processes

MDPI

Article

Development of a Calibration Model for Real-Time Solute Concentration Monitoring during Crystallization of Ceritinib Using Raman Spectroscopy and In-Line Process Microscopy

Matea Gavran [1], Željka Ujević Andrijić [1,*], Nenad Bolf [1], Nikola Rimac [1], Josip Sacher [1] and Damir Šahnić [2]

[1] Department of Measurements and Process Control, Faculty of Chemical Engineering and Technology, University of Zagreb, 1000 Zagreb, Croatia
[2] PLIVA Croatia Ltd., Teva Pharmaceutical Industries Ltd., 10000 Zagreb, Croatia
* Correspondence: zujevic@fkit.unizg.hr

Abstract: Raman spectroscopy is a useful tool for polymorphic form-monitoring during the crystallization process. However, its application to solute concentration estimation in two-phase systems like crystallization is rare, as the Raman signal is influenced by various changing factors in the crystallization process. The development of a robust calibration model that covers all variations is complex and represents a major challenge for the implementation of Raman spectroscopy for in-line monitoring and control of the solution crystallization process. This paper describes the development of a Raman-based calibration model for estimating the solute concentration of the active pharmaceutical ingredient ceritinib. Several different calibration approaches were tested, which included both temperature and spectra of clear solutions and slurries/suspensions. It was found that the concentration of the ceritinib solution could not be accurately predicted when suspended crystals were present. To overcome this challenge, the approach was enhanced by including additional variables related to crystal size and solid concentration obtained via in-line process microscopy (chord-length distribution percentiles D10, D50 and D90) and turbidity. Partial least squares regression (PLSR) and artificial neural network (ANN) models were developed and compared based on root mean square error (RMSE). ANN models estimated the solute concentration with high accuracy, with the prediction error not exceeding 1% of the nominal solute concentration.

Keywords: process analytical technology; Raman spectroscopy; turbidity measurements; in-line process microscopy; crystallization; ceritinib; calibration models; pharmaceuticals

Citation: Gavran, M.; Ujević Andrijić, Ž.; Bolf, N.; Rimac, N.; Sacher, J.; Šahnić, D. Development of a Calibration Model for Real-Time Solute Concentration Monitoring during Crystallization of Ceritinib Using Raman Spectroscopy and In-Line Process Microscopy. *Processes* **2023**, *11*, 3439. https://doi.org/10.3390/pr11123439

Academic Editor: Iqbal M. Mujtaba

Received: 29 November 2023
Revised: 12 December 2023
Accepted: 13 December 2023
Published: 16 December 2023

1. Introduction

Crystallization is a fundamental process in the pharmaceutical industry for the purification and isolation of substances. As the vast majority of active pharmaceutical ingredients (APIs) are produced in crystalline form, crystallization plays a crucial role in achieving the desired chemical purity and physical properties of the product [1]. When done properly, it facilitates precise control of crystal properties like the polymorphic form, particle size distribution (PSD), aggregation and/or agglomeration. These properties in turn directly impact downstream processes like filtration, drying and particle reduction unit operations, as well as final product characteristics like solubility, stability, dissolution rate and bioavailability [2,3].

Given the strict manufacturing process quality control and the requirements for high-quality pharmaceutical products, in 2004, the U.S. Food and Drug Administration (FDA) proposed an innovative approach to the research and development, manufacturing and quality control of pharmaceutical products that utilizes process analytical technology (PAT) [4]. PAT combines scientific and engineering approaches through the development and implementation of analytical methods, advanced techniques and tools for the measurement and control in real time of critical quality attributes (CQA) and critical process

parameters (CPP). For the monitoring and subsequent control of crystallization processes, in-line PAT probes have been developed to measure and determine the concentration, particle size distribution, particle shape and polymorphic form directly in the process.

Accurate solute concentration measurements are essential to define process conditions for effective process control strategies, as the driving force of the process is supersaturation, which is defined as the difference between the current solute concentration and the equilibrium solute concentration [5–7]. Zhang et al. [8] provide a detailed overview of different techniques for concentration measurement in crystallization processes. The most common in situ method for solute concentration determination is the ATR technique in combination with FTIR and UV/Vis spectroscopy. This is due to the small penetration depth of the beam into the sample (i.e., solution or suspension), which makes the ATR technique insensitive to the presence of particles, i.e., the heterogeneity of the crystallization system [9–11]. On the other hand, Raman spectroscopy is more commonly used for monitoring the composition of the solid phase [12–14] in terms of polymorphic form determination or transformation monitoring. Due to the inherent properties of the Raman probe designs (focal point probes and coherent beam probes), the Raman signal is far more complex, as it depends on many factors, including the composition of the solid and liquid phases, the size and shape of the suspended crystals and the temperature [15,16]. Due to the dynamic nature of the crystallization process, during which all these variables change, the development of a suitable calibration model facilitated by Raman spectra is challenging. Nonetheless, the quantitative use of Raman spectroscopy for the measurement of solute and solid concentration has been reported. The concentration of a solute can be estimated based on a specific solute peak [17,18]. However, the univariate approach is not always feasible, as it cannot account for peak shifts and signal overlapping. Cornel et al. [15] demonstrated that despite weak and overlapping liquid- and solid-phase signals, solute concentration can be estimated from data collected from suspensions considering multivariate approaches. Several authors propose the collection of Raman spectra across a variety of temperatures in solid-free experiments, and in suspensions with varying concentrations of solids for solid-phase experiments [16,19,20]. However, the effects of temperature on Raman spectra can be considered negligible compared to the size and quantity of the suspended crystals, thus making this calibration step quite extensive. Various modeling approaches have been employed, including classical least squares (CLS), multiple linear regression (MLR), principal component regression (PCR), partial least square regression (PLSR) and artificial neural networks (ANN), for solute and solid concentration estimation. Table 1 gives an overview of the literature on quantitative applications of Raman spectroscopy for real-time measurements in multiphase systems over the past decade.

Table 1. Recent literature on the application of in-line Raman spectroscopy for multiphase systems.

Application	Model System	Description	Modeling Approach	Ref.
Solid concentration	Nuclear waste slurries	Suspended solids in dense nuclear waste slurries were estimated using Raman spectra. The concentration of solutes could not be accurately predicted from Raman spectra at high solids concentration.	PLSR	[21]
Solute concentration	Taurine	The kinetic parameters for nucleation and growth were estimated by combining the solid-phase data from the image analysis results with the liquid-phase concentration data from Raman spectra and temperature.	PLSR	[22]

Table 1. *Cont.*

Application	Model System	Description	Modeling Approach	Ref.
Solute and solid concentration	L-glutamic acid and paracetamol	The solution concentration and slurry density for two model–systems were determined simultaneously and quantitatively based on multivariate models. ANNs predicted the solute concentration with high accuracy.	CPR, PCR, PLSR, ANN	[20]
Solid composition and concentration	Carvedilol	Raman spectroscopy was used to evaluate the composition of the solid phase. In combination with ATR/UV-Vis to measure the solute concentration, the polymorphic concentration of each polymorph was calculated.	MLR, CLS	[23]
Solid concentration	Paracetamol	Online Raman spectroscopy was used to monitor the concentration of form II crystals as a function of time.	PCR	[24]
Solute concentration	Carbamazepine	The concentrations of undersaturated solutions of carbamazepine were estimated based on Raman spectra and temperature.	univariate	[18]
Solute concentration	Paracetamol	The solute concentration during the continuous crystallization of paracetamol with additives was monitored using Raman spectroscopy.	PLSR	[19]
Solute concentration	D-mannitol	The liquid-phase concentration in an aqueous polymorphic system of D-mannitol was estimated and monitored during the solvent-mediated polymorphic transformation from the α- to the β-form of mannitol.	PLSR	[16]
Solute concentration and polymorphic ratio	*o*-aminobenzoic acid	Raman spectroscopy was used to accurately determine the solute concentration in a solids-free experiment. Raman spectroscopy was used to measure the polymorphic ratio of forms I and II.	PLSR, PCR	[25,26]
Solid concentration ratio	Carvedilol	Based on the ratio of the solids concentrations of forms I and II measured using Raman spectroscopy, a feedback control of the crystallization of the desired form II was performed.	CLS	[27]

Recent literature on crystallization shows that Raman spectroscopy has been promoted to one of the most effective quantitative and qualitative in-line spectroscopic measurement methods for crystallization process monitoring and control [28,29]. This stems from its versatility in determining both solid- and solute-phase composition and concentration.

In this work, a novel calibration approach was investigated. The substantial calibration steps were reduced to measure suspensions of various solid concentrations within the metastable zone, as no solid-free solutions exist in a seeded crystallization process. The model system for the developed calibration approach was ceritinib form A in tetrahydrofuran. Ceritinib is an active pharmaceutical ingredient that targets metastatic non-small lung cancer cells by inhibiting the anaplastic lymphoma kinase (ALK) protein. The recrystalliza-

tion process of ceritinib in tetrahydrofuran is likely to optimize the crystal size distribution (CSD) of ceritinib, contributing to its improved granulometric properties. However, it is important to note that the detailed exploration of CSD improvement through recrystallization is not the scope of this work. Instead, this study concentrates on a novel calibration model approach to effectively monitor this recrystallization process. The dataset for model development consisted of Raman spectra, temperature and variables measured via in-line process microscopy that are related to crystal size (chord-length distribution percentiles D10, D50, D90) and solid concentrations (turbidity) measured in suspensions of different solute and solids concentrations.

2. Materials and Methods

In this work, the model system for the proposed calibration approach is the active pharmaceutical ingredient ceritinib of polymorphic form A in tetrahydrofuran. The IUPAC name of the active pharmaceutical ingredient ceritinib is 5-chloro-2-N-(5-methyl-4-piperidin-4-yl-2-propan-2-yloxyphenyl)-4-N-(2-propan-2-ylsulfonylphenyl)pyrimidine-2,4-diamine [30]. The molecular structure of the compound is shown in Figure 1. Ceritinib has three known polymorphic forms. Of these three solid forms, form A and form C are anhydrous, and form B is a hydrate [31].

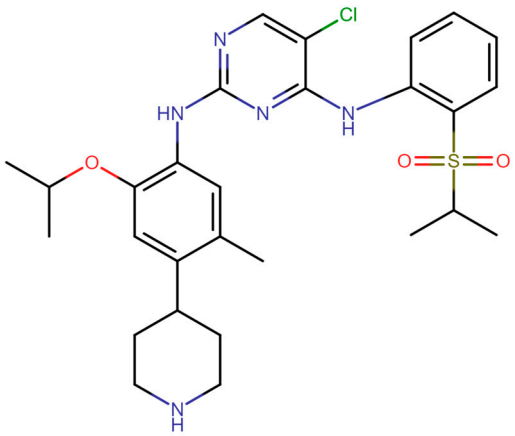


Figure 1. The molecular structure of ceritinib.

All solvents were purchased from Lachner (Neratovice, Czech Republic). Ceritinib form A was synthesized following the preparation steps described by Zokić et al. [32]. Ceritinib dihydrochloride (Hui Chem Co., Ltd., Shanghai, China) was dissolved in an acetone–water solvent mixture, followed by pH modification with sodium hydroxide and cooling crystallization. Crystals were dried under a vacuum.

Tetrahydrofuran was chosen as a suitable solvent for the purification and granulometric improvement of the synthesized ceritinib. The proposed calibration approach was developed as the recrystallization process of form A ceritinib in tetrahydrofuran was to be monitored and controlled based on the calibration model.

2.1. Offline Characterization Methods

The offline characterization of the prepared crystals was carried out using X-ray powder diffraction (XRPD). The samples were recorded using D8 Advance (Bruker, Billerica, MA, USA) with Cu Kα radiation at an accelerating voltage of 40 kV and a current of 25 mA in a Bragg–Brentano configuration in the range of 2–55° 2Θ, with a step of 0.02° and a step duration of 0.6 s.

The distinction of the characteristic Raman peaks of solid ceritinib form A and tetrahydrofuran was recorded offline using a fiber-optic probe (MarqMetrix Inc., Seattle, WA, USA) connected to a Raman spectrometer (WP 785 nm, Wasatch Photonics, Logan, UT, USA).

Raman spectra were collected in the range 2015–241 cm $^{-1}$ with a resolution of 2 cm^{-1}. The exposure time was altered to increase the signal-to-noise ratio for each sample.

2.2. Solubility Determination

The solubility determination of ceritinib form A in tetrahydrofuran was carried out in a CrystalSCAN (E2153, h.e.l Ltd., Borehamwood, UK) batch reactor with turbidity and temperature probes. Tetrahydrofuran was kept at a constant temperature, ranging from 5–55 °C. A defined, small mass of form A ceritinib crystals was added until a consistent turbidity plateau was reached, indicating that no further dissolution was occurring. An aliquot was withdrawn from the reactor, filtered and the solubility was gravimetrically determined using a moisture analyzer (MLS 50-3C, Kern & Sohn, Balingen, Germany).

A total of 10 solubility data points were measured as a function of temperature, and the solubility curve was fitted using the solubility regress design module in Dynochem software (version 2.2.). Within the solubility regress design module, several solubility expressions were proposed. The solubility expression:

$$c^* = \exp\left(lnA - B/(R \times T) - C/(R \times T)^2\right). \quad (1)$$

was selected based on the highest coefficient of determination, R^2. The solubility is measured in g/kg$_{solution}$, T is the temperature in K and R is the ideal gas constant in kJ/mol K. A, B and C are solubility model parameters.

2.3. Calibration Experiments

All experiments were carried out in a shaded 500 mL jacketed reactor coupled with a cooling/heating circulator (Magio MS-1000F, Julabo, Seelbach, Germany), a Pt-100 temperature sensor, an overhead stirrer and a PTFE propeller with three blades inclined at 45° (Bohlender, Grünsfeld, Germany).

The Raman spectrum was measured in-line, using the same Raman spectrometer and immersion probe as mentioned in the section as previously mentioned. The laser power was set to 450 mW, and a consistent exposure time of 1500 ms was used for all sample measurements, ensuring an optimal signal-to-noise ratio and avoiding saturation of the CCD detector. Chord-length distribution percentiles and turbidity were recorded using an in-line process microscope (Blaze 900, BlazeMetrics, Marysville, WA, USA). The agitation speed and laser strength were kept constant throughout the experiments, as they could cause a substantial amount of variation in the predicted solute concentration [18].

The Raman spectrometer, in-line process microscope and cooling/heating circulator were integrated through data-linking, enabling synchronous data acquisition (Figure 2). Dedicated software was developed for the systematic archiving of all measurement data using the OPC UA protocol and a standard USB.

Tetrahydrofuran (THF) was heated to the desired saturation temperature, ranging from 9 to 50 °C. The calculated mass of ceritinib form A was added to the tetrahydrofuran at the saturation temperature. The suspension was heated slightly above the saturation temperature until all crystals dissolved, which was observed using an in-line process microscope. When complete dissolution occurred, the solution was cooled back to saturation temperature. Suspensions with different solid concentrations were prepared by adding the known mass of form A ceritinib crystals to the saturated solutions (Figure 3). The amount of crystals needed for the preparation of the suspensions was calculated based on the solubility curve. For lower concentrations, a greater amount of crystals was added, considering the crystallization yield.

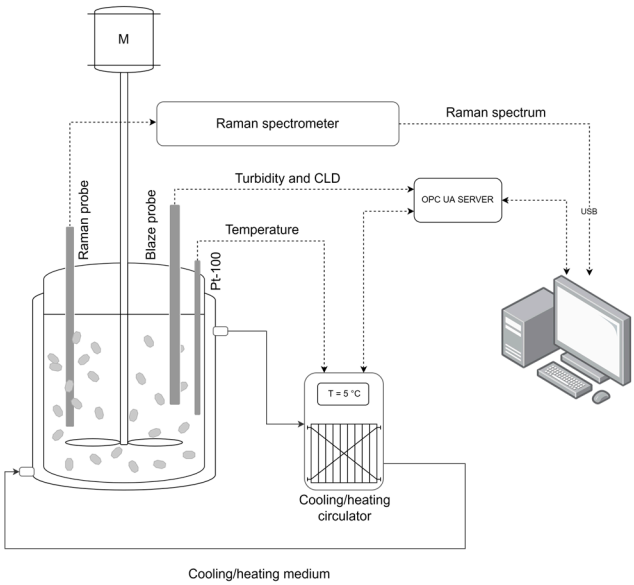


Figure 2. Experimental setup.

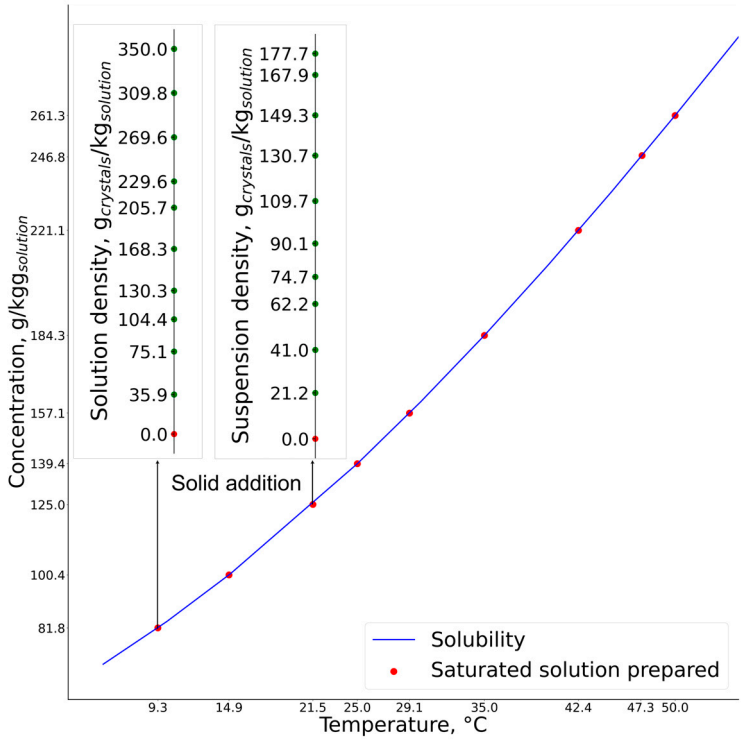


Figure 3. Calibration experiment for two distinct saturated solutions of different concentrations and their prepared suspensions.

In total, nine different saturated solutions were prepared, and for each distinct saturated solution up to ten suspensions of different solid concentrations were prepared. The Raman spectrum, temperature, turbidity and chord-length distribution (CLD) percentiles (D10, D50 and D90) were collected simultaneously after each addition of crystals to the saturated solution.

2.4. Model Development

Solute concentration estimation based on spectroscopy is carried out indirectly via the interpretation and quantification of the spectral data using calibration models. The

development of calibration models involves data collection, analysis and visualization, the removal of outliers, signal preprocessing, variable selection or reduction, choosing the type of model and training and validation of the model. As stated in the previous section, the collected experimental data consisted of the Raman spectrum, temperature, turbidity and chord-length distribution for a given suspension with varying solute ($g/kg_{\ solution}$) and solid concentrations ($g_{crystals}/kg_{\ solution}$).

All computations were performed in the Python programming language using the Sklearn and Keras libraries for the development of partial least squares regression (PLSR) and artificial neural network (ANN) models, respectively.

The Raman spectrum is prone to baseline shifts, scattering effects and noise, which affect its interpretability. In this work, several preprocessing methods were applied to the spectral data. The aim was to identify and apply preprocessing algorithms that could improve the model's performance. Baseline effects were corrected using asymmetric least squares smoothing (ALS) and a first-order Savitzky–Golay derivative. The Savitzky–Golay filter was also used for smoothing and noise reduction, with careful parameter selection to retain features of the spectra characteristic of ceritinib. Scattering correction was performed using standard normal variate (SNV) and multiplicative scatter correction (MSC). In addition, the spectral range for the model's development was reduced to 1800–400 cm^{-1} after a visual inspection and analysis of the offline and in-line collected data. The application of various preprocessing algorithms resulted in the generation of an additional four datasets.

To ensure the robustness of the model across various solute concentrations, a leave-one-out cross-validation adapted to the solute concentration was used. A distinct subset of the data was represented by the unique solute concentration value. As the collected data contained nine subsets (i.e., nine different solute concentrations), the model was trained on eight subsets and validated on one. This process was repeated iteratively for each subset to be used as a validation set, and the average root mean square error of cross-validation (RMSECV) was calculated.

The partial least squares regression model was used to estimate the solute and solid concentrations. When estimating multiple outputs, the partial least squares regression is referred to as PLSR2. The algorithm calculates the PLS components that explain the maximum covariance between the inputs and the outputs. The number of PLS components was optimized using the adapted leave-one-out cross-validation based on the RMSECV and the coefficient of determination (R^2) for both outputs simultaneously. The final PLSR model was fitted to the entire dataset.

Additionally, feedforward artificial neural networks with backpropagation algorithms were developed. The neural network's input layer initially consisted of 705 variables: Raman spectrum, temperature, turbidity and chord-length distribution percentiles. The output layer contained two neurons with linear activation functions. An optimal ANN architecture and parameters were determined through an extensive analysis, in which different numbers of neurons, hidden layers, activation functions and optimizers were investigated. To minimize the RMSE in training, the Adam optimization algorithm was utilized, along with dropout layers to avoid overfitting.

3. Results and Discussion

Ceritinib can exhibit three different crystalline forms. Ceritinib form A has characteristic X-ray Powder Diffraction (XRPD) peaks at angles 10.63°, 12.78°, 13.25°, 15.60° and 17.58°, which are absent in both forms B and C. No peaks characteristic of forms B and C at the angles 5.05°, 5.42°, 9.37°, 9.61°, 10.09°, 15.04° and 15.11° were observed [31,32]. The results of the XRPD analysis verify that the crystallization of ceritinib dihydrochloride in acetone–water yielded ceritinib form A (Figure 4). Hence, the prepared ceritinib was used for calibration experiments in tetrahydrofuran. The studied ceritinib form A is the most stable form; therefore, under the given conditions, form A is the most likely to form crystals and remain unchanged. The recrystallization of ceritinib in tetrahydrofuran did

not change the crystal structure, as the peaks of the crystals complied with the standard. This demonstrates that the use of tetrahydrofuran as a solvent in the recrystallization of ceritinib is effective for purification and can also aid to achieve an improved crystal size distribution after the initial crystallization in an acetone–water solvent mixture.

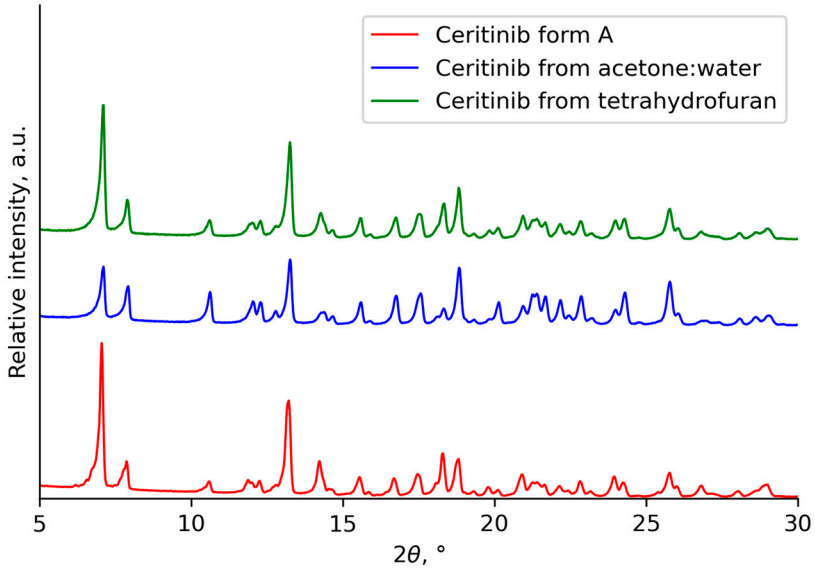


Figure 4. XRPD patterns of the prepared and recrystallized ceritinib.

The offline Raman spectrum of ceritinib form A and tetrahydrofuran shown in Figure 5 helped identify the spectral region for model development. Tetrahydrofuran had broad bands in the ranges 1500–1400 cm^{-1}, 1300–1200 cm^{-1}, 1050–1000 cm^{-1} and 950–850 cm^{-1}. The peaks of ceritinib form A did not overlap with tetrahydrofuran in the spectral range of 1650–1550 cm^{-1} and at the lower wavenumbers.

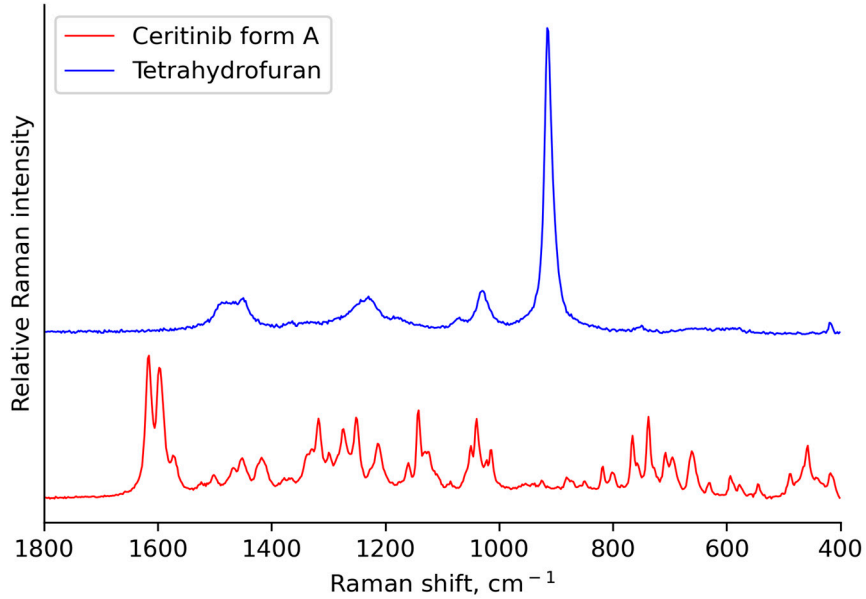


Figure 5. Offline Raman spectrum of ceritinib form A and tetrahydrofuran.

3.1. Solubility Determination

The determination of the solubility curve is a prerequisite for crystallization process development, as it defines the crystallization method (e.g., cooling, anti-solvent) and operating conditions and determines the yield [33]. Thus, for the proposed calibration

method, an accurate solubility curve is crucial to avoid the dissolution of crystals at a given temperature.

Table 2 contains the fitted solubility parameters. Solubility model 5 (Figure 6) displayed the highest correlation coefficient (R^2 = 0.995) among the other proposed Dynochem models when fitting the experimental data points, with an estimated average percent error of the measurements of 3.81%.

Table 2. Estimated solubility equation parameters.

ln*A*	*B*, kJ/mol	*C*, (kJ/mol)2
3.02	−31.74	66.89

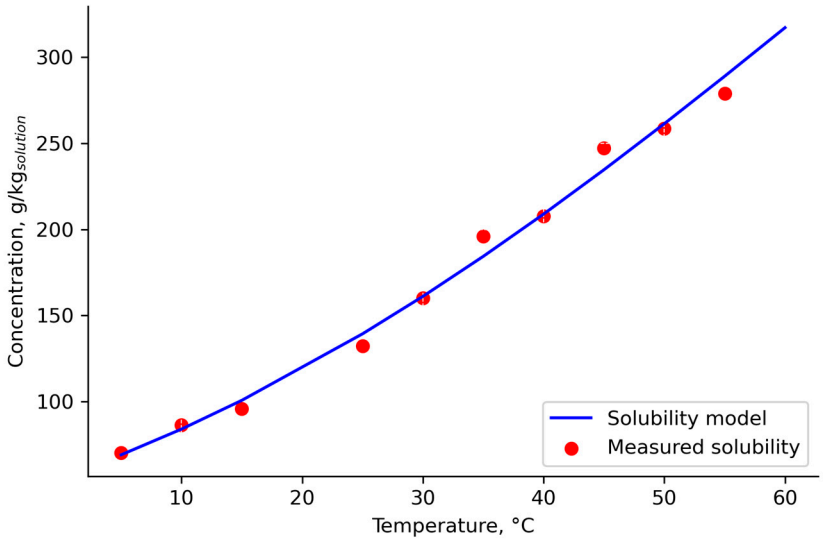


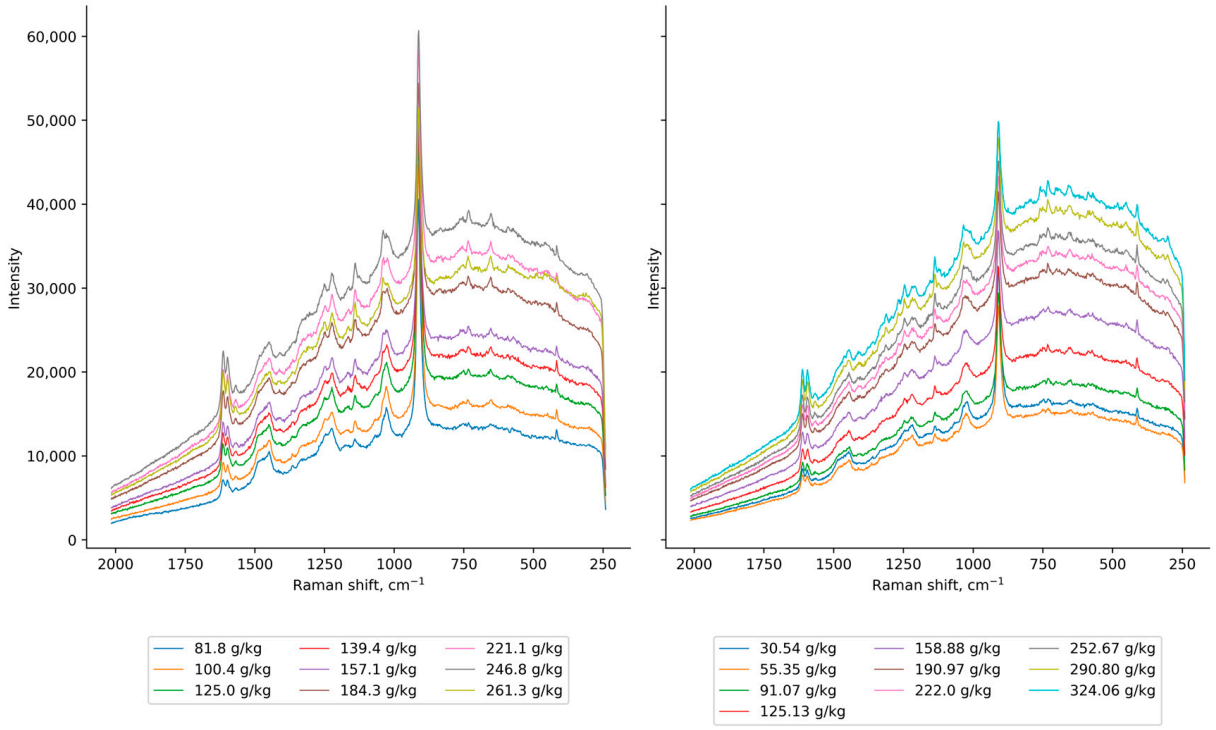
Figure 6. The estimated solubility curve and experimentally determined data.

Performing calibration experiments from saturated solutions confirmed the accuracy of the solubility model, as no dissolution of crystals was observed upon the addition of solids (indicated by the turbidity measurement).

3.2. Calibration Experiments and Raman Spectra Analysis

The visual inspection of the Raman spectra is a crucial step in the development of the calibration model. It provides useful information about the characteristic spectral features and variability of the data. It also helps to detect underlying data artifacts that influence Raman spectra, such as baseline shift, fluorescence background, noise and possible cosmic spikes, that need to be preprocessed before model development [34,35]. The influence of the changes in solute and solid concentrations on the Raman spectra is shown in Figure 7.

An increase in the intensity of the Raman spectra was observed for the increase in the solute concentration, although three of the most concentrated solutions did not follow the trend (Figure 7a). With the addition of crystals, an increase in the relative intensity of the characteristic ceritinib peaks to the tetrahydrofuran peaks was observed (Figure 7b). With the increase in both solute and solid concentrations, the suppression of tetrahydrofuran wide bands (1530–1410 cm^{-1}, 1140 cm^{-1} and 970–870 cm^{-1}) revealed additional ceritinib peaks. This also indicated that the solid and solute peaks overlap. To address this issue, turbidity was introduced as an input variable for both training and prediction, as it correlates strongly with solid concentration [36,37]. The calculated Pearson correlation coefficient supported this calibration approach, with a correlation of 0.91 indicating that turbidity should indirectly provide solid concentration information to the model. Additionally, as turbidity depends on particle size [38], the percentiles of the chord-length distribution were chosen as input variables.



(**a**) (**b**)

Figure 7. The raw Raman spectra of (**a**) saturated solutions with different solute concentrations and (**b**) saturated solution (c_{sat} = 100.4 g/kg $_{solution}$) with various solid concentrations.

3.3. Model Development

The preprocessing of spectral data is essential for noise removal and the reduction and elimination of variability in the data that is not related to the property of interest, all to enable the examined spectra to be further modeled more effectively [39]. Careful selection of the spectral preprocessing algorithm can improve the robustness and quality of the final model.

The raw Raman spectra (Figure 7) are noisy and show a baseline shift. Therefore, asymmetric least squares (ALS), a first-derivative Savitzky–Golay filter, standard normal variate (SNV) and multiplicative scatter correction (MSC) were applied to correct these effects (Figure 8). However, it is important to establish a balance between the various preprocessing steps to avoid overfitting and the potential risk of distorting the original data to the extent of losing essential information from the spectrum. In addition, when setting the parameters for preprocessing, particular attention was taken to ensure that the prominent peak of ceritinib form A was preserved, especially in the spectral range of 1650–1550 cm^{-1}. This consideration was crucial to ensure that this key spectral feature was not lost or distorted, maintaining the integrity and relevance of the spectral data for accurate analysis. All the preprocessing techniques are listed in Table 3.

The preprocessing steps aim to increase the linearity of the data; this is particularly important for PLSR models, as it facilitates their ability to more effectively identify and exploit linear correlations within the data, thus enabling a better solute concentration prediction. The application of standard normal variate (SNV), Savitzky–Golay smoothing and both asymmetric least squares (ALS) and first-derivative baseline correction reduced the variation in the data caused by light scattering effects [40]. These preprocessing steps significantly enhanced the linearity of the data, namely, of the P2 and P3 datasets. As a result, the prediction performance of the PLSR models was improved compared to the unprocessed data. On the other hand, for the developed ANN models, none of the preprocessing methods significantly improved their prediction performance (results shown in Table 3).

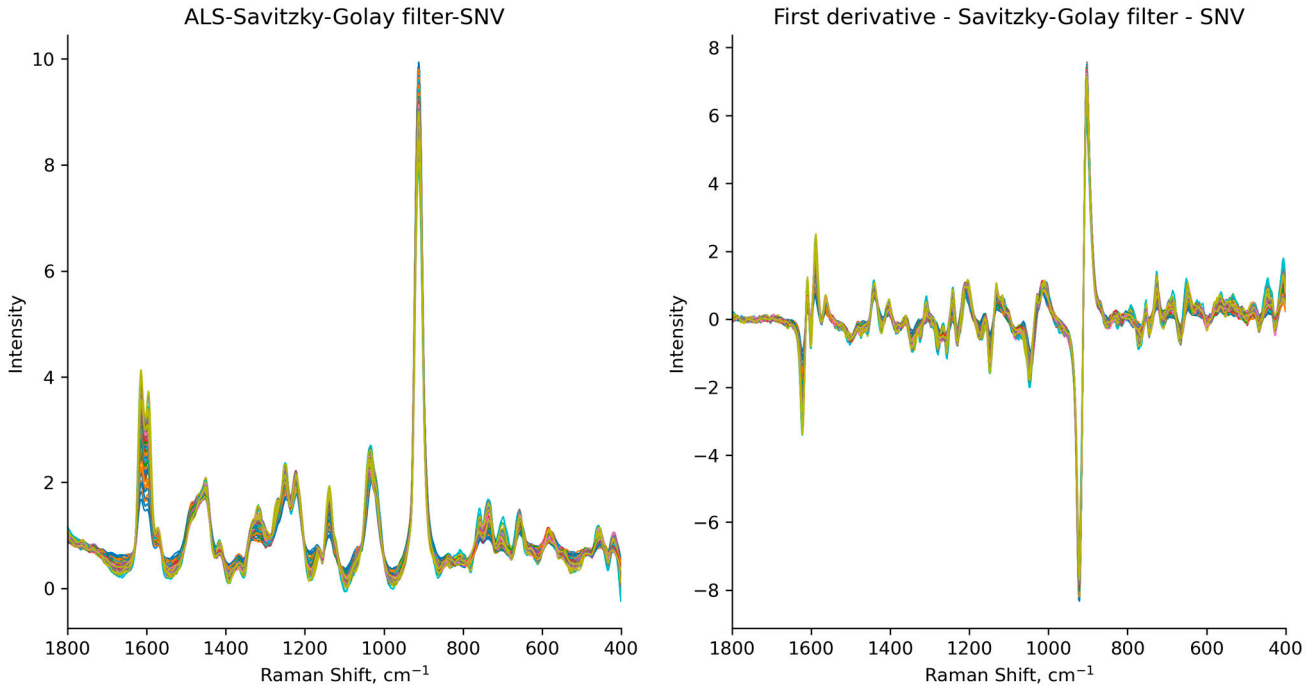


Figure 8. Spectra of all samples preprocessed using: (**a**) ALS algorithm, Savitzky–Golay filter and SNV, i.e., datasets P2 and A2; (**b**) first-derivative Savitzky–Golay smoothing and SNV, i.e., datasets P3 and A3.

Table 3. Preprocessing techniques and RMSECV.

Model	Dataset	Baseline Correction	Noise Reduction	Scatter Correction	RMSECV g/kg	
					Solute	Solids
PLSR	P1		none		18.97	20.99
	P2	ALS	Savitzky–Golay filter	SNV	9.82	20.99
	P3	1st derivative		SNV	10.22	21.46
	P4	1st derivative		MSC	22.49	22.29
	P5	ALS		MSC	23.33	22.29
ANN	A1		none		0.79	48.55
	A2	ALS	Savitzky–Golay filter	SNV	0.60	36.64
	A3	1st derivative		SNV	0.61	40.20
	A4	1st derivative		MSC	0.48	20.93
	A5	ALS		MSC	0.49	22.12

ALS—asymmetric least squares smoothing algorithm; ANN—artificial neural network; A1–A5 denote the developed artificial neural network models; MSC—multiplicative scatter correction; PLSR—partial least squares regression; P1–P5 denote the developed partial least squares regression models; SNV—standard normal variate.

The PLSR model's performance was evaluated by varying the number of PLS components from 1 to 10 and observing the changes in R^2 and RMSECV for each of the response variables. This approach allowed the selection of an optimal number of components based on a balance between model complexity and prediction accuracy. To reduce the possibility of overfitting, for the dataset P2, a PLSR model with three PLS components was used to fit the data, as additional components did not increase predictive performance (Figure 9).

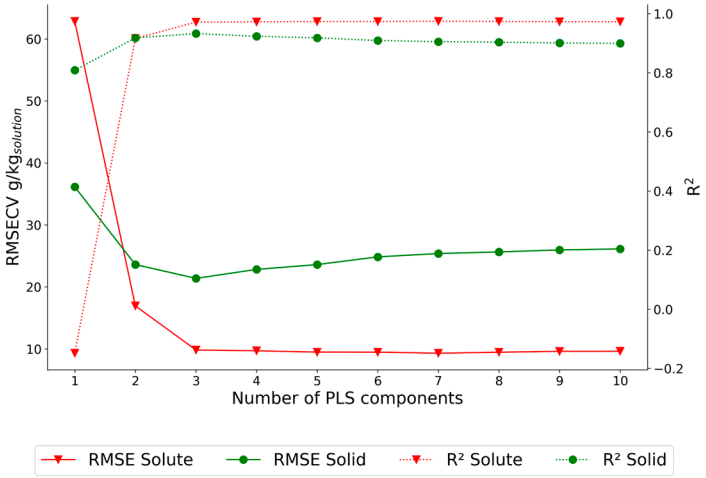
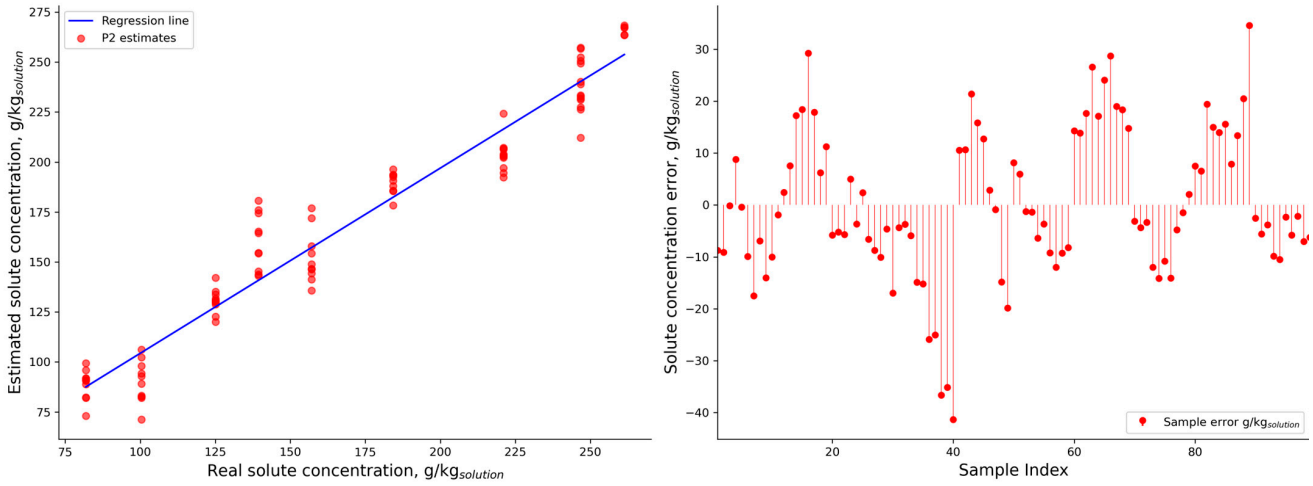


Figure 9. PLSR model optimization for the dataset P2.

The optimized PLSR P2 model overestimates and underestimates solute concentration (Figure 10) and the errors are not equally distributed, which indicates a non-linear relationship in the dataset.



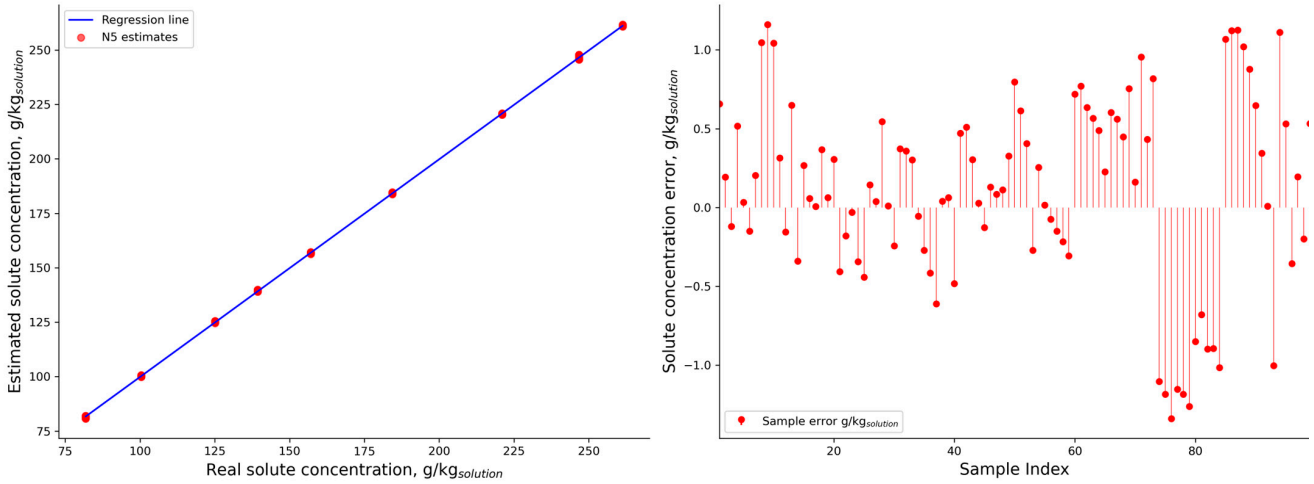
(**a**) (**b**)

Figure 10. P2 (**a**) results on the validation dataset and its (**b**) residuals.

Due to the complex and non-linear relationship in the data that PLSR models could not deconvolve, ANN models were developed, as discussed by Lin et al. [20]. In multivariate modeling, linear models are generally employed for systems with linear behavior, while non-linear models are applied to systems with non-linear relationships. Linear models are often preferred for linear systems because they are reliable, simple and result from basic physico-chemical principles such as the Beer–Lambert law.

Through comprehensive analysis involving various numbers of hidden layers, neurons, activation functions and optimizers, the optimal neural network architecture was established. The most effective ANN configuration included a first hidden layer with 705 neurons, followed by a second hidden layer comprising 100 neurons with a sigmoid activation function. The model's performance was evaluated through leave-one-out cross-validation across nine distinct solute concentration subsets, ensuring robustness and accurate predictions.

The average RMSECV did not exceed 2 g/kg$_{solution}$, which is less than 1% of the lowest solution concentration used in the model development (Figure 11).



(**a**) (**b**)

Figure 11. NN A5 (**a**) results on the validation dataset and its (**b**) residuals.

4. Conclusions

This research highlights the challenges of accurately predicting solute concentration using Raman spectroscopy in the presence of suspended crystals, considering the complex non-linear relationship of the solute concentration and the Raman spectrum in a multiphase crystallization system. This complexity is a major challenge and limitation to the application of Raman spectroscopy for in-line monitoring and control of the solution crystallization process. By integrating additional variables that are related to the crystal size and solid concentration using an advanced data-driven approach, a calibration model for solute concentration was developed. It was demonstrated that artificial neural networks (ANNs) estimated the solute concentration with higher accuracy compared to partial least squares regression (PLSR). The prediction error did not exceed 1% of the nominal solution concentration.

This research demonstrates the practical applicability of the proposed calibration approach for real-time monitoring and control of crystallization processes. The results indicate that this approach could also be suitable for other multiphase systems.

Author Contributions: Conceptualization, M.G., Ž.U.A., N.B., N.R. and D.Š.; methodology, M.G. and N.R.; software, M.G., N.R. and J.S.; validation, M.G. and N.R.; formal analysis, M.G.; investigation, M.G.; writing—original draft preparation, M.G.; writing—review and editing, M.G., Ž.U.A., N.B. and D.Š.; visualization, M.G.; supervision, Ž.U.A., N.B. and D.Š. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Structural and Investment Funds, grant number KK.01.1.1.07.0017 (CrystAPC—Crystallization Advanced Process Control).

Data Availability Statement: Data available on request.

Acknowledgments: We would like to express our sincere gratitude to our colleague Marko Sejdić for the technical support and colleagues Katarina Mužina and Arijeta Bafti from the Dept. of Inorganic Chemical Technology and Non-Metals.

Conflicts of Interest: Damir Šahnić was employed by the company PLIVA Croatia Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Orehek, J.; Teslić, D.; Likozar, B. Continuous Crystallization Processes in Pharmaceutical Manufacturing: A Review. *Org. Process Res. Dev.* **2021**, *25*, 16–42. [CrossRef]
2. Gao, Y.; Zhang, T.; Ma, Y.; Xue, F.; Gao, Z.; Hou, B.; Gong, J. Application of Pat-based Feedback Control Approaches in Pharmaceutical Crystallization. *Crystals* **2021**, *11*, 221. [CrossRef]

3. Perini, G.; Salvatori, F.; Ochsenbein, D.R.; Mazzotti, M.; Vetter, T. Filterability Prediction of Needle-like Crystals Based on Particle Size and Shape Distribution Data. *Sep. Purif. Technol.* **2019**, *211*, 768–781. [CrossRef]
4. U.S. Department of Health and Human Services. *Food and Drug Administration Guidance for Industry: PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance*; U.S. Food and Drug Administration: Silver Spring, MD, USA, 2004.
5. Nagy, Z.K.; Braatz, R.D. Advances and New Directions in Crystallization Control. *Annu. Rev. Chem. Biomol. Eng.* **2012**, *3*, 55–75. [CrossRef] [PubMed]
6. Herceg, T.; Ujević Andrijić, Ž.; Gavran, M.; Sacher, J.; Vrban, I.; Bolf, N. Primjena Neuronskih Mreža Za Procjenu Koncentracije Otopine Ksilometazolin Hidroklorida u N-Butanolu Primjenom ATR-FTIR Spektroskopije in Situ. *Kem. Ind.* **2023**, *72*, 639–650. [CrossRef]
7. Sacher, J.; Sejdić, M.; Gavran, M.; Bolf, N.; Ujević Andrijić, Ž. Proračun Optimalnog Temperaturnog Profila Hlađenja Šaržnog Kristalizatora. *Kem. Ind.* **2023**, *72*, 443–453. [CrossRef]
8. Zhang, F.; Du, K.; Guo, L.; Huo, Y.; He, K.; Shan, B. Progress, Problems, and Potential of Technology for Measuring Solution Concentration in Crystallization Processes. *Meas. J. Int. Meas. Confed.* **2022**, *187*, 110328. [CrossRef]
9. Lewiner, F.; Klein, J.P.; Puel, F.; Févotte, G. On-Line ATR FTIR Measurement of Supersaturation during Solution Cystallization Processes. Calibration and Applications on Three Solute/Solvent Systems. *Chem. Eng. Sci.* **2001**, *56*, 2069–2084. [CrossRef]
10. Zhang, F.; Liu, T.; Wang, X.Z.; Liu, J.; Jiang, X. Comparative Study on ATR-FTIR Calibration Models for Monitoring Solution Concentration in Cooling Crystallization. *J. Cryst. Growth* **2017**, *459*, 50–55. [CrossRef]
11. Billot, P.; Couty, M.; Hosek, P. Application of ATR-UV Spectroscopy for Monitoring the Crystallisation of UV Absorbing and Nonabsorbing Molecules. *Org. Process Res. Dev.* **2010**, *14*, 511–523. [CrossRef]
12. Qu, H.; Alatalo, H.; Hatakka, H.; Kohonen, J.; Louhi-Kultanen, M.; Reinikainen, S.P.; Kallas, J. Raman and ATR FTIR Spectroscopy in Reactive Crystallization: Simultaneous Monitoring of Solute Concentration and Polymorphic State of the Crystals. *J. Cryst. Growth* **2009**, *311*, 3466–3475. [CrossRef]
13. Simone, E.; Saleemi, A.N.; Nagy, Z.K. Application of Quantitative Raman Spectroscopy for the Monitoring of Polymorphic Transformation in Crystallization Processes Using a Good Calibration Practice Procedure. *Chem. Eng. Res. Des.* **2014**, *92*, 594–611. [CrossRef]
14. Schöll, J.; Bonalumi, D.; Vicum, L.; Mazzotti, M.; Müller, M. In Situ Monitoring and Modeling of the Solvent-Mediated Polymorphic Transformation of L-Glutamic Acid. *Cryst. Growth Des.* **2006**, *6*, 881–891. [CrossRef]
15. Cornel, J.; Lindenberg, C.; Mazzotti, M. Quantitative Application of in Situ ATR-FTIR and Raman Spectroscopy in Crystallization Processes. *Ind. Eng. Chem. Res.* **2008**, *47*, 4870–4882. [CrossRef]
16. Su, W.; Li, C.; Hao, H.; Whelan, J.; Barrett, M.; Glennon, B. Monitoring the Liquid Phase Concentration by Raman Spectroscopy in a Polymorphic System. *J. Raman Spectrosc.* **2015**, *46*, 1150–1156. [CrossRef]
17. Hu, Y.; Liang, J.K.; Myerson, A.S.; Taylor, L.S. Crystallization Monitoring by Raman Spectroscopy: Simultaneous Measurement of Desupersaturation Profile and Polymorphic Form in Flufenamic Acid Systems. *Ind. Eng. Chem. Res.* **2005**, *44*, 1233–1240. [CrossRef]
18. Acevedo, D.; Yang, X.; Mohammad, A.; Pavurala, N.; Wu, W.L.; O'Connor, T.F.; Nagy, Z.K.; Cruz, C.N. Raman Spectroscopy for Monitoring the Continuous Crystallization of Carbamazepine. *Org. Process Res. Dev.* **2018**, *22*, 156–165. [CrossRef]
19. Powell, K.A.; Saleemi, A.N.; Rielly, C.D.; Nagy, Z.K. Monitoring Continuous Crystallization of Paracetamol in the Presence of an Additive Using an Integrated PAT Array and Multivariate Methods. *Org. Process Res. Dev.* **2016**, *20*, 626–636. [CrossRef]
20. Lin, M.; Wu, Y.; Rohani, S. Simultaneous Measurement of Solution Concentration and Slurry Density by Raman Spectroscopy with Artificial Neural Network. *Cryst. Growth Des.* **2020**, *20*, 1752–1759. [CrossRef]
21. Prasad, R.; Crouse, S.H.; Rousseau, R.W.; Grover, M.A. Quantifying Dense Multicomponent Slurries with In-Line ATR-FTIR and Raman Spectroscopies: A Hanford Case Study. *Ind. Eng. Chem. Res.* **2023**, *62*, 15962–15973. [CrossRef]
22. Wu, Y.; Gao, Z.; Rohani, S. Deep Learning-Based Oriented Object Detection for in Situ Image Monitoring and Analysis: A Process Analytical Technology (PAT) Application for Taurine Crystallization. *Chem. Eng. Res. Des.* **2021**, *170*, 444–455. [CrossRef]
23. Tacsi, K.; Gyürkés, M.; Csontos, I.; Farkas, A.; Borbás, E.; Nagy, Z.K.; Marosi, G.; Pataki, H. Polymorphic Concentration Control for Crystallization Using Raman and Attenuated Total Reflectance Ultraviolet Visible Spectroscopy. *Cryst. Growth Des.* **2020**, *20*, 73–86. [CrossRef]
24. Nicoud, L.; Licordari, F.; Myerson, A.S. Polymorph Control in Batch Seeded Crystallizers. A Case Study with Paracetamol. *CrystEngComm* **2019**, *21*, 2105–2118. [CrossRef]
25. Simone, E.; Saleemi, A.N.; Nagy, Z.K. Raman, UV, NIR, and Mid-IR Spectroscopy with Focused Beam Reflectance Measurement in Monitoring Polymorphic Transformations. *Chem. Eng. Technol.* **2014**, *37*, 1305–1313. [CrossRef]
26. Simone, E.; Saleemi, A.N.; Tonnon, N.; Nagy, Z.K. Active Polymorphic Feedback Control of Crystallization Processes Using a Combined Raman and ATR-UV/Vis Spectroscopy Approach. *Cryst. Growth Des.* **2014**, *14*, 1839–1850. [CrossRef]
27. Pataki, H.; Csontos, I.; Nagy, Z.K.; Vajna, B.; Molnar, M.; Katona, L.; Marosi, G. Implementation of Raman Signal Feedback to Perform Controlled Crystallization of Carvedilol. *Org. Process Res. Dev.* **2013**, *17*, 493–499. [CrossRef]
28. Kim, E.J.; Kim, J.H.; Kim, M.S.; Jeong, S.H.; Choi, D.H. Process Analytical Technology Tools for Monitoring Pharmaceutical Unit Operations: A Control Strategy for Continuous Process Verification. *Pharmaceutics* **2021**, *13*, 919. [CrossRef]

29. Cote, A.; Erdemir, D.; Girard, K.P.; Green, D.A.; Lovette, M.A.; Sirota, E.; Nere, N.K. Perspectives on the Current State, Challenges, and Opportunities in Pharmaceutical Crystallization Process Development. *Cryst. Growth Des.* **2020**, *20*, 7568–7581. [CrossRef]
30. Burns, M.W.; Kim, E.S. Profile of Ceritinib in the Treatment of ALK+ Metastatic Non-Small-Cell Lung Cancer. *Lung Cancer Targets Ther.* **2015**, *6*, 35–42. [CrossRef]
31. Chennuru, R.; Koya, R.T.; Kommavarapu, P.; Narasayya, S.V.; Muthudoss, P.; Vishweshwar, P.; Babu, R.R.C.; Mahapatra, S. In Situ Metastable Form: A Route for the Generation of Hydrate and Anhydrous Forms of Ceritinib. *Cryst. Growth Des.* **2017**, *17*, 6341–6352. [CrossRef]
32. Zokić, I.; Prlić Kardum, J. Crystallization Behavior of Ceritinib: Characterization and Optimization Strategies. *ChemEngineering* **2023**, *7*, 84. [CrossRef]
33. Tung, H.H.; Paul, E.L.; Midler, M.; McCauley, J.A. *Crystallization of Organic Compounds: An Industrial Perspective*; Pearson: London, UK, 2008; ISBN 9780471467809.
34. Rinnan, Å. Pre-Processing in Vibrational Spectroscopy-When, Why and How. *Anal. Methods* **2014**, *6*, 7124–7129. [CrossRef]
35. Bocklitz, T.; Walter, A.; Hartmann, K.; Rösch, P.; Popp, J. How to Pre-Process Raman Spectra for Reliable and Stable Models? *Anal. Chim. Acta* **2011**, *704*, 47–56. [CrossRef] [PubMed]
36. Rabesiaka, M.; Porte, C.; Bonnin-Paris, J.; Havet, J.L. An Automatic Method for the Determination of Saturation Curve and Metastable Zone Width of Lysine Monohydrochloride. *J. Cryst. Growth* **2011**, *332*, 75–80. [CrossRef]
37. Raphael, M.; Rohani, S. On-Line Estimation of Solids Concentrations and Mean Particle Size Using a Turbidimetry Method. *Powder Technol.* **1996**, *89*, 157–163. [CrossRef]
38. Crawley, G.; Cournil, M.; Benedetto, D. Di Size Analysis of Fine Particle Suspensions by Spectral Turbidimetry: Potential and Limits. *Elements* **1997**, *91*, 197–208.
39. Rinnan, Å.; Nørgaard, L.; van den Berg, F.; Thygesen, J.; Bro, R.; Engelsen, S.B. Chapter 2—Data Pre-Processing. *Infrared Spectrosc. Food Qual. Anal. Control.* **2009**, *3*, 29–50.
40. Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M.C. Breaking with Trends in Pre-Processing? *TrAC Trends Anal. Chem.* **2013**, *50*, 96–106. [CrossRef]