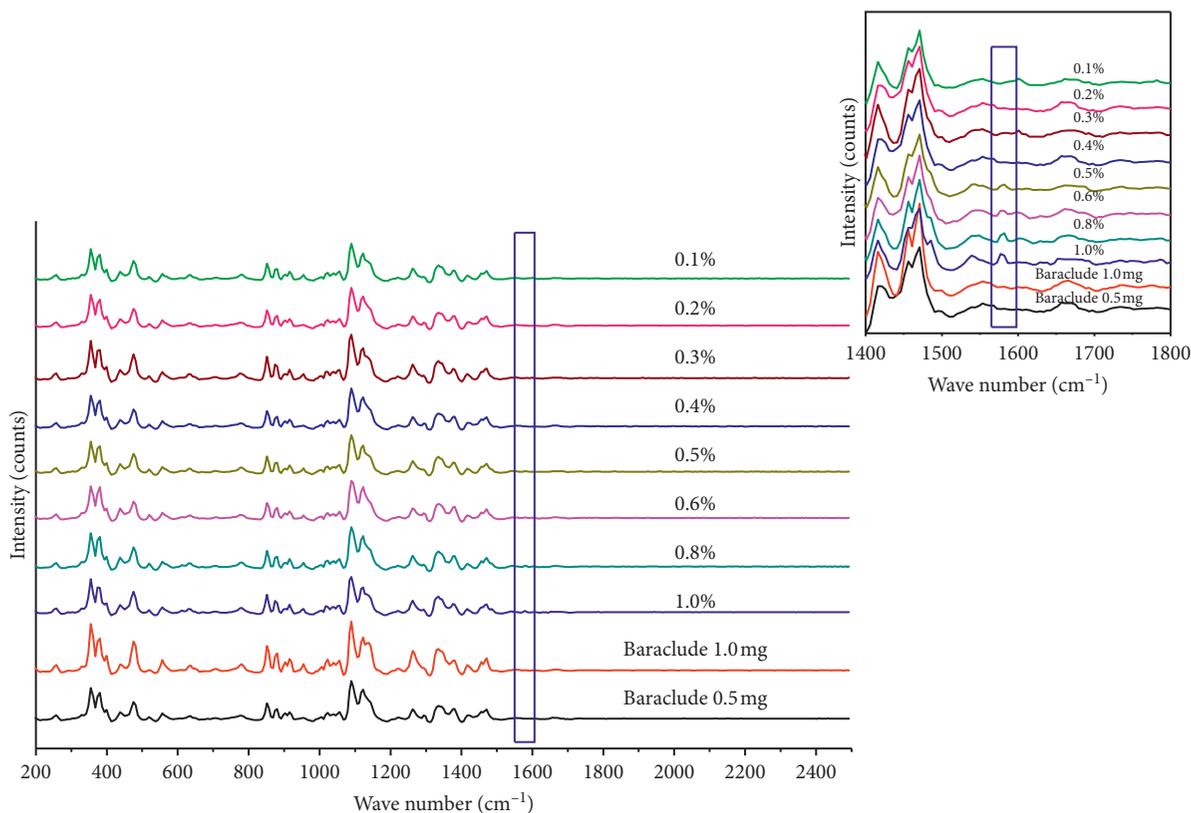


FIGURE 2: Raman spectra of calibration samples of different API mass concentrations (w/w%) and test set samples (Baraclude tablets and 0.5% w/w%). The blue box shows the intensity change of the Raman peak at 1580 cm^{-1} .

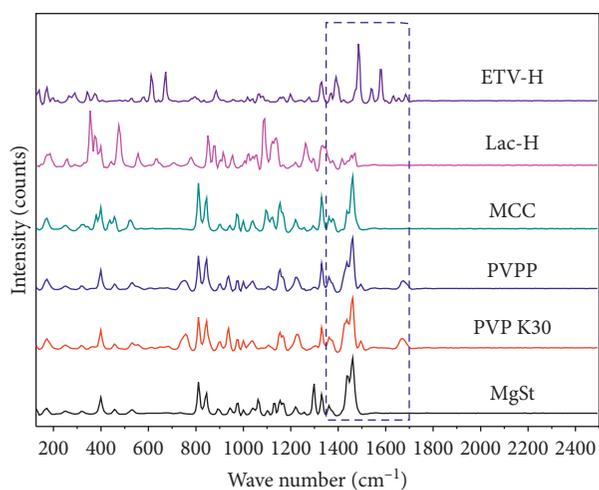


FIGURE 3: Raman spectra of raw ETV-H and all excipients used in Baraclude's formulation.

cm^{-1} , and 1686 cm^{-1} were produced by ETV-H. Other correlated peaks at 355 cm^{-1} , 398 cm^{-1} , 473 cm^{-1} , 850 cm^{-1} , 874 cm^{-1} , 1083 cm^{-1} , 1138 cm^{-1} , 1260 cm^{-1} , and 1467 cm^{-1} were originated from lactose monohydrate (Lac-H). The large positive coefficient produced by Lac-H was caused by its highest content in tablet and the content varied slightly with the change of API concentration.

The positively correlated peaks in LV2 were 379 cm^{-1} , 435 cm^{-1} , 457 cm^{-1} , 612 cm^{-1} , 670 cm^{-1} , 965 cm^{-1} ,

992 cm^{-1} , 1067 cm^{-1} , 1094 cm^{-1} , 1124 cm^{-1} , 1154 cm^{-1} , 1387 cm^{-1} , 1485 cm^{-1} , 1540 cm^{-1} , and 1576 cm^{-1} . Referred to the data in Table 2, the positively correlated peaks at 379 cm^{-1} , 435 cm^{-1} , 457 cm^{-1} , 965 cm^{-1} , 992 cm^{-1} , 1067 cm^{-1} , 1094 cm^{-1} , 1124 cm^{-1} , 1154 cm^{-1} , and 1387 cm^{-1} were produced by microcrystalline cellulose (MCC). Other correlated peaks at 612 cm^{-1} , 670 cm^{-1} , 1485 cm^{-1} , 1540 cm^{-1} , and 1576 cm^{-1} were originated from ETV-H. The reason for the large positive coefficient produced by MCC was the second highest content in tablet and the content changed slightly for the samples of different API concentrations.

The total scores of the three variables reached 99%. All the PLS components for other models were chosen for the same method.

Then, the total 21 calibration set samples were divided into a calibration set (14 samples) and a validation set (7 samples). Combining different preprocessing methods and spectral ranges, we obtained 16 models (such as the model with baseline correction as the pretreatment method in the full spectral range). As shown in Table 3, PLS LVs of all the models were calculated by the method mentioned above. The accuracy of the PLS calibration model was evaluated by assessing the correlation coefficient (R^2) and RMSEP. So, the RMSECV, R^2 , RMSEC, and RMSEP values of all the models were calculated by using the corresponding LV value. The best performing model was marked in italics in Table 3, which was the model combining the baseline correction with

TABLE 2: The assignments of Raman characteristic peaks for ETV-H and all excipients used in Baraclude tablets.

Raman peaks	Vibrational assignments
Entecavir monohydrate (ETV-H)	
288 m ^a	Torsion vibration of NCNH, OHOC, and OHOH and bending vibration of H-O-H and OCN
341 m	Bending vibration of CCO
372 m	Torsion vibration of CCOH and out-of-plane deformation of NNNC and NCNC
614 s	Bending vibration of CCC
672 s	Torsion vibration of NCNC and NCNH and out-of-plane deformation of NNNC and OCNC
885 m	Stretching vibration of OC and CC and bending vibration of CCH and NCN
1065 m	Stretching vibration of OC
1196 m	Stretching vibration of CC, NC, and OC and bending vibration of NCH
1329 s	Bending vibration of OCH and torsion vibration of HOCH
1387 s	Bending vibration of CCH and torsion vibration of CCCH
1487 vs	Stretching vibration of NC
1538 m	Stretching vibration of NC and bending vibration of CNC
1580 s	Stretching vibration of NC and bending vibration of HNH
1684 m	Stretching vibration of CC and bending vibration of HCH
Lactose monohydrate (Lac-H)	
355 vs	Bending vibration of O-C-C
376 s	Bending vibration of H-O-H and bending vibration of O-C-C
398 s	Bending vibration of O-C-C and out-of-plane deformation of O-C-C-C
473 vs	Bending vibration of O-C-C
852 s	Stretching vibration of O-C and out-of-plane deformation of O-C-C-C
879 s	Stretching vibration of O-C
915 s	Stretching vibration of C-C and O-C and out-of-plane deformation of O-C-C-C
954 s	Stretching vibration of O-C
1023 m	Stretching vibration of O-C
1056 m	Stretching vibration of O-C
1090 vs	Stretching vibration of O-C
1124 s	Stretching vibration of O-C
1138 s	Stretching vibration of O-C
1220 m	Bending vibration of H-C-O and torsion vibration of HCCC
1261 s	Bending vibration of H-C-O
1334 s	Bending vibration of H-C-O
1346 s	Bending vibration of H-C-O
1472 s	Bending vibration of H-C-H
Microcrystalline cellulose (MCC)	
379 s	Bending vibration of H-O-H and bending vibration of O-C-C
439 s	Out-of-plane deformation of O-C-C-C
455 s	Out-of-plane deformation of O-C-C-C
899 m	Out-of-plane deformation of O-C-C-C
972 m	Stretching vibration of O-C
998 m	Stretching vibration of O-C
1036 m	Stretching vibration of O-C
1063 m	Stretching vibration of O-C
1094 vs	Stretching vibration of O-C
1123 vs	Stretching vibration of O-C
1154 s	Stretching vibration of O-C
1380 s	Bending vibration of H-C-O
Crospovidone (PVPP)	
359 m	Bending vibration of O-C-C, C-C-N
552 m	Stretching vibration of C-C and out-of-plane vibration of O-C-N-C
752 s	Bending vibration of H-C-C and torsion vibration of H-C-C-C
852 m	Torsion vibration of H-C-C-N
899 m	Stretching vibration of C-C and bending vibration of CNC and CCN
932 s	Torsion vibration of HCNC and CCNC
1023 m	Stretching vibration of N-C and C-C and bending vibration of H-C-C, CNC, and CCN
1227 s	Stretching vibration of N-C, bending vibration of H-C-C, and torsion vibration of H-C-C-C
1298 m	Stretching vibration of N-C and bending vibration of H-C-C
1369 m	Bending vibration of H-C-C and H-C-H, stretching vibration of N-C, and torsion vibration of H-C-N-C
1425 vs	Bending vibration of H-C-H

TABLE 2: Continued.

Raman peaks	Vibrational assignments
1451 s	Bending vibration of H-C-H
1493 s	Bending vibration of H-C-H
1673 s	Stretching vibration of C=C and O=C
Povidone K30 (PVP K30)	
364 m	Bending vibration of OCC and CCN
559 m	Bending vibration of OCC and CCN and out-of-plane vibration of OCNC
759 s	Stretching vibration of NC and CC and torsion vibration of HCCC
852 m	Bending vibration of CCC and torsion vibration of HCCC and HCNC
899 m	Stretching vibration of CC and bending vibration of CNC and NCC
936 s	Torsion vibration of HCNC
1003 m	Stretching vibration of CC and NC and bending vibration of HCC and OCC
1025 m	Stretching vibration of NC and bending vibration of HCC and NCC
1212 m	Bending vibration of HCN and torsion vibration of HCCN and HCNC
1229 s	Stretching vibration of NC and bending vibration of HCC and HCN
1347 m	Stretching vibration of NC and torsion vibration of HCNC
1371 m	Bending vibration of HCC and HCH
1427 vs	Bending vibration of HCH
1456 s	Bending vibration of HCH
1493 s	Bending vibration of HCC and HCH
1664 s	Stretching vibration of C=C and bending vibration of HCC and HCH
Magnesium stearate (MgSt)	
890 m	Stretching vibration of C-C=O
946 m	Stretching vibration of C-C
1061 vs	Stretching vibration of O-C and C-C
1100 m	Stretching vibration of C-C
1127 s	Stretching vibration of C-C=O and C-C-O
1174 m	Stretching vibration of C-C=O and CCO
1300 vs	Out-of-plane rocking vibration of OCCC
1370 m	Out-of-plane rocking vibration of OCCC
1440 vs	Deformation vibration of H-C-H
1460 s	Deformation vibration of H-C-H

a: Raman peaks intensity, vs- very strong, s- strong, m- middle, w- weak

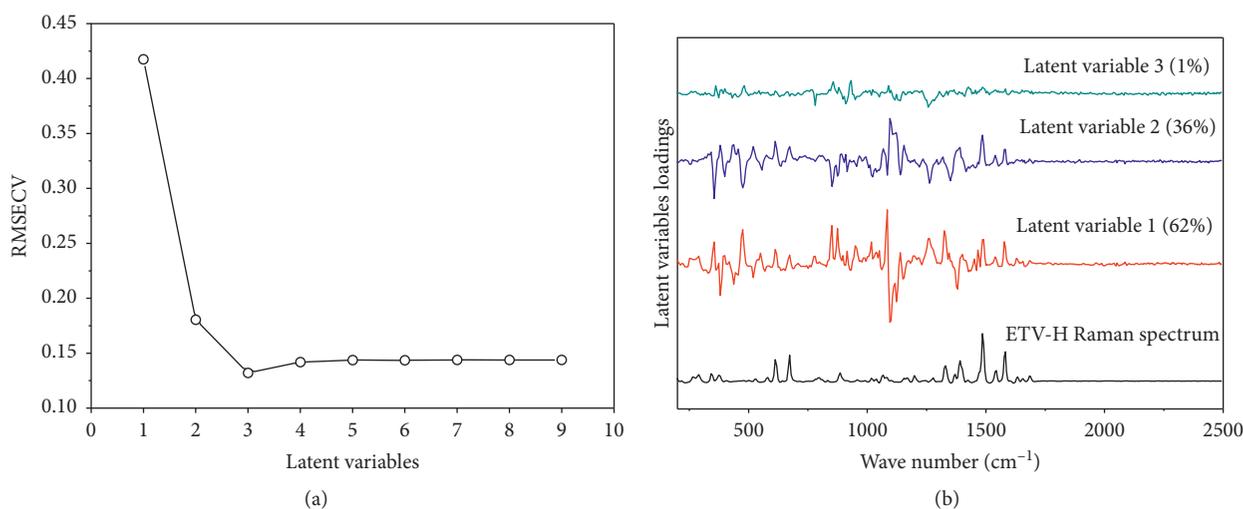


FIGURE 4: (a) RMSECV values vs. the latent variables for the PLS model. (b) Latent variables loadings plot for the PLS model.

the SNV method for the data pretreatment in the wave number range of 1350–1700 cm⁻¹. As shown in Figure 5, the squared correlation coefficient of the best model was 0.970, RMSEC was 0.05%, and RMSEP was 0.05%. These results represented the reliability and accuracy of the model.

Then, all the calibration samples (21 samples) were used to build PLS models using the best performing model parameters, and the ETV-H content in Baraclude tablets and homemade tablets was successfully predicted. The best quantitative model was obtained from the spectral region

TABLE 3: Performance of the regression models using raw Raman data of calibration set samples and validation set samples which were obtained by randomly divided the calibration samples.

Preprocessing	Region	LV	RMSECV	R^2	RMSEC	RMSEP
Raw data	Full	2	0.14	0.896	0.09	0.28
	1350–1700 cm^{-1}	3	0.13	0.925	0.08	0.30
Baseline	Full	3	0.14	0.877	0.10	0.25
	1350–1700 cm^{-1}	3	0.15	0.941	0.07	0.28
MSC	Full	3	0.16	0.896	0.09	0.28
	1350–1700 cm^{-1}	4	0.14	0.925	0.08	0.30
SNV	Full	3	0.13	0.938	0.07	0.26
	1350–1700 cm^{-1}	2	0.13	0.962	0.06	0.10
Baseline + MSC	Full	2	0.14	0.896	0.09	0.28
	1350–1700 cm^{-1}	2	0.15	0.925	0.08	0.30
<i>Baseline + SNV</i>	Full	3	0.14	0.905	0.09	0.20
	1350–1700 cm^{-1}	3	0.13	0.970	0.05	0.05
Baseline + D1 data	Full	2	0.25	0.797	0.12	0.33
	1350–1700 cm^{-1}	3	0.13	0.892	0.09	0.25
Baseline + D2 data	Full	3	0.16	0.839	0.11	0.27
	1350–1700 cm^{-1}	4	0.18	0.899	0.09	0.31

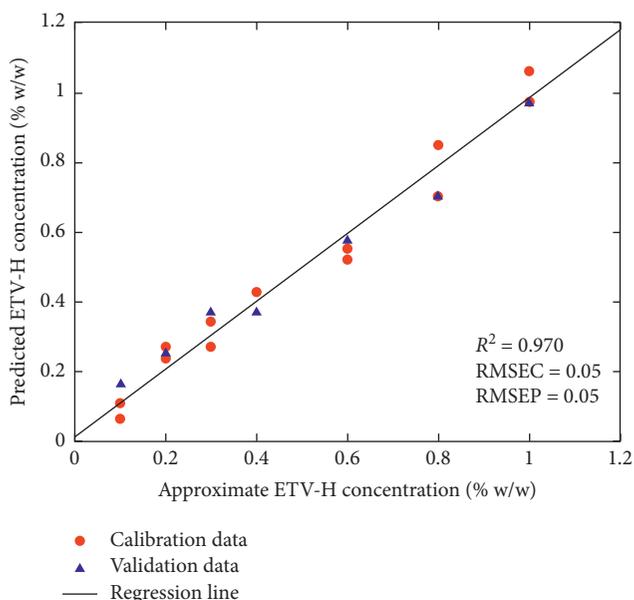


FIGURE 5: Approximate vs. predicted ETV-H concentrations in calibration set samples and validation set samples obtained by the best performance PLS models (baseline correction and the SNV method for data pretreatment in the wave number range of 1350–1700 cm^{-1}). Despite a long mixing process, the homemade tablets still could not obtain a uniform content, so it was impossible to obtain the exact concentration of ETV-H in each tablet. Approximate ETV-H concentration referred to the approximate value of the real ETV-H concentration.

1350–1700 cm^{-1} using baseline correction and SNV as the preprocessing method with the results of an R^2 of 0.973, RMSEC of 0.05%, and RMSEP of 0.03% (Figure 6). As shown in Figure 6, due to the content uniformity margin of the tablets in which the API content would fluctuate, the surface of the prepared tablets was not uniform despite the long mixing process. Raman spectroscopy could not overcome

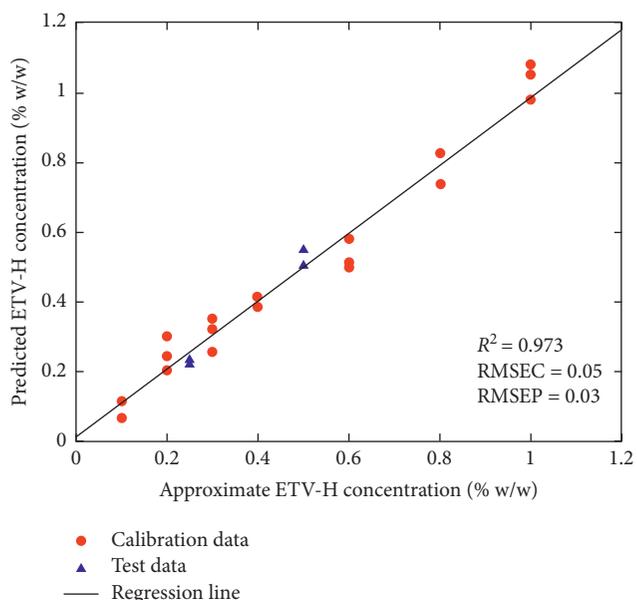


FIGURE 6: Approximate vs. predicted ETV-H concentrations in test set samples obtained by the best performance PLS models (baseline correction and the SNV method for data pretreatment in the wave number range of 1350–1700 cm^{-1}). Approximate ETV-H concentration referred to the approximate value of the real ETV-H concentration.

the problem of uneven surface content, so the exact ETV-H content in each tablet could not be determined by Raman spectroscopy. Therefore, for all samples, only approximate values for the real ETV-H concentration were used for the quantitation model. The error of the approximate value plus the error of the measured value would produce error transmission and affected the statistical indicators of the models R^2 , RMSEC, and RMSEP.

It could be seen that the original data could be used to quantify ETV-H in Baraclude tablets after baseline

TABLE 4: The actual and predicted API content and the percentage of predicted API in the drug content calculated by the best performance model.

Test set sample	Acceptance API content	Predicted API content (mg)
Baraclude 0.5 mg (90%–105%)	0.45–0.53 mg	0.47
Baraclude 0.5 mg (90%–105%)	0.45–0.53 mg	0.45
Baraclude 1.0 mg (90%–105%)	0.9–1.05 mg	0.93
Baraclude 1.0 mg (90%–105%)	0.9–1.05 mg	0.91
0.5% (w/w%)	0.5%	0.55
0.5% (w/w%)	0.5%	0.50

processing and SNV pretreatment. In addition, the prediction results of the models were not bad when using SNV pretreatment only. From this point of view, SNV performed better than other pretreatment methods for Raman spectral data. According to equation (3), the σ value of the LOD equation referred to RMSEC of the model, and S referred to slope of the calibration curve. So, the LOD of the best performance model was 0.17%. According to US Pharmacopeia guidelines, the acceptable percentage of the labeled amount of ETV-H in the drug content ranged from 90.0% to 105.0%. As shown in Table 4, the predicted ETV-H contents in commercial tablets were all within the above range.

4. Conclusions

Raman spectroscopy with 1064 nm excitation wavelength was successfully employed as an analytical tool for the nondestructive, and no sample preparation required determination of low content ETV-H in Baraclude tablets. By applying different preprocessing methods (baseline correction, SNV, MSC, and Savitzky–Golay first and second derivatives), PLS quantitative models were built to predict the concentration of ETV-H in Baraclude tablets. The calibration samples were divided into two sets, and the best performing model was chosen. Then, the best chemometrics model using all the calibration samples to build the model and predict the ETV-H content in the test set samples. It showed a good result with an R^2 of 0.973, RMSEC of 0.05%, and RMSEP of 0.03% on the spectral region 1350–1700 cm^{-1} with baseline processing and SNV as preprocessing methods for the raw data. The LOD of the best performance model was 0.17% w/w%. The predicted ETV-H contents in Baraclude tablets were all in the range defined in US Pharmacopeia.

In addition to the backscattering mode, Raman technology also had other two modes, transmission mode Raman and spatially offset Raman spectroscopy. Transmission mode Raman spectroscopy had the ability to penetrate the entire tablet to obtain the whole information which included API concentration and the spectra of all ingredients. Spatially offset Raman spectroscopy had the ability to obtain deep feature information inside the sample through nontransparent packaging or surface. In comparison, backscattering Raman could only obtain information on the sample surface, which would be insufficient when analyzing low-content API tablets. However, compared with the backscattering Raman, spatially offset Raman and transmission Raman spectroscopy both require larger laser

power, which might affect the measured sample. So, this study chose backscattering Raman spectroscopy for the quantification of low-content API tablets. The results showed that 1064 nm Raman spectroscopy had the capability to predict the low-content API in pharmaceutical tablets in the market.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Yanlei Kang and Yushan Zhou contributed equally to this work.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant nos. 2016YFC0800900, 2016YFC0800905, 2016YFC0800905-Z03, and 2016YFC0800902); the Key Research and Development Program Projects of Zhejiang Province (grant no. 2018C03G2011156); and the Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (grant no. ICT1806).

References

- [1] H. Cho, H. Ahn, D. H. Lee et al., “Entecavir and tenofovir reduce hepatitis B virus-related hepatocellular carcinoma recurrence more effectively than other antivirals,” *Journal of Viral Hepatitis*, p. 25, 2018.
- [2] G. V. Papatheodoridis, V. Sypsa, G. Dalekos et al., “8-year survival in chronic hepatitis B patients under long-term entecavir or tenofovir therapy is similar to the general population,” *Search Results*, vol. 68, 2018.
- [3] J. Vial, M. Cohen, P. Sassi, and D. Thiébaud, “Pharmaceutical quality of docetaxel generics versus originator drug product: a comparative analysis,” *Current Medical Research and Opinion*, vol. 24, no. 7, pp. 2019–2033, 2008.
- [4] N. L. Calvo, R. M. Maggio, and T. S. Kaufman, “Characterization of pharmaceutically relevant materials at the solid state employing chemometrics methods,” *Journal of Pharmaceutical and Biomedical Analysis*, vol. 147, pp. 538–564, 2018.

- [5] C. Ortega-Zúñiga, C. P.-D. la Rosa, A. D. Román-Ospino, A. Serrano-Vargas, R. J. Romañach, and R. Méndez, "Development of near infrared spectroscopic calibration models for in-line determination of low drug concentration, bulk density, and relative specific void volume within a feed frame," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 164, pp. 211–222, 2019.
- [6] N. L. Calvo, R. M. Maggio, and T. S. Kaufman, "Characterization of pharmaceutically relevant materials at the solid state employing chemometrics methods," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 147, pp. 538–564, 2018.
- [7] FDA, *Powder Blend and Finished Dosage Form Units Stratified in Process Dosage Unit Sampling and Assessment, Draft Guidance for Industry*, FDA, White Oak, MD, USA, 2003.
- [8] M. Kuentz, R. Holm, and D. P. Elder, "Methodology of oral formulation selection in the pharmaceutical industry," *European Journal of Pharmaceutical Sciences*, vol. 87, pp. 136–163, 2016.
- [9] N. M. El-Abasawi, K. A. M. Attia, A. A. M. Abo-serie, and A. Abdel-Fattah, "Simultaneous determination of rosuvastatin and propranolol by HPLC in pharmaceutical dosage form," *Analytical Chemistry Letters*, vol. 8, no. 1, pp. 46–53, 2018.
- [10] U. Holzgrabe, "Quantitative NMR spectroscopy in pharmaceutical applications," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 57, no. 2, pp. 229–240, 2010.
- [11] J. J. A. van Kampen, P. C. Burgers, R. de Groot, and T. M. Luider, "Qualitative and quantitative analysis of pharmaceutical compounds by MALDI-TOF mass spectrometry," *Analytical Chemistry*, vol. 78, no. 15, pp. 5403–5411, 2006.
- [12] Y. Gut, M. Boiret, L. Bultel et al., "Application of chemometric algorithms to MALDI mass spectrometry imaging of pharmaceutical tablets," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 105, pp. 91–100, 2015.
- [13] Ł. Komsta and J. K. Maurin, "Recognition of active ingredients in tablets by chemometric processing of X-ray diffractometric data," *Talanta*, vol. 82, no. 2, pp. 850–853, 2010.
- [14] H. Terada, Y. Hattori, T. Sasaki, and M. Otsuka, "Quantitation of trace amorphous solifenacin succinate in pharmaceutical formulations by transmission Raman spectroscopy," *International Journal of Pharmaceutics*, vol. 565, pp. 325–332, 2019.
- [15] T. Casian, S. Iurian, A. Gavan et al., "Near Infra-Red spectroscopy for content uniformity of powder blends - focus on calibration set development, orthogonality transfer and robustness testing," *Talanta*, vol. 188, pp. 404–416, 2018.
- [16] T. Casian, A. Reznik, A. L. Vonica-Gligor, J. Van Renterghem, T. De Beer, and I. Tomuța, "Development, validation and comparison of near infrared and Raman spectroscopic methods for fast characterization of tablets with amlodipine and valsartan," *Talanta*, vol. 167, pp. 333–343, 2017.
- [17] Y. V. Zontov, K. S. Balyklova, A. V. Titova, O. Y. Rodionova, and A. L. Pomerantsev, "Chemometric aided NIR portable instrument for rapid assessment of medicine quality," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 131, pp. 87–93, 2016.
- [18] D. Andrews, K. Geentjens, B. Igne et al., "Analytical method development using transmission Raman spectroscopy for pharmaceutical assays and compliance with regulatory guidelines—part I: transmission Raman spectroscopy and method development," *Journal of Pharmaceutical Innovation*, vol. 13, no. 2, pp. 121–132, 2018.
- [19] Z. D. Harms, Z. Shi, R. A. Kulkarni, and D. P. Myers, "Characterization of near-infrared and Raman spectroscopy for in-line monitoring of a low-drug load formulation in a continuous manufacturing process," *Analytical Chemistry*, vol. 91, no. 13, pp. 8045–8053, 2019.
- [20] K. A. Esmonde-White, M. Cuellar, C. Uerpmann, B. Lenain, and I. R. Lewis, "Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing," *Analytical and Bioanalytical Chemistry*, vol. 409, no. 3, pp. 637–649, 2017.
- [21] J. A. Griffen, A. W. Owen, J. Burley, V. Taresco, and P. Matousek, "Rapid quantification of low level polymorph content in a solid dose form using transmission Raman spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 128, pp. 35–45, 2016.
- [22] B. Li, A. Calvet, Y. Casamayou-Boucau, C. Morris, and A. G. Ryder, "Low-content quantification in powders using Raman spectroscopy: a facile chemometric approach to sub 0.1% limits of detection," *Analytical Chemistry*, vol. 87, no. 6, pp. 3419–3428, 2015.
- [23] S. Assi, R. A. Watt, and A. C. Moffat, "On the quantification of ciprofloxacin in proprietary Ciproxin tablets and generic ciprofloxacin tablets using handheld Raman spectroscopy," *Journal of Raman Spectroscopy*, vol. 43, no. 8, pp. 1049–1057, 2012.
- [24] D. Riolo, A. Piazza, C. Cottini et al., "Raman spectroscopy as a PAT for pharmaceutical blending: advantages and disadvantages," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 149, pp. 329–334, 2018.
- [25] L. Makraduli, P. Makreski, K. Goracinova, S. Stefov, M. Anevskaja, and N. Geskovski, "A comparative approach to screen the capability of Raman and infrared (mid- and near-) spectroscopy for quantification of low-active pharmaceutical ingredient content solid dosage forms: the case of alprazolam," *Applied Spectroscopy*, vol. 74, no. 6, pp. 661–673, 2020.
- [26] M. G. Orkoula and C. G. Kontoyannis, "Non-destructive quantitative analysis of risperidone in film-coated tablets," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 47, no. 3, pp. 631–635, 2008.
- [27] C. Muehlethaler, J. R. Lombardi, and M. Leona, "Sample self-absorption in surface-enhanced Raman spectroscopy (SERS): influence of the resonance, dilution and depth of the measurements," *Journal of Raman Spectroscopy*, vol. 48, no. 5, pp. 647–652, 2017.
- [28] M. A. d. S. Farias, F. L. F. Soares, and R. L. Carneiro, "Crystalline phase transition of ezetimibe in final product, after packing, promoted by the humidity of excipients: monitoring and quantification by Raman spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 121, pp. 209–214, 2016.
- [29] M. Korang-Yeboah, Z. Rahman, D. A. Shah, and M. A. Khan, "Spectroscopic-based chemometric models for quantifying low levels of solid-state transitions in extended release theophylline formulations," *Journal of Pharmaceutical Sciences*, vol. 105, no. 1, pp. 97–105, 2016.
- [30] M. C. Hennigan and A. G. Ryder, "Quantitative polymorph contaminant analysis in tablets using Raman and near infrared spectroscopies," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 72, pp. 163–171, 2013.
- [31] P. T. Mah, S. J. Fraser, M. E. Reish, T. Rades, K. C. Gordon, and C. J. Strachan, "Use of low-frequency Raman spectroscopy and chemometrics for the quantification of crystallinity in amorphous griseofulvin tablets," *Vibrational Spectroscopy*, vol. 77, pp. 10–16, 2015.

- [32] H. Vakili, H. Wickström, D. Desai, M. Preis, and N. Sandler, "Application of a handheld NIR spectrometer in prediction of drug content in inkjet printed orodispersible formulations containing prednisolone and levothyroxine," *International Journal of Pharmaceutics*, vol. 524, no. 1-2, pp. 414–423, 2017.
- [33] C. De Bleye, P.-F. Chavez, J. Mantanus et al., "Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 69, pp. 125–132, 2012.
- [34] I. Tomuta, R. Iovanov, E. Bodoki, and L. Vonica, "Development and validation of NIR-chemometric methods for chemical and pharmaceutical characterization of meloxicam tablets," *Drug Development and Industrial Pharmacy*, vol. 40, no. 4, pp. 549–559, 2014.