Author(s): Lipiäinen, Tiina; Pessi, Jenni; Movahedi, Parisa; Koivistoinen, Juha; Kurki, Lauri; Tenhunen, Mari; Yliruusi, Jouko; Juppo, Anne M.; Heikkonen, Jukka; Pahikkala, Tapio; Strachan, Clare J.

Title: Time-Gated Raman Spectroscopy for Quantitative Determination of Solid-State Forms of Fluorescent Pharmaceuticals

Year: 2018

Version:

Time-Gated Raman Spectroscopy for Quantitative Determination of Solid-State Forms of Fluorescent Pharmaceuticals

Tiina Lipiäinen, Jenni Pessi, Parisa Movahedi, Juha Tapio Koivistoinen, Lauri Kurki, Mari Tenhunen, Jouko Yliruusi, Anne M. Juppo, Jukka Heikkonen, Tapio Pahikkala, and Clare J. Strachan
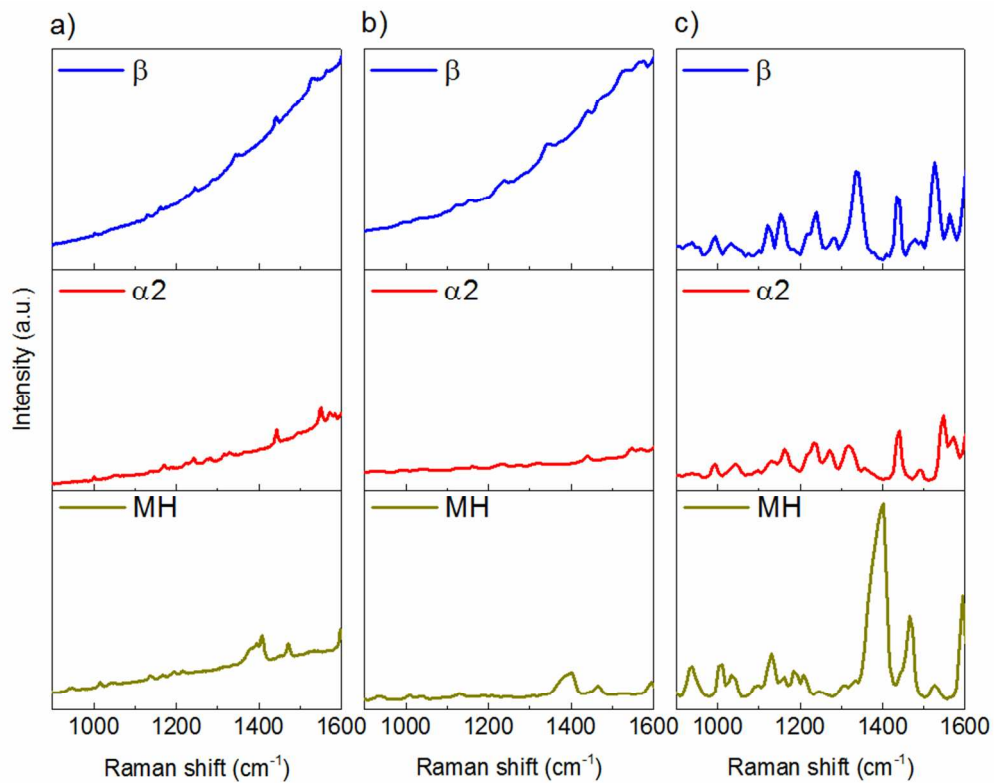
Just Accepted

Figure 4. Raman spectra obtained with a) the CW Raman setup, b) the time-gated Raman instrument, presented as sum spectra from 0-5.5 ns and c) the time-gated Raman instrument, presented as spectra after fluorescence rejection. The Raman intensity scale is the same for each solid-state form, but different for each of the three columns for clarity.

84x66mm (300 x 300 DPI)

Time-Gated Raman Spectroscopy for Quantitative Determination of Solid-State Forms of Fluorescent Pharmaceuticals

Tiina Lipiäinen[†,‡], Jenni Pessi[†,‡], Parisa Movahedi[#], Juha Koivistoinen[⊥], Lauri Kurki[‖], Mari Tenhunen[‖], Jouko Yliruusi[†], Anne M. Juppo[†], Jukka Heikkonen[#], Tapio Pahikkala[#], Clare J. Strachan[†,*]

[†]Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, Viikinkaari 5 E, FI-00790 Helsinki, Finland

[#]Department of Future Technologies, University of Turku, Vesilinnantie 5, FI-20500 Turku, Finland

[⊥]Nanoscience Center, Department of Chemistry, P.O. Box 35, FI-40014, University of Jyväskylä, Finland

[‖]TimeGate Instruments, Teknologiantie 5, FI-90590 Oulu, Finland

ABSTRACT: Raman spectroscopy is widely used for quantitative pharmaceutical analysis, but a common obstacle to its use is sample fluorescence masking the Raman signal. Time-gating provides an instrument-based method for rejecting fluorescence through temporal resolution of the spectral signal, and allows Raman spectra of fluorescent materials to be obtained. An additional practical advantage is that analysis is possible in ambient lighting. This study assesses the efficacy of time-gated Raman spectroscopy for the quantitative measurement of fluorescent pharmaceuticals. Time-gated Raman spectroscopy with a $128 \times (2) \times 4$ CMOS SPAD detector was applied for quantitative analysis of ternary mixtures of solid-state forms of the model drug, piroxicam (PRX). Partial least squares (PLS) regression allowed quantification, with Raman-active time domain selection (based on visual inspection) improving performance. Model performance was further improved by using kernel-based regularized least squares (RLS) regression with greedy feature selection, in which the data use in both the Raman shift and time dimensions was statistically optimized. Overall, time-gated Raman spectroscopy, especially with optimized data analysis in both the spectral- and time-dimensions, shows potential for sensitive and relatively routine quantitative analysis of photoluminescent pharmaceuticals during drug development and manufacturing.

Most (90%) active pharmaceutical ingredients (API) crystallize as solid particles.[1] Different inter- and intra-molecular bonding and conformations in solid-state forms of a substance, such as polymorphs, amorphous solids, salts, and solvates, result in different physicochemical properties.[2,3] Dissolution rate, solubility, stability, and bioavailability, among other properties, depend on the solid-state structure of the substance. This poses challenges to the pharmaceutical industry in terms of material characterization, formulation, processing and end product quality control, and has therapeutic, legal and commercial implications.[4]

Effective methods for evaluating the possible changes in solid state structure during research and development, manufacturing, and storing are needed.[5,6] Raman spectroscopy is an established method for qualitative and quantitative analysis of APIs exhibiting different solid-state forms and often enables rapid, non-destructive measurements with no sample preparation needed.[7-9] The spectra can be measured through container walls, blisters, plastic bags, and in an aqueous environment since Raman spectroscopy has low sensitivity for water.[10] The form of the sample is also flexible; powders, slurries, pellets, emulsions, films are all suitable for Raman spectroscopy. These properties make Raman spectroscopy well-suited for diverse real-time process monitoring applications.

Raman spectra are obtained by measuring the intensity distribution of Raman scattered photons from a monochromatic light source as a function of wavelength.[10,11] Quantitative determination is based on the concentration of the substance of interest being proportional to the integrated intensity of its characteristic Raman bands.[12] Overlapping peaks of different compounds in a mixture and experimental effects that are not related to sample concentration complicate the analysis.[13] In such cases multivariate analysis, where a large amount of spectral data can be included, is more reliable than methods where only one or a few spectral features are considered. Several multivariate methods have been established for the interpretation of Raman spectra.[14,15] The aims of such methods are to: i) extract spectral information that quantifies substances of interest, ii) estimate the uncertainties of the quantification, and iii) evaluate the performance of the built model.[14] Partial least squares (PLS) regression is one of the most widely used chemometric methods for quantitative analysis.[16] PLS relates the information in two data matrices, X (e.g. the spectral variation) and Y (e.g. the sample composition), in a multivariate model by maximizing their covariance.[17] Kernel-based regularized least-squares (kernel-based RLS) regression is another approach that has the ability to learn functions from the nonlinear data features which, when combined with feature selection algorithms, such as greedy forward feature selection, optimizes the use of information provided by the data features.[18,19] PLS and RLS are quite similar in

1

that they aim to shrink the solution away from the ordinary least squares solution toward the directions of the variable space of large sample spread with lower variability.[20]

Error sources in the quantitative analysis of powder mixtures using Raman spectroscopy include variance in intra- and inter-day reproducibility of the Raman instrument, changes in room temperature and humidity, sample fluorescence, mixing, packing and positioning, as well as sample particle size and compactness.[21,22] While most issues can be addressed with suitable spectral processing and data analysis approaches, complete subtraction of fluorescence without any instrument-based methods is difficult, even with sophisticated algorithms.[10]

Complete or partial rejection of the fluorescence signal from the Raman signal is possible with various time-resolved techniques.[23] The ability to detect the arrival time and energy of each photon allows assessment of the lifetime of both the fluorescence and Raman signals. Due to the lifetime differences, rejecting the fluorescence background is possible (**Figure 1**). Time-gated devices employ short, intensive laser pulses and the sample response is recorded simultaneously with the pulses. This also means that analysis in ambient lighting is possible.[24]
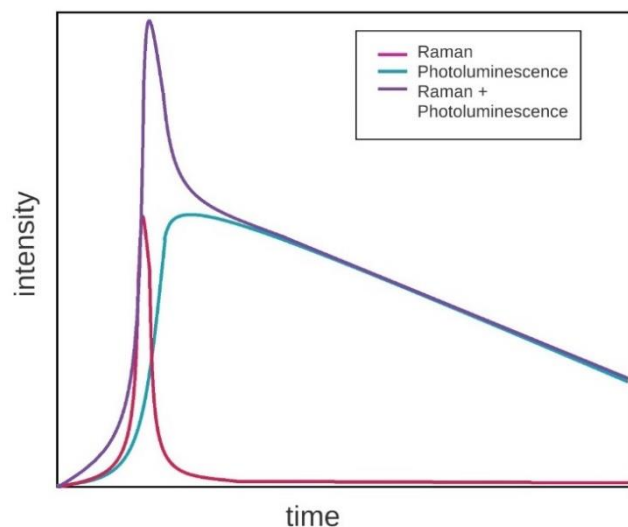

Figure 1. Relative lifetimes (not to scale) of Raman and photoluminescence (including fluorescence) signals (adapted from [25]).

Time-gating can be realized with various detection systems such as time-resolved photomultiplier tubes,[26,27] high-speed optical shutters based on a Kerr cells,[28,29] intensified charge-coupled devices,[30] quantum dot resonant tunneling diodes,[31] and complementary metal-oxide semiconductor single-photon avalanche diodes (CMOS SPADs).[24] One of the essential advantages of CMOS SPADs is the ability to reject both the photoluminescence tail and the photon noise.[32] SPADs are realized in standard CMOS technology and contain a pn junction which is reverse-biased above its breakdown voltage, meaning that entry of even a single photon can trigger avalanche breakdown that can then be recorded.[33-35] The width and position of the time gate need to be properly selected.[36]

The current CMOS SPADs are compact and inexpensive while being able to achieve adequate temporal resolutions (sub-nanosecond).[37-39] CMOS SPAD detectors have been used to evaluate fluorescence lifetimes.[40] More recently the applicability of CMOS SPADs for fluorescence rejection in Raman spectroscopy in pharmaceuticals has also been shown.[25,36,41]

The aim of this study was to investigate the potential of time-gated Raman spectroscopy for quantitative analysis of fluorescent pharmaceutical solids. A time-gated Raman setup using a fast CMOS SPAD detector[39] was employed for the first time for quantitative analysis of powder mixtures. This instrument allows the separation of the photoluminescence signal from the Raman signal in ambient lighting and enables stronger Raman signal generation compared to traditional instruments.[38,39] The data, with and without prior time-domain selection (based on visual inspection), was analyzed using PLS regression, the most well established multivariate quantitative spectral analysis method in pharmaceutics. Quantitative analysis was also performed using kernel-based RLS with greedy feature selection, which statistically optimized data use in both the spectral and time domains.

MATERIALS AND METHODS

Materials. Piroxicam (PRX) (Hawkins, USA), a non-steroidal anti-inflammatory drug, was the fluorescent model compound in this study. PRX has six reported polymorphs (β (I), α1 and α2 (both also referred as form II), III, IV, and V)) and one hydrated form (monohydrate, MH).[42-47] Ternary powder mixtures used in this study consisted of the most commonly observed forms: β, α2 and MH.

The PRX was purchased in form β and this form was used as received. PRX form α2 was prepared by recrystallization from a saturated solution in absolute ethanol.[46] PRX MH was prepared by recrystallization from saturated aqueous solution.[48] The aqueous solution was heated to 80 °C, the ethanol solution to 70 °C and the solutions were slowly cooled to room temperature before vacuum filtration.

Evaluating polymorph conversion. X-ray powder diffractometry (XRPD) analysis was performed using a Bruker D8 Advance diffractometer (Bruker, Germany) with a Cu Kα radiation source (λ = 1.5418 Å) over a 2θ range of 5° to 40°, using a step size of 0.01°, step time of 0.5 s, voltage of 40 kV, and current of 40 mA. The results were compared to the patterns in the Cambridge Structural Database (CSD). Fourier transform infrared spectroscopy (FTIR) measurements were performed with a Bruker Vertex 70 spectrometer (Bruker Optik, Germany) and an ATR accessory with a single reflection diamond crystal (MIRacle, Pike Technologies, Madison, WI, USA). The obtained spectra were the mean of 64 scans and have a spectral range from 650 to 4000 cm^{-1} with a resolution of 4 cm^{-1}. The ATR spectra were converted to absorbance spectra with OPUS software (v. 5.0, Bruker Optik, Ettlingen, Germany). Differential scanning calorimetry (DSC) was performed with a differential scanning calorimeter (DSC823e, Mettler Toledo AG) in sealed perforated aluminum pans under dry nitrogen purge (50 mL/min) at a heating rate of 10 °C/min from 30 to 210 °C. Particle size and morphology of the PRX solid-state forms were examined by scanning electron microscopy (SEM) with a Quanta™ 250 FEG (FEI Inc., U.S.). Samples for SEM were mounted on carbon-coated double-sided tape (Agar Scientific,

2

Germany) and sputter-coated with a 5 nm layer of platinum (Q150T Quomm, Turbo-Pumped Sputter Coater, China).

Mixture design. The powder mixtures were prepared according to a special cubic mixture design (**Figure 2**).[49] The mass ratio of each form was varied between 0, 1:6, 1:3, 2:3, and 1 in the mixtures, the (1:3, 1:3, 1:3) mixture was prepared in triplicate. A ternary mixture was preferred over a binary mixture since often more than two solid-state forms are potentially present in a process environment. The solid-state forms of PRX were carefully mixed using geometric dilution with a card to avoid inducing changes in the solid state.
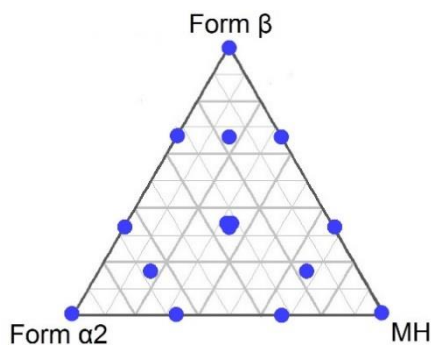

Figure 2. Mixture design employed in the experiments.

Time-gated Raman spectroscopy. Raman spectra of the mixtures of different solid-state forms of PRX were collected with a TimeGated® TG532 M1 Raman spectrometer (TimeGate Instruments Oy, Finland) coupled with a BWTek sampling probe with a focal spot size of approximately 85 µm (**Figure 3**). The Raman instrument was equipped with a picosecond pulsed laser, a CMOS SPAD array detector, and the sampling probe. The excitation source was a 532 nm Nd:YVO microchip pulsed laser, the average power used was 14 mW (2.235 mW after the probe), repetition rate 40 kHz, pulse width 150 ps, focus diameter 50 µm, pulse energy 0.35 µJ, peak power 2 kW, and maximum irradiance 28 MW cm^{-2}.

The detector was a $128 \times (2) \times 4$ CMOS SPAD matrix detector.[39] The internal time histogram of the detector consisted of four bins accumulating single-photon arrivals. Bin 3 provided the strongest Raman signal with the present setup (**Figure 3**). The signals collected with bin 3 were used for the data analysis. The time-resolved spectral datasets were collected by sequentially moving the gate in 50 ps steps using the electronic delay generator. Raman spectra with fluorescence rejection and time-resolved fluorescence spectra were acquired simultaneously. The spectra were obtained from the Raman shift range of 700 cm^{-1} to 1700 cm^{-1} up to 5.5 ns.

The measurements were conducted in triplicate, with continuous sample rotation, and the focal point was moved between each measurement to acquire a more representative signal over a larger area of the sample. The measurements were carried out at ambient temperature, lighting, and humidity. Cyclohexane was used as a reference standard to monitor wavenumber accuracy. Data acquisition and setup control were performed with the instrument software (TimeGated® Model 1).

Continuous wave (CW) Raman spectroscopy. Raman measurements were executed with a home-built Raman setup in a backscattering geometry using 532 nm excitation produced with a CW single frequency laser (Alphalas, Monolas-532-100-SM). The beam was focused onto the sample and subsequently collected with a 100 x microscope objective (Olympus 100x with 0.70 N.A.). The scattered light was dispersed in a 0.5 m imaging spectrograph (Acton, SpectraPro 2500i) using a 600 g/mm grating (resolution: ~5 - 6 cm^{-1}). The signal was detected with EMCCD camera (Andor Newton EM DU971N-BV) using 60 µm slit width. The Rayleigh scattering was attenuated with a notch-filter (Semrock). The sample positioning was performed with an XYZ-piezo scanner (Attocube, ANPxyz101) with the smallest step of 100 nm in each direction. The laser power was ~0.5 mW and two 5 s measurements were averaged for each accumulation.

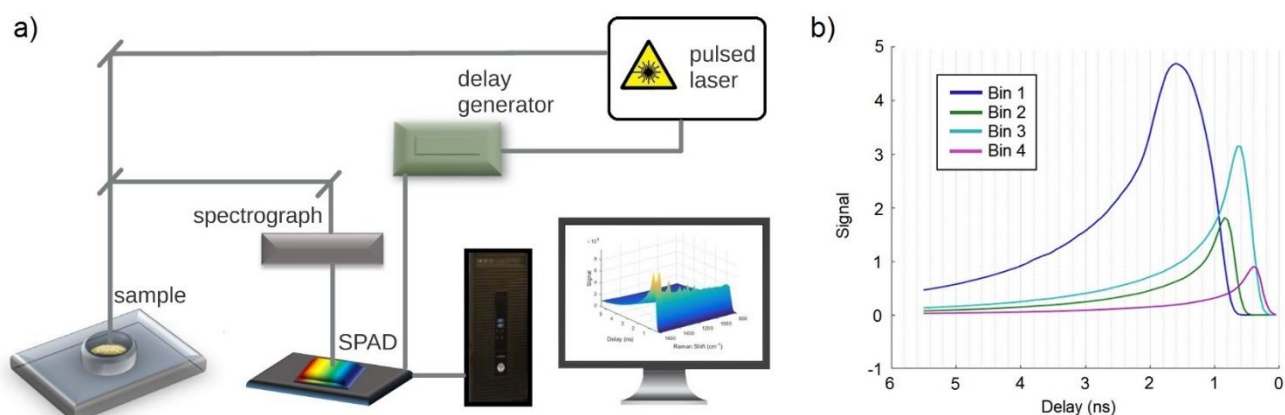**Figure 3.** a) Schematic of the time-gated Raman instrument used for obtaining the Raman spectra and performing fluorescence rejection and b) basis for bin 3 selection. The four bins collect the scattered photons with different delays and the intensity of the obtained signal varies. Bin 3 provided the strongest signal at the optimal time-frame for detection of Raman scattered photons for PRX.

3

Partial least squares (PLS). Part of the fluorescence was rejected from the signal by the time-gated detection system using the data obtained from bin 3 data. Residual photoluminescence (elevated baseline) signal was removed using the software provided with the instrument (TimeGated® Model 1). The time-frame for analysis was selected manually based on visual appearance of the signal. The location of the Raman peaks in the time-domain was found to be at the delay of 0.4 - 0.8 ns. Baseline correction was performed using adaptive iteratively reweighted penalized least squares (airPLS) and local minima fitting (Lmin) algorithms. Data from the whole time-domain without selecting a specific time-frame (0.0 - 5.5 ns) was processed identically with the selected time-frame data for comparison.

PLS is widely used for quantitative Raman spectral analysis of pharmaceutical samples. In general, PLS finds components known as latent factors in variable matrix **X** which best predict the response matrix **Y**. PLS regression searches for a set of factors that simultaneously decompose **X** and **Y** where these factors explain the covariance between the two matrices as much as possible.[50] The spectral data was standard normal variate (SNV) transformed[51] and mean centered (without scaling) prior to PLS analysis. SNV and mean-centering have been shown to be suitable algorithms for quantitative analyses of solid-state mixtures by vibrational spectroscopy.[21] PLS regression[52] for quantitative analysis was carried out with the NIPALS algorithm[17] using SIMCA-P software (v. 13.0.3, Umetrics AB, Sweden).

The performance of the model was evaluated using R^2X, R^2Y, and the root-mean-square error of cross-validation (RMSECV). RMSECV values were obtained with leave-one-out cross validation (LOOCV), with the leave-one-out procedure performed with all mixtures, except the pure forms (since there is no mixing error associated with the pure forms), where in each CV round all replicates of one mixture are left out. The reported RMSECV values are the average of the root-mean-square error of prediction (RMSEP) values which were obtained for the left out mixtures for each cross-validation round (Equation 1):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}} \qquad (1)$$

Here, $y-\hat{y}$ is the predicted residual for each mixture form of an observation.

Kernel-based regularized squares (RLS). Part of the fluorescence was rejected from the signal by the time-gated detection system using the data obtained from bin 3 data, as in the previous section. To further investigate the quantification potential of the 3D spectra in both the spectral and time dimensions, fast kernel-based RLS analysis with multi-target greedy feature selection was applied. All predictive models were trained with the Python-based machine learning software library RLScore [19]. RLS with a Gaussian kernel was built as the prediction model. Given a training set $\{(x_i, y_i)\}_{i=1}^{n}$ where the feature vector $x_i \in \mathbb{R}^p$ and the class labels $y_i \in \mathbb{R}^q$, the multivariate RLS formulation finds A such that (Equation 2):

$$A = arg \min_{A} \frac{1}{n}\|Y - KA\|_F^2 + \lambda\, tr(A^T KA) \qquad (2)$$

where A is the $n \times q$ weight matrix, Y is the $n \times q$ label matrix, $\|.\|_F$ is the Frobenius norm of a matrix, K is the $n \times n$ kernel matrix, λ is the regularization parameter and tr is the trace of a matrix. The following Gaussian kernel function was used in the models (**Equation 3**):

$$K(x_i, x) = exp\left(\frac{\|x_i - x\|^2}{2\sigma^2}\right) \qquad (3)$$

where $\|.\|$ is the ℓ_2 norm and σ is the kernel width parameter.

A kernel-based RLS model was obtained by carrying out the following procedure. A hyper-parameter combination consisting of the kernel width parameter, σ, the regularization parameter, λ, and the time interval for averaging with SNV and mean centering, was selected from a three-dimensional grid with LOOCV on a training set. In addition to the hyper-parameter values, a multi-target greedy RLS algorithm was built to select a predictive subset of Raman shifts.[18] Greedy RLS starts from the empty set, and on each iteration adds the feature (Raman shift) whose addition provides the best LOOCV performance. To avoid selection bias, the prediction performance of the obtained kernel-based RLS model was estimated with the standard nested cross-validation approach, in which the selection procedure described above was separately carried out during each round of an outer cross-validation, and the performance estimate was the average of the prediction errors of these models on the data withheld in the corresponding rounds of the outer cross-validation.[53]

In addition, to ensure that the performance estimate would reflect the real-world conditions under which the model is expected to be used, the fold-partition of the cross-validation was performed similar to PLS analysis as follows. A LOOCV was applied to the PRX mixtures, indicating that every replication of each mixture was simultaneously used as test data and the pure forms were not used for testing.

Given the input vector of a new measurement unseen during the training phase (left-out mixtures for testing), kernel-based RLS makes a prediction of its corresponding output vector. The real-value vectors (y = [predicted value of form β, predicted value of form α2, predicted value of MH]) predicted by the kernel-RLS model were post-processed as follows, with the ith entry of the vector, y, set as (**Equation 4**):

$$\frac{max(0, y_i)}{\sum_i max(0, y_i)} \qquad (4)$$

The purpose of this setting was to restrict the mixture proportions between zero and one and prevent impossible predictions. Later, **Equation 1** was used as described earlier to calculate the RMSECV values of each of the three solid-state forms.

RESULTS AND DISCUSSION

Polymorph conversion. XRPD, FTIR, and DSC analyses confirmed complete polymorph conversion of form β of PRX (CSD: BIYSEH13)[54] to form α2 (CSD: BIYSEH06)[46] and MH (CSD: CIDYAP02).[54] No solid-state impurities were detected. SEM images show clear morphological differences between the solid-state forms (**Figure S1** (Supplementary data)). Additionally, PCA of the Raman data also showed very clear differences

4

for all the mixtures with no overlap of the sample clusters observed.

Raman spectra and fluorescence rejection. Fluorescence, as indicated by the elevated baselines, was observed in both the CW Raman spectra and the time-gated spectra that were the sum of the raw signal recorded over the whole time-scale (0 - 5.5 ns) (**Figure 4a and 4b**). Form β fluoresced more strongly than form α2 and the MH. The baseline increased with increasing Raman shift for all three solid state forms.

Fluorescence rejection with the time-gated data (using bin 3, 0.4-0.8 ns time-frame, and residual airPLS and Lmin for baseline correction) resulted in 2D Raman spectra with fluorescence-free baselines (Figure 4c). The characteristic peaks of the solid-state forms of PRX match those previously published.[55] The vibrational modes for piroxicam have previously been predicted and assigned using density functional theory calculations.[56]
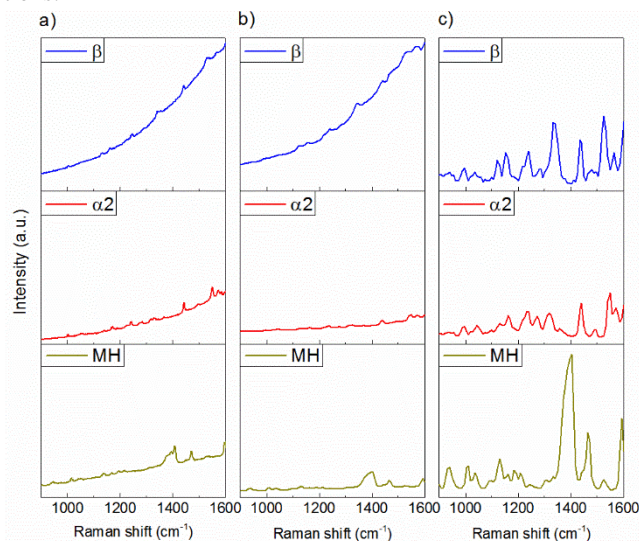

Figure 4. Raman spectra obtained with a) the CW Raman setup, b) the time-gated Raman instrument, presented as sum spectra from 0-5.5 ns and c) the time-gated Raman instrument, presented as spectra after fluorescence rejection. The Raman intensity scale is the same for each solid-state form, but different for each of the three columns for clarity.

The raw 3D spectra recorded with the time-gated instrument (bin 3 data), the subtracted 3D baseline spectra (representing the fluorescence) and the 3D Raman spectra after baseline rejection from PRX form β, form α2 and the MH are presented in **Figure 5**. The 3D data indicates the starting point of the Raman signal immediately after the laser pulse, as well as the fluorescence starting-point and the fluorescence tail. Consistent with the spectra in **Figure 4b**, the 3D spectra also suggest the three solid-state forms of PRX fluoresced to varying degrees over the

presented Raman shift range, with form β exhibiting the strongest baseline intensity maxima, as well as largest baseline profile change as a function of Raman shift. The 3D plots also reveal the changing baselines over time: a rapid initial increase (at all Raman shifts) is followed by a more gradual decay over several nanoseconds for all three forms. It is important to note that, since the presented data are from bin 3 only, the baseline signal cannot be expected to represent the total fluorescence signal over the presented time range, with detected signal intensity biased toward time delays close to the Raman-active time-frame. Bin selection for biased detection was appropriate in this case, since avoiding fluorescence through instrumental means for improved quantification was one of the aims of the study. Despite this, it is interesting to note that different baseline decay profiles are visible for the three different solid-state forms, supporting previous evidence that not only relative fluorescence intensity (as a function of Raman shift), but also the fluorescence signal lifetime profiles, can also be solid-state specific. Differences in such decay profiles have previously been observed using time-gated Raman spectroscopy with the amorphous and γ-crystalline forms of the drug indomethacin.[25]

After subtracting the detected baseline spectra from the raw spectra, very little fluorescence signal was observed and Raman peaks were clearly visible at time delays of less than 1 ns. Overall, the time-gated Raman instrument and with baseline processing enabled robust fluorescence rejection without any requirement for substance specific calibration or suppression methods. This provided a suitable basis for applying chemometric data analysis for quantitative solid-state determination.

PLS regression. The PLS regression used to quantify the mixtures on the basis of the associated Raman spectra using the 0.4 - 0.8 ns window was successful. Traditional PLS models with four PLS factors resulted in an $R^2X(cum)$ of 0.997, $R^2Y(cum)$ of 0.982, and a mean RMSECV of 4.1%, whereas the data from the whole time-domain without selection of a specific time-frame (0.0 - 5.5 ns) resulted in a mean RMSECV of 6.7%, $R^2X(cum)$ of 0.997, and $R^2Y(cum)$ with four PLS factors (**Table 1**).

Table 1. Data analysis performed on the Raman data with PLS indicating time-frame, method for baseline removal and RMSECV values obtained for each crystal form.

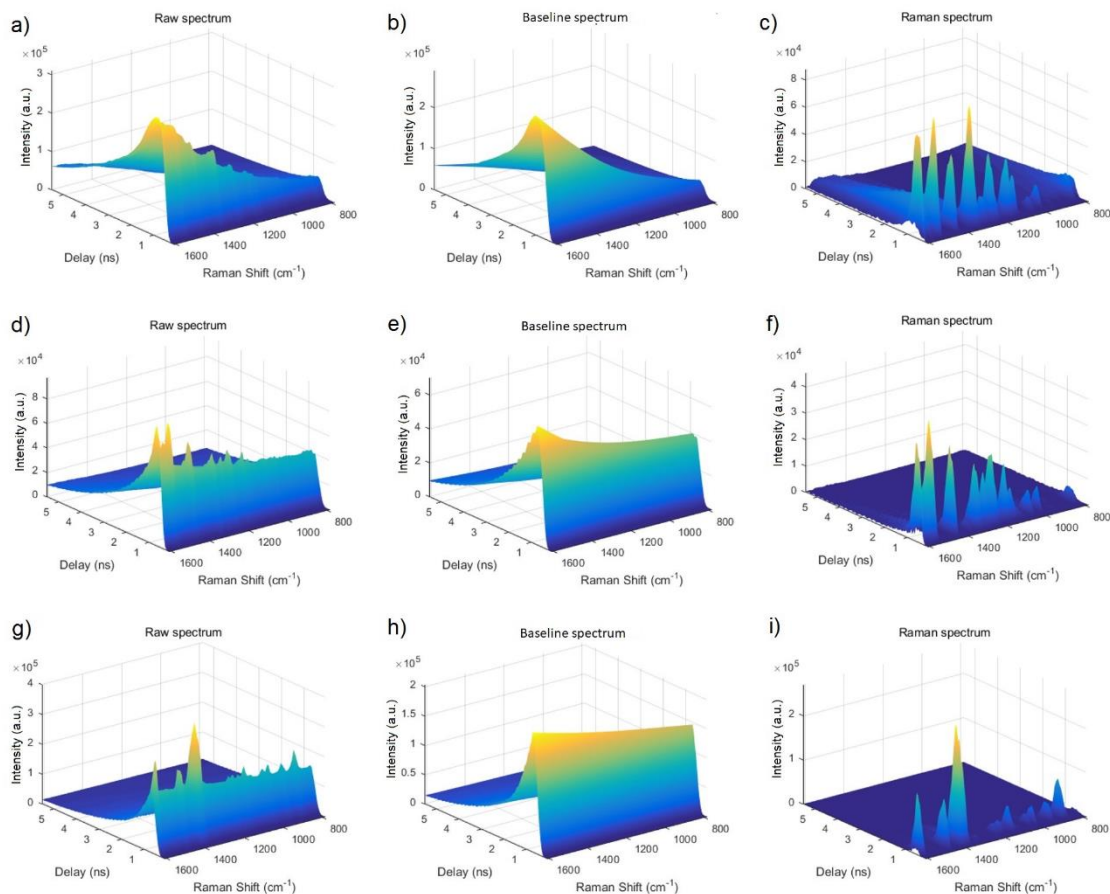| Time-frame (ns)	Baseline removal	RMSECV Form β	RMSECV Form α2	RMSECV MH
0.4 - 0.8	airPLS, Lmin	4.1%	4.5%	3.8%
0.0 - 5.5	airPLS, Lmin	7.5%	6.6%	6.0%

5

Figure 5. 3D spectra obtained with time-gated Raman of a) raw spectrum (form β), b) baseline spectrum (form β), c) Raman spectrum (form β), d) raw spectrum (form α2), e) baseline spectrum (form α2), f) Raman spectrum (form α2), g) raw spectrum (MH), h) baseline spectrum (MH), and i) Raman spectrum (MH).

The Raman signals of piroxicam were able to be detected over the fluorescence backgrounds. However, in the case of more extreme or complete Raman signal masking, an instrumental means to avoid fluorescence becomes essential. Time-gated Raman spectroscopy is one such approach.[25]

Kernel-based data analysis. Iterative optimization of the time-frame (an example of the process is presented in **Figure 6**) with the kernel-based RLS and greedy forward feature selection strongly affected the quantitative performance. Clear differences were observed in the quantitative performance between the optimized and non-optimized time-frames (**Table 2**). If the full time-frame data was used, mean RMSECV values of 6.2%, at best, were obtained. However, when the time-frame was optimized, the predictions improved, down to 1.4%. AirPLS (optimized $\lambda = 10$) was found most efficient with or without time-frame selection. Overall, this result suggests that kernel-based RLS analysis is a valid alternative to the PLS approach in this study for quantitative analysis of time-gated Raman spectra, as indicated by at best approximately three-fold lower RMSECV values.
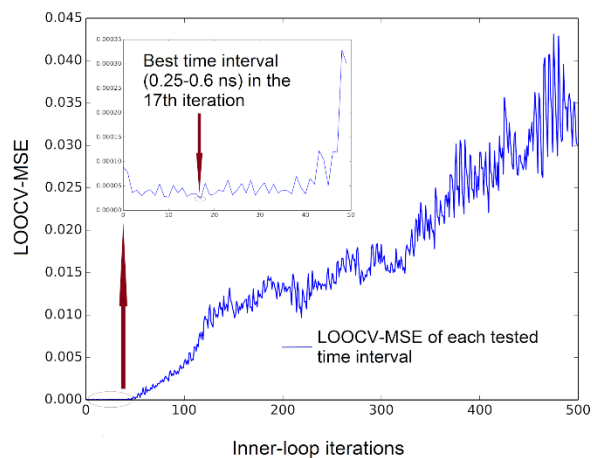
Figure 6. Leave-one-out cross-validation mean squared error (LOOCV-MSE) results from one round of the inner-loop of the kernel-based RLS model, where the model tries to find optimal parameters (time-interval, σ^2, λ) based on the LOOCV-MSE. The X-axis represents the number of different time intervals tested during each round of the model construction to find the optimal time interval along with the other optimal model parameters. The time interval corresponding to the lowest LOOCV-MSE was 0.25 - 0.6 ns in this example.

6

Table 2. Data analysis performed on the Raman data with kernel-based RLS indicating time-frame, method for baseline removal, use (+) (or absence (-)) of greedy feature selection, and RMSECV values obtained for each crystal form.

Time-frame (ns)	Baseline removal	Pre-treatment	Greedy feature selection	RMSECV Form β	RMSECV Form α2	RMSECV MH
0.35 - 0.60	airPLS, Lmin	SNV, mean centering	+	2.6%	2.5%	1.9%
0.25 - 0.60	airPLS	SNV, mean centering	+	1.6%	1.2%	1.5%
0.20 - 0.65	none	SNV, mean centering	+	2.1%	2.2%	1.6%
0.0 - 5.5	airPLS, Lmin	SNV, mean centering	-	8.4%	6.5%	5.9%
0.0 - 5.5	airPLS	SNV, mean centering	-	7.5%	6.2%	4.9%
0.0 - 5.5	none	SNV, mean centering	-	8.4%	7.9%	6.8%

The Gaussian kernel-based RLS model used in this study has the ability to learn target functions from the data capturing the nonlinearity of its features. The kernel-based RLS model accompanied by careful selection of the Raman shifts, time interval and the models' hyper-parameters utilizing a nested cross-validation resulted in improved prediction of the different drug forms in the mixtures. The result of this study supports exploration of the possibilities of efficient optimization of the time-frame as well as selection of the best Raman shifts for Raman analysis using kernel based methods and feature selection.

Overall, this study demonstrates that quantitative analysis with time-gated Raman spectroscopy can be suitable for solid-state analysis of photoluminescent pharmaceuticals during drug development and manufacturing. Raman spectroscopy is especially applicable for focusing on the properties of the API in mixtures and pharmaceutical products. This is because the functional moieties present in common APIs typically involve aromatic and pi-bonded structures which produce stronger Raman signals than the aliphatic and polar structures typical of common excipients. However, in addition to some APIs, many excipients (e.g. cellulose-based polymers) also fluoresce, which further restricts conventional Raman analysis for the analysis of pharmaceutical processing and dosage forms. An additional advantage of the time-gated measurements is that they can be performed in ambient lighting which facilitates analysis during pharmaceutical processing. These advantages mean that the time-gated Raman spectroscopy approach used in this study has much potential for process monitoring in pharmaceutical manufacturing.

Even though the PRX Raman bands were still observable without the fluorescence rejection, quantification was improved by the fluorescence rejection. Furthermore, the quantitative analysis approach in this study is applicable to more strongly fluorescing systems, as well as, for example, samples with high water contents, such as proteins, biological and biochemical samples. Altogether, the capability of the time-resolved Raman and fluorescence measurements with a CMOS SPAD detector for quantitative analysis shows promise in diverse areas, including fundamental chemical research, the pharmaceutical setting, process analytical technology (PAT), and the life sciences.

CONCLUSIONS

This study demonstrates that time-gated Raman spectroscopy is a useful tool for quantifying mixtures of fluorescent materials when conventional Raman spectroscopy could fail. PLS analysis of the time-gated spectra allowed quantitative analysis and demonstrated the benefit of time-domain selection. In this case, statistical optimization of model parameters using kernel-based RLS further improved the quantitative results. Overall, the time-gated Raman spectroscopy approach employed shows potential for relatively routine quantitative solid-state analysis of photoluminescent pharmaceuticals during drug development and manufacturing.

ASSOCIATED CONTENT

Supporting Information

Figure S-1. SEM images of the PRX crystal forms used in this study (.pdf).
The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

* clare.strachan@helsinki.fi

Author Contributions

‡These authors contributed equally.

Conflicts of interest

Lauri Kurki and Mari Tenhunen are affiliated with the company that commercialized the CMOS SPAD detector technology used in this research. The other authors declare no conflict of interest.

7

REFERENCES

(1) Gupta, R. B.; Kompella, U. B. *Nanoparticle Technology for Drug Delivery*; CRC Press: New York, 2006.

(2) Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. Crystalline solids. *Adv. Drug Delivery Rev.* **2001**, *48*, 3-26.

(3) Haleblian, J.; McCrone, W. Pharmaceutical applications of polymorphism. *J. Pharm. Sci.* **1969**, *58*, 911-929.

(4) Vankeirsbilck, T.; Vercauteren, A.; Baeyens, W.; Van der Weken, G.; Verpoort, F.; Vergote, G.; Remon, J. P. Applications of Raman spectroscopy in pharmaceutical analysis. *TrAC, Trends Anal. Chem.* **2002**, *21*, 869-877.

(5) Brittain, H. G. Spectral methods for the characterization of polymorphs and solvates. *J. Pharm. Sci.* **1997**, *86*, 405-412.

(6) Wartewig, S.; Neubert, R. H. Pharmaceutical applications of mid-IR and Raman spectroscopy. *Adv. Drug Delivery Rev*. **2005**, *57*, 1144-1170.

(7) Strachan, C. J.; Pratiwi, D.; Gordon, K. C.; Rades, T. Quantitative analysis of polymorphic mixtures of carbamazepine by Raman spectroscopy and principal components analysis. *J. Raman Spectrosc.* **2004**, *35*, 347-352.

(8) Das, R. S.; Agrawal, Y. Raman spectroscopy: Recent advancements, techniques and applications. *Vib. Spectrosc.* **2011**, *57*, 163-176.

(9) Tian, F.; Zhang, F.; Sandler, N.; Gordon, K.; McGoverin, C.; Strachan, C.; Saville, D.; Rades, T. Influence of sample characteristics on quantification of carbamazepine hydrate formation by X-ray powder diffraction and Raman spectroscopy. *Eur. J. Pharm. Biopharm.* **2007**, *66*, 466-474.

(10) Jestel, N. L. In *Process Analytical Technology Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*; Bakeev, K. A., Ed.; Blackwell Publishing Ltd.: Oxford, 2005; pp 133-169.

(11) Raman, C. V.; Krishnan, K. S. A new type of secondary radiation. *Nature* **1928**, *121*, 501-502.

(12) McCreery, R. L. In *Raman Spectroscopy for Chemical Analysis;* Wiley: Hoboken, 2000, pp 15-34.

(13) Aaltonen, J.; Gordon, K. C.; Strachan, C. J.; Rades, T. Perspectives in the use of spectroscopy to characterise pharmaceutical solids. *Int. J. Pharm*. **2008**, *364*, 159-169.

(14) Pelletier, M. Quantitative analysis using Raman spectrometry. *Appl. Spectrosc*. **2003**, *57*, 20-42.

(15) Strachan, C. J.; Rades, T.; Gordon, K. C.; Rantanen, J. Raman spectroscopy for quantitative analysis of pharmaceutical solids. *J. Pharm. Pharmacol.* **2007**, *59*, 179-192.

(16) Eriksson, L.; Byrne, T.; Johansson, E.; Trygg, J.; Vikström, C. *Multi- and megavariate data analysis - Basic principles and applications*, 3rd ed.; MKS Umetrics AB: Malmö, 2013.

(17) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109-130.

(18) Naula, P.; Airola, A.; Salakoski, T.; Pahikkala, T. Multi-label learning under feature extraction budgets. *Pattern Recog. Lett.* **2014**, *40*, 56-65.

(19) Pahikkala, T.; Airola, A. RLScore: Regularized least-squares learners. *J. Mach. Learn. Res*. **2016**, *17*, 1-5.

(20) Frank, I. E.; Friedman, J. H. A statistical view of some chemometrics regression tools. *Technometrics*. **1993**, 109-135

(21) Heinz, A.; Savolainen, M.; Rades, T.; Strachan, C. J. Quantifying ternary mixtures of different solid-state forms of indomethacin by Raman and near-infrared spectroscopy. *Eur. J. Pharm. Sci.* **2007**, *32*, 182-192.

(22) Chen, Z.-P.; Li, L.-M.; Jin, J.-W.; Nordon, A.; Littlejohn, D.; Yang, J.; Zhang, J.; Yu, R.-Q. Quantitative analysis of powder mixtures by Raman spectrometry: the influence of particle size and its correction. *Anal. Chem.* **2012**, *84*, 4088-4094.

(23) Fishburn, M. W. Fundamentals of CMOS single-photon avalanche diodes. Doctoral thesis, Delft University of Technology, The Netherlands, 2012.

(24) Patounakis, G.; Shepard, K. L.; Levicky, R. Active CMOS array sensor for time-resolved fluorescence detection. *IEEE J. Solid-State Circuits* **2006**, *41*, 2521-2530.

(25) Rojalin, T.; Kurki, L.; Laaksonen, T.; Viitala, T.; Kostamovaara, J.; Gordon, K. C.; Galvis, L.; Wachsmann-Hogiu, S.; Strachan, C. J.; Yliperttula, M. Fluorescence-suppressed time-resolved Raman spectroscopy of pharmaceuticals using complementary metal-oxide semiconductor (CMOS) single-photon avalanche diode (SPAD) detector. *Anal. Bioanal. Chem.* **2016**, *408*, 761-774.

(26) Iams, H.; Salzberg, B. The secondary emission phototube. *Proc. IRE* **1935**, *23*, 55-64.

(27) Van Duyne, R. P.; Jeanmaire, D. L.; Shriver, D. Mode-locked laser Raman spectroscopy. A new technique for the rejection of interfering background luminescence signals. *Anal. Chem.* **1974**, *46*, 213-222.

(28) Matousek, P.; Towrie, M.; Parker, A. Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shifted excitation Raman difference techniques. *J. Raman Spectrosc.* **2002**, *33*, 238-242.

(29) Matousek, P.; Towrie, M.; Ma, C.; Kwok, W.; Phillips, D.; Toner, W.; Parker, A. Fluorescence suppression in resonance Raman spectroscopy using a high-performance picosecond Kerr gate. *J. Raman Spectrosc.* **2001**, *32*, 983-988.

(30) Efremov, E. V.; Buijs, J. B.; Gooijer, C.; Ariese, F. Fluorescence rejection in resonance Raman spectroscopy using a picosecond-gated intensified charge-coupled device camera. *Appl. Spectrosc.* **2007**, *61*, 571-578.

(31) Blakesley, J.; See, P.; Shields, A.; Kardynał, B.; Atkinson, P.; Farrer, I.; Ritchie, D. Efficient single photon detection by quantum dot resonant tunneling diodes. *Phys. Rev. Lett.* **2005**, *94*, 067401.

(32) Kabuss, J.; Werner, S.; Hoffmann, A.; Hildebrandt, P.; Knorr, A.; Richter, M. Theory of time-resolved Raman scattering and fluorescence emission from semiconductor quantum dots. *Phys. Rev. B* **2010**, *81*, 075314-1-8.

(33) Stoppa, D.; Pancheri, L.; Scandiuzzo, M.; Gonzo, L.; Dalla Betta, G.-F.; Simoni, A. A CMOS 3-D imager based on single photon avalanche diode. *IEEE Trans. Circuits Syst. I* **2007**, *54*, 4-12.

(34) Stoppa, D.; Mosconi, D.; Pancheri, L.; Gonzo, L. Single-photon avalanche diode CMOS sensor for time-resolved fluorescence measurements. *IEEE Sens. J.* **2009**, *9*, 1084-1090.

(35) Rochas, A.; Gani, M.; Furrer, B.; Besse, P.; Popovic, R.; Ribordy, G.; Gisin, N. Single photon detector fabricated in a complementary metal–oxide–semiconductor high-voltage technology. *Rev. Sci. Instrum.* **2003**, *74*, 3263-3270.

(36) Nissinen, I.; Nissinen, J.; Keränen, P.; Kostamovaara, J. On the effects of the time gate position and width on the signal-to-noise ratio for detection of Raman spectrum in a time-gated CMOS single-photon avalanche diode based sensor. *Sens. Actuators, B* **2017**, *241*, 1145-1152.

(37) Mosconi, D.; Stoppa, D.; Pancheri, L.; Gonzo, L.; Simoni, A. CMOS single-photon avalanche diode array for time-resolved fluorescence detection. *2006 Proc. Eur. Solid-State Circuits Conference;* IEEE, 2006, pp 564-567.

(38) Nissinen, I.; Nissinen, J.; Keränen, P.; Länsman, A.-K.; Holma, J.; Kostamovaara, J. A 2×(4)×128 multitime-gated SPAD line detector for pulsed Raman spectroscopy. *IEEE Sens. J.* **2015**, *15*, 1358-1365.

(39) Kostamovaara, J.; Tenhunen, J.; Kögler, M.; Nissinen, I.; Nissinen, J.; Keränen, P. Fluorescence suppression in Raman spectroscopy using a time-gated CMOS SPAD. *Opt. Express* **2013**, *21*, 31632-31645.

8

(40) Schwartz, D. E.; Charbon, E.; Shepard, K. L. A single-photon avalanche diode imager for fluorescence lifetime applications. *2007 IEEE Symposium on VLSI Circuits*; IEEE, 2007, pp 144-145.

(41) Maruyama, Y.; Blacksberg, J.; Charbon, E. A 1024×8, 700-ps time-gated SPAD line sensor for planetary surface exploration with laser Raman spectroscopy and LIBS. *IEEE J. Solid-State Circuits* **2014**, *49*, 179-189.

(42) Lavrič, Z.; Pirnat, J.; Lužnik, J.; Puc, U.; Trontelj, Z.; Srčič, S. (14)N nuclear quadrupole resonance study of piroxicam: confirmation of new polymorphic form V. *J. Pharm. Sci.* **2015**, *104*, 1909-1918.

(43) Naelapää, K.; van de Streek, J.; Rantanen, J.; Bond, A. D. Complementing high-throughput X-ray powder diffraction data with quantum-chemical calculations: Application to piroxicam form III. *J. Pharm. Sci.* **2012**, *101*, 4214-4219.

(44) Sheth, A. R.; Bates, S.; Muller, F. X.; Grant, D. Polymorphism in piroxicam. *J. Cryst. Growth Des.* **2004**, *4*, 1091-1098.

(45) Upadhyay, P. P.; Bond, A. D. Crystallization and disorder of the polytypic α1 and α2 polymorphs of piroxicam. *CrystEngComm* **2015**, *17*, 5266-5272.

(46) Vrečer, F.; Vrbinc, M.; Meden, A. Characterization of piroxicam crystal modifications. *Int. J. Pharm.* 2003, 256, 3-15.

(47) Thomas, L. H.; Wales, C.; Wilson, C. C. Selective preparation of elusive and alternative single component polymorphic solid forms through multi-component crystallisation routes. *Chem. Commun.* **2016**, *52*, 7372-7375.

(48) Kogermann, K.; Aaltonen, J.; Strachan, C. J.; Pöllänen, K.; Heinämäki, J.; Yliruusi, J.; Rantanen, J. Establishing quantitative in-line analysis of multiple solid-state transformations during dehydration. *J. Pharm. Sci.* **2008**, *97*, 4983-4999.

(49) Eriksson, L.; Johansson, E.; Wikström, C. Mixture design—design generation, PLS analysis, and model usage. *Chemom. Intell. Lab. Syst.* **1998**, *43*, 1-24.

(50) Abdi, H. Partial least squares (PLS-regression). In *Encyclopedia for Social Science Research Methods*. Lewis-Beck, M.; Bryman, A.; Futing, T., Eds.; SAGE: Thousand Oaks, CA. 792–5, 2003

(51) Barnes, R.; Dhanoa, M. S.; Lister, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *J. Appl. Spectrosc.* **1989**, *43*, 772-777.

(52) Martens, H.; Naes, T. *Multivariate calibration;* John Wiley & Sons: New York, 1992.

(53) Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinf.* **2006**, *7*, 91.

(54) Shi, X.; El Hassan, N.; Ikni, A.; Li, W.; Guiblin, N.; de-Biré, A. S.; Ghermani, N. Experimental electron densities of neutral and zwitterionic forms of the drug piroxicam. *CrystEngComm* **2016**, *18*, 3289-3299.

(55) Redenti, E.; Zanol, M.; Ventura, P.; Fronza, G.; Comotti, A.; Taddei, P.; Bertoluzza, A. Raman and solid state ^{13}C-NMR investigation of the structure of the 1 : 1 amorphous piroxicam : β-cyclodextrin inclusion compound. *Biospectroscopy* **1999**, *5*, 243-251.

(56) Suresh, S.; Gunasekaran, S.; Srinivasan, S. Vibrational spectra (FT-IR, FT-Raman), frontier molecular orbital, first hyperpolarizability, NBO analysis and thermodynamics properties of piroxicam by HF and DFT methods. *Spectrochim. Acta, Part A* **2015**, *138*, 447-459.

9

Analytical Chemistry

Table of Contents artwork