Application of reverse engineering in the field of pharmaceutical tablets using Raman mapping and chemometrics

Application of reverse engineering in the field of pharmaceutical tablets using Raman mapping and chemometrics

Tereza Čapková[a,b,c,*], Tomáš Pekárek[b], Barbora Hanulíková[a], Pavel Matějka[c]

[a]*Centre of Polymer Systems, Tomas Bata University in Zlin, tř. Tomáše Bati 5678, 760 01 Zlín, Czech Republic*

[b]*Zentiva, k.s., Praha, U Kabelovny 130, 102 37 Prague 10, Czech Republic*

[c]*University of Chemistry and Technology Prague, Technická 5, 166 28 Prague 6, Czech Republic*

[*]*Corresponding author: Centre of Polymer Systems, Tomas Bata University in Zlin, tř. Tomáše Bati 5678, 760 01 Zlín, Czech Republic. E-mail address: capkova@utb.cz (T. Čapková).*

ABSTRACT

Raman micro-spectroscopy technique offers a combination of relatively high spatial resolution with identification of components or mixtures of components in different sample areas, e.g. on the surface or the cross-section of a sample. This study is focused on the analysis of the tablets from pharmaceutical development with different technological parameters: (1) the manufacturing technology, (2) the particle size of the input API (active pharmaceutical ingredient) and (3) the quantitative composition of the individual excipients.

These three mentioned parameters represent the most frequently solved problems in the field of reverse engineering in pharmacy. The investigation aims to distinguish tablets with the above-described technological parameters with limited subjective steps by Raman microscopy. Furthermore, non-subjective methods of Raman data analysis using advanced statistical analysis have been proposed, namely Principal Component Analysis, Soft Independent Modelling of Class Analogy and Linear Discriminant Analysis. The methods successfully distinguished and identified even very small differences in the analysed tablets within our study and provided objective statistic evaluation of Raman maps. The information on component and particle size distribution including their small differences, which is the critical parameter in the development of the original and generic products, was obtained due to combination of these methods. Even though each of these chemometric methods evaluates the data set from a different perspective, their mutual application on the problem of Raman maps evaluation confirmed and specified results on level that would be unattainable with the use of only one them.

Keywords: Raman mapping, multivariate analysis, tablets, Principal Component Analysis, Soft Independent Modelling of Class Analogy, Linear Discriminant Analysis

1. Introduction

The physicochemical properties of a dosage form affect the bioavailability of the drug product, which is one of the pharmacokinetic parameters driving the amount of drug in the active form in the systemic circulation and that is also crucial aspect in the development of new dosage forms or their generics. The generic drug is the equivalent of original medicinal product that contains the same active pharmaceutical ingredient (API), in the same type of dosage form, route of administration, safety, strength, quality and performance characteristics, and bioequivalence [**1,2**].

Reverse engineering/deformulation that is a complex process consisting of several phases, such as identification and quantification of formulation components, is used at the beginning of the development of a generic drug product. Another stage of reverse engineering is characterisation of the API, which can be divided into several levels: (1) molecular level, (2) particle level and (3) bulk level that has been described in the publication [**3**]. It would be the most appropriate to use the same form of API as the original product to ensure the same stability and dissolution profile [**3**]. In most cases, this strategy is blocked by the patents protecting the form of API used in the drug formulation [**3**]. Therefore, generic companies are forced to develop an alternative form of the same API for their product that may exhibit different physical and chemical properties. Particle size distribution (PSD), mainly of API, directly affects dissolution rate and bioavailability in the case of poorly soluble molecules [**4**], specifically for compounds of class II and IV of BCS (Biopharmaceutics Classification System) [**5**]. Thus, PSD must be taken into consideration to achieve the required similarity between the generic and the original product [**6**].

Table 1 The summary of the differences between the tablet formulations and their batches.

Formulation	1 batch	2 batch
A - difference in the manufacturing technology of API	API-1 and API-2 wet granulation	API-1 the granules, API-2 the powder form
B - difference in particle size of the input API	d_{90} = 81 μm	d_{90} = 50 μm
C - difference in quantitative composition	API 40% and 60% excipients	API 13% and 87% excipients

Reverse engineering is a useful tool for generic product development to achieve bioequivalence and stability of a drug product. The appropriate reverse engineering strategy, which includes decoding the qualitative and quantitative composition of the original medicinal product, characterization of the API and manufacturing process, can reduce development time and costs. Furthermore, the reverse engineering can be applied to improve API properties, minimise drug side effects, improve the composition and design of the product, identify defective or contaminated products. The development of pharmaceutical formulations requires a combination of analytical techniques, e.g. chromatography, X-ray diffraction, dissolution techniques as well as spectral methods. Raman microspectroscopy is an important technique in the reverse engineering process. The combination of chemometric methods with Raman mapping/imaging has allowed to visualise the distribution of the component [**7,8,9,10**] or study the interaction of API and excipient [**11,12**], as well as the mapping of the occurrence of different polymorphic forms [**13,12,14**] and stability during storage [**15**] of the solid dosage forms. It has also been applied to characterization of homogeneity [**16,17,18**] or identification of contamination [**19,20**], the analysis of tablet production technologies [**21**], when studying particle size [**22,23**] or determination of low API concentrations [**24,25**], in the study of morphology [**26**] and other physicochemical properties of the dosage form.

Our study is focused on the reverse engineering of several tablets from pharmaceutical development using Raman micro-spectroscopy. A comparison of the obtained Raman maps to assess, whether the

distribution of components is identical, similar, or completely different, might be challenging. The image analysis has often been used for such comparison, however, it depends on the subjective evaluation of the obtained Raman maps. The new procedure has been proposed for non-subjective assessment of Raman maps using Principal Component Analysis (PCA) [**27,28,29**] as a descriptive method and classification methods, particularly Soft Independent Modelling of Class Analogy (SIMCA) [**30,31**] and Linear Discriminant Analysis (LDA) [**32,33**]. Above-mentioned multivariate chemometric methods look at the data sets or Raman spectra from a different point of view. As for the information obtained from these statistical methods, they should match. PCA enables to reduce the dimensionality and minimises information loss [**34**]. Contrary to LDA and SIMCA methods, PCA calculation does not need training data set (so-called prediction model). PCA and LDA are partly similar, because they both examine linear combination to evaluate the data sets. PCA creates new uncorrelated variables called principal components and preserves maximum variance [**35**], whereas LDA finds a new space direction to project the data in order to maximise classes separation [**36**]. On the other hand, SIMCA is a class-modelling method like LDA but it is based on disjoint PCA, i.e. PCA is performed on each of the predefined classes from the prediction model [**37**]. The soft modelling allows to classify complex systems and samples that belong to more classes (samples can be assigned to more classes) which is the advantage over LDA.

The aim of the present research is the use of Raman microspectroscopy as a tool for the study of technological parameters of the pharmaceutical tablets using chemometric analysis. The tablets with different technological parameters were studied: (1) the manufacturing technology, (2) the particle size of the input API and (3) the quantitative composition of the individual excipients. The results show that the developed procedure involving Raman microspectroscopy analysis and consequent statistic evaluation can clearly reveal even small differences in formulations of tablets. The seemingly minor differences can have a significant impact on the behaviour of the tablet in solution, its disintegration and subsequent release and dissolution of API from the formulation. These factors are crucial in the development of a new generic tablet.

2. Materials and methods

2.1. Materials

Experiments were performed on three pairs of batches of pharmaceuticals tablets with different technological parameters. The tablets were provided by pharmaceutical company Zentiva, k. s., Czech Republic. Individual tablet components were of pharmaceutical quality. Each pair of tablets mutually differed by a specific technological parameter such as manufacturing technology, particle size of the input active pharmaceutical ingredient (API) and quantitative composition of tablets.

2.2. Tablet formulation

Individual tablets were manufactured under the good manufacturing practice procedure. Tablets of formulation type A, B and C are used in several indications including HIV/AIDS, treatment of inflammation or pain (non-steroidal anti-rheumatics) and reduction of thrombotic complications in cardiovascular events, resp. The first pair of batches labelled A had the same quantitative composition, although the batches 1A and 2A differed in the manufacturing technology. Formulation A contained two APIs and five excipients. The batch 1A was manufactured by mixing granules of both APIs, whereas the batch 2A was prepared from two different forms of APIs, i.e. granules (API-1) and powder (API-2).

The granulation process of APIs was the wet granulation in both cases. Formulation B contained one API and six excipients. This second type of tablets, 1B and 2B, had the same qualitative and quantitative composition. They were produced by the same technology but differed in particle size of the input API, which was d_{90} = 81 µm and d_{90} = 50 µm for batches 1B and 2B, resp. The third type of tablets, C, consisted of one API and three excipients of different quantitative proportion in two batches 1C and 2C.

The technological parameters of the tablet formulations and their batches are summarised in **Table 1**.

2.3. Tablet preparation for analysis

The representative tablet cross-sections were prepared for following analyses. The tablets were cut lengthwise using a scalpel (sc); manual cut. The tablets were also prepared in the form of a paraffin block and cut transversely by microtome (mc; microtome cut). Commercial Paraplast bulk (Leica Biosystems, Germany) was used to prepare paraffin blocks with tablets. The prepared blocks were allowed to solidify at room temperature and then the tablet in a paraffin block was cut with a microtome (Leica RM2265, Leica Microsystems, Germany).

The two tablets from each batch of formulations A, B and C were cut to obtain cross-sections suitable for analyses. Tablets prepared this way were placed on a glass slide which was inserted into the sample compartment equipped with a software-controlled motorised xyz stage and subjected to Raman analysis.

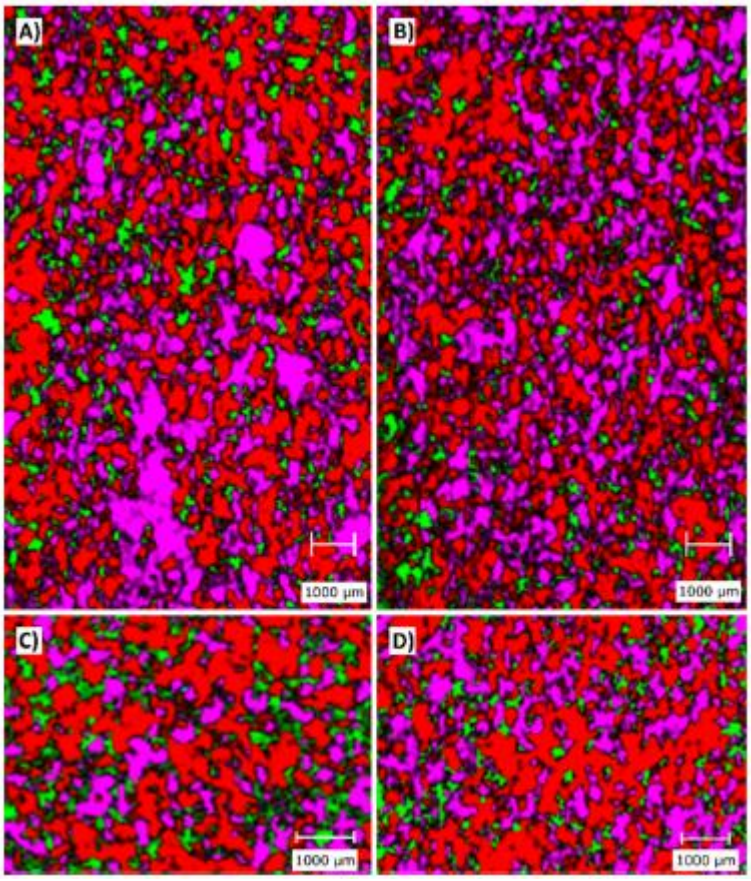

Fig. 1. Raman maps of formulation A; longitudinal cut with a scalpel - A) batch 1A, B) batch 2A and transversal cut with microtome - C) batch 1A D) batch 2A

2.4. Raman micro-spectroscopy

The cross-section of tablets was analysed on a dispersive inVia Reflex Raman microscope (Renishaw, UK) with an integrated Leica microscope. Raman microscope was operating with the Peltier-cooled CCD detector (Renishaw RenCam, 1040 x 256 pixels) and controlled by Wire 4.1 (Renishaw, UK) software. Raman spectra were collected using NIR laser with excitation wavelength of 785 nm in a line arrangement with the objective 5x (NA 0.12). Raman mapping was performed in StreamLine mode. Spectra were acquired in the range 1800 - 730 cm^{-1}. The size of the map was adjusted according to the real size of the tablet. The approximate map size was 625 x 620 pixels with x, y-step size 56.7 µm. The exposure time was 1 s and laser power was set up to 100%. The spectra of the C type formulation had contained a high noise level at an exposure time of 1 s. Therefore, the accumulation time was increased to 3 s to achieve a compliant signal-to-noise ratio for the C formulation.

2.5. Data evaluation and chemometric methods

Data sets of Raman spectra were processed to the chemical maps. The maps were created using the chemometric method Direct Classical Least Squares (DCLS) in the Wire programme (4.1, Renishaw, UK). The principle of creating Raman maps by DCLS was described in our last publication [**39**].

The challenge was to process the tens of thousands of spectra contained in each map to chemometric evaluation. For this reason, the in-house built programme Spectra Helper [**39**] has been designed and implemented to spectra processing. The Raman spectral sets were converted to an appropriate format (*. CSV) and averages spectra systematically. From each Raman map, we obtained 800 averaged spectra. In total, 12 tablets cross-sections were compared, which involved 9600 spectra (e.g. 1A_sc_sp1, 1A_sc_sp2, 2A_sc_sp1, 2A_sc_sp2, 1B_sc_sp1, 1B_sc_sp2, 2B_sc_sp1, 2B_sc_sp2, 1C_sc_sp1, 1C_sc_sp2, 2C_sc_sp1, 2C_sc_sp2). The spectra were processed by three multivariate chemometric methods, PCA, SIMCA and LDA, in the programme The Unscrambler X (10.5, CAMO Software AS, Norway) and in a free software R. The whole sets of spectra (800 x 12 spectra) were used for PCA. The NIPALS (Non-linear Iterative Partial Least Squares) algorithm, cross validation and 95% confidence intervals were set within PCA calculation. In case of SIMCA and LDA training data (spectra) of each tablet cross-section were needed for calculation of the prediction model. For this purpose, half of each spectral set (400 x 12 spectra) was used as predictive and the remaining half (400 x 12 spectra) as classification spectra. All three methods evaluated the same data. Generally, SIMCA performs PCA on each data set of the calibration set. SIMCA as PCA uses cross validation criterion. The PCA calibration sets were calculated under the same conditions as for the stand-alone PCA. Before calculating the LDA, it is necessary to define a category variable and create a classification model. LDA use linear method (Fisher's linear discriminant) for calculation.

3. Result and discussion

3.1. Analysis of Raman maps

In all cases of tablets with different technological parameters, Raman maps were measured from longitudinal and transversal cross-sections of tablets prepared with a scalpel and a microtome, respectively.

The first type of tablet formulation A contained two APIs. The distribution of components is distinguished by the colours in the Raman maps; the red colour represents API-1 and magenta colour

stands for API-2. Both tablets consisted of approx. 73% of APIs. The most abundant excipient (EX-1) representing 20% of the tablet composition is displayed green in the maps. Other excipients were present in very small amounts ranging from 4% to 0.4%, therefore, we do not observe them in the Raman maps. Batches 1A and 2A were prepared by different manufacturing technology (see chapter 2.2. Tablet formulation).

The representation of the individual components in the map of both batches is very similar (**Fig. 1**). In the longitudinal cross-section of tablet 1A (Raman map **Fig. 1**A), API-1 and API-2 are observed to form distinct particle clusters. This indicates that both APIs were prepared with the same technique, in this case by wet granulation. Whereas, API-2 in the tablets of batch 2A (**Fig. 1**B) was distributed more evenly and did not form clusters of significant size. This can be attributed to the powder form of API-2 mixed in the granulate of API-1. The space between the APIs is filled with the EX-1 similarly in both cases. This fact is less pronounced in the Raman maps of tablets cut transversely with a microtome (**Fig. 1**C and D). This is probably due to the smaller cross-sectional area of the tablets. Despite this fact, a clear difference in particle cluster size of API-2 can still be observed corresponding with its manufacturing process.

The tablet formulations B differed in the input particle size of the API. API is represented in yellow in the maps. The two most abundant excipients EX-1 and EX-2 were shown in magenta and green. Batches 1B and 2B had different distribution of API (**Fig. 2**A and B longitudinal cross-section, C and D transversal cross-section). In the map of batch 1B (Fig. 2A), the API aggregated into larger clusters homogeneously distributed in the tablet. In contrast, the API of batch 2B created more small clusters in the Raman map (**Fig. 2**B). However, the fact that the size of the clusters is not necessarily related to the input particle size as large clusters can contain either large or small particles, the difference in batch 1B and 2B is evident and indicates that smaller input particle size has led to creation of smaller API clusters in the tablet. Moreover, it seems that the smaller clusters of API in 2B have facilitated a creation of larger clusters of both excipients. Due to the porosity of tablets B (uncoated tablets), paraffin penetrated their surface partially and it was difficult to evaluate transversal cross-section (black places - paraffin) of these tablet maps by Raman software (**Fig. 2**C and D). In Raman maps on **Fig. 2**C and D, we can only observe the centre of the tablet, which is not sufficient for proper tablet assessment. Therefore, it is not possible to visually evaluate the technological parameters on these maps.

The third type of tablet with formulation C differed in the quantitative composition of the individual excipients, and in the total weight of the tablets. Although API was added in equal amounts to both batches, an overall tablet weight of batch 2C was more than 50% higher than the weight of tablets of batch 1C. The excipients formed about 60% of tablet (API 40%) and 87% of tablet (API 13%) of batch 1C and 2C, respectively. The obtained Raman maps of batch C are given in Fig. 3. API is represented in yellow and the two excipients EX-1 and EX-2 are shown in green and red colour, respectively. The higher API percentage is apparent from Raman maps of batch 1C (**Fig. 3**A and C), which is evident from both a longitudinal and a transversal tablets cross-section. The smaller amount of API, based on the total weight of the tablet, is observed in the map of tablet 2C (**Fig. 3**B and D). In both cases (batches 1C and 2C), a proportional increase/decrease of the green-labelled EX-1 and red-labelled EX-2 can be observed. Raman maps thus correspond to the amount of API as well as excipients in the tablets.

3.2. Statistical multivariate methods

To assess the similarity or difference of Raman maps without a subjective analysis, a procedure using PCA and classification methods, specifically SIMCA and LDA, was proposed. The Raman spectra of tablets cross-section with different technological parameters were processed and compared. The

extracted Raman spectra from Raman maps were processed in our in-house built programme Spectra Helper as shown in our last publication [**39**]. The newly processed data contained 800 Raman spectra averages per map, see chapter 2.5. Data evaluation and chemometric methods.

Raman maps of microtome cut tablets were difficult to interpret, e.g. in case of tablets of formulation B. Therefore, the statistical results were focused only on the spectra measured from the longitudinal cross-section tablets cut with a scalpel. Statistical results of microtome cross-sections of tablet A, B and C are presented in the **supplementary data**. The results obtained from the microtome-cut transversal cross-section (PCA - **Fig. S1**, SIMCA - **Table S4**, LDA -**Table S6**) were similar to the results gained from scalpel-cut crosssections.

3.2.1. Principal component analysis

PCA method allows data analysis without prior comparison with prediction classes. The importance of the original variables (their contribution rate) is presented by the spectral loadings and projection of objects (spectra) along the new PCs (Principal Components) in the so-called PC score plots.

PCA calculations were carried out for comparison of particular batches of test tablets sets. The PC score plots of tablets with formulations A, B and C are shown in **Fig. 4**. The average spectra in PC score plots of different tablet batches are represented by dots while individual batches are distinguished with colours. The first principal component (PC-1) for all three formulations describes the most variability in the data (corresponding to the technological parameters); 95%, 81% and 100%, for formulations A, B and C, resp. The second principal component (PC-2) of formulation A and C describes only a very small part of the overall data variability. The exception is the formulation B with higher PC-2 value 13%. The information about the origin of the main influence of distribution points in PC score plot was obtained from PC-1 loading, which is similar to the spectrum of the API in all cases of formulations A, B and C. The spectra of formulation A labelled with dark blue and red colour dots in the PC score plot (**Fig. 4**A) belong to the batch 1A and dots of the lighter colours belong to spectra of batch 2A. The average spectra (tablet samples of each batch; sp1 and sp2) of a batches 1A and 2A form separate clusters. Distance between the clusters is the smallest of all tested tablet formulations because the tablets 1A and 2A had the same composition.

The dark green and dark red dots in the score plot (**Fig. 4**B) indicate the average spectra of batch 1B and the light shaded dots represent the spectra of batch 2B. The calculated PC model distances for these batches (**Fig. 4**) are higher than for batches of formulations A. The formulation B has different arrangement of score plot in comparison with A and C with clusters separated into the upper and lower part of the plot. The PC loadings explain this separation (**Fig. 5**). PC-1 loadings show (**Fig. 5**A) that the first principal component has the same course as the API Raman spectrum. These loading-spectra are overlaid and highlighted in red (**Fig. 5**A). It can be seen that the courses of PC-1 loadings of batches 1B and 2B are identical. PC-2 loading (PC-2 13%) course contains some bands that can be also assigned to the API (**Fig. 5**B). This indicates that PC-2 has in case of different API's particle size distribution influence that is not negligible and therefore encourages the separation of batches (dots in score plot) of formulation B. For other formulations the effect of PC-2 is not pronounced substantially because its value reached 3% or less than 1% of overall variability for A and C, respectively. This implies that the individual clusters of average spectra 1B and 2B are more different from each other than the clusters of the averaged spectra obtained for formulation A and C.

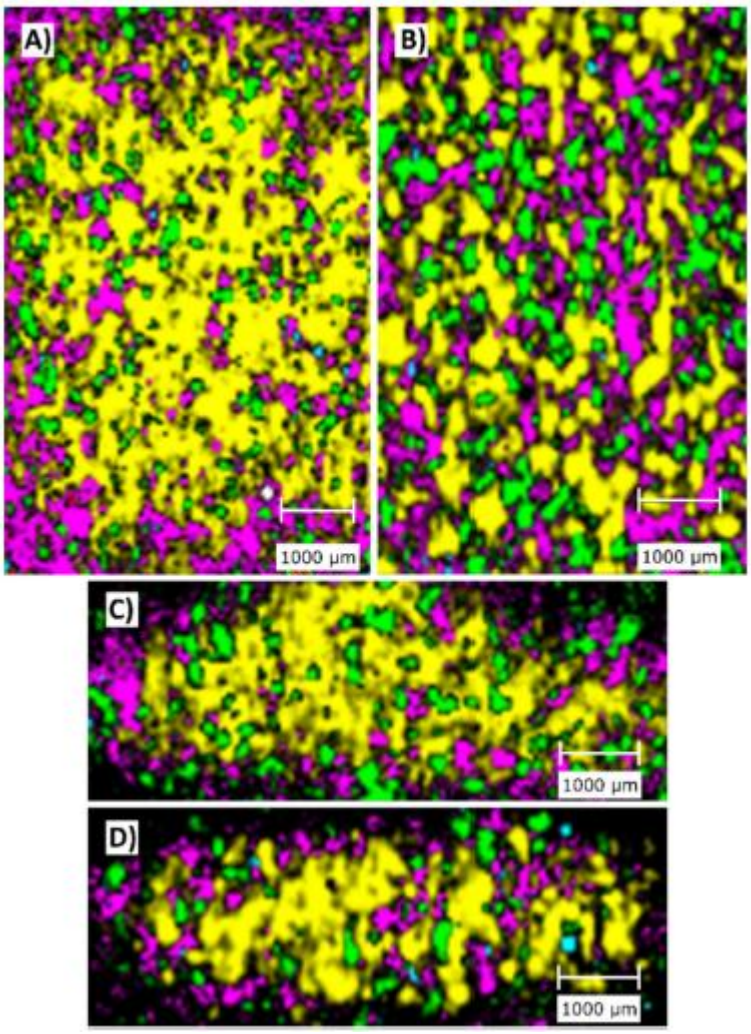

Fig. 2. Raman maps of formulation B; longitudinal cut with a scalpel - A) batch 1B, B) batch 2B and transversal cut with microtome - C) batch 1B D) batch 2B.

The last tablet type of formulation C differed in the quantitative composition. In the score plot (**Fig. 4**C), the spectra (dots) of individual batches are divided into two clusters, as in the previous cases. Dark violet and green dots represent the average spectra of batch 1C and analogously their light shades batch 2C. The spectra of batch 1C are more variable than 2C spectra, clusters of which are more compact. The spectra of both batches C showed strong fluorescence that could not have been minimised. Nevertheless, the vibration bands belonging to the API are observed in the PC-1 loading. Therefore, API amount in the tablets (API 1C 40% and API 2C 13%) has the most significant influence on the distribution in the score plot.

3.2.2. Soft independent modelling of class analogy

SIMCA method allows to classify spectra into predefined groups (prediction model) and calculate distances between data. SIMCA does not search for differences between classes but it describes the similarities between the samples within a class. The mutual distances between the individual tablet formulations of their batches with different technological parameters, computed from the PCA models, were calculated using the SIMCA. The average spectra (800 spectra per map) of the tested

tablets were divided exactly in to two halves. Individual PCA models (prediction model) were constructed from the 400 average spectra per map for each tested tablet, e.g. 1A_sc_sp1, 1A_sc_sp2, 2A_sc_sp1, 2A_sc_sp2 (sample, sp; scalpel cut, sc) etc., see chapter 2.5. Data evaluation and chemometric methods. The remaining part of the spectra was used as classification spectra. Outliers (spectra) were excluded in the prediction model and classification spectra and the model was recalculated without these outliers.

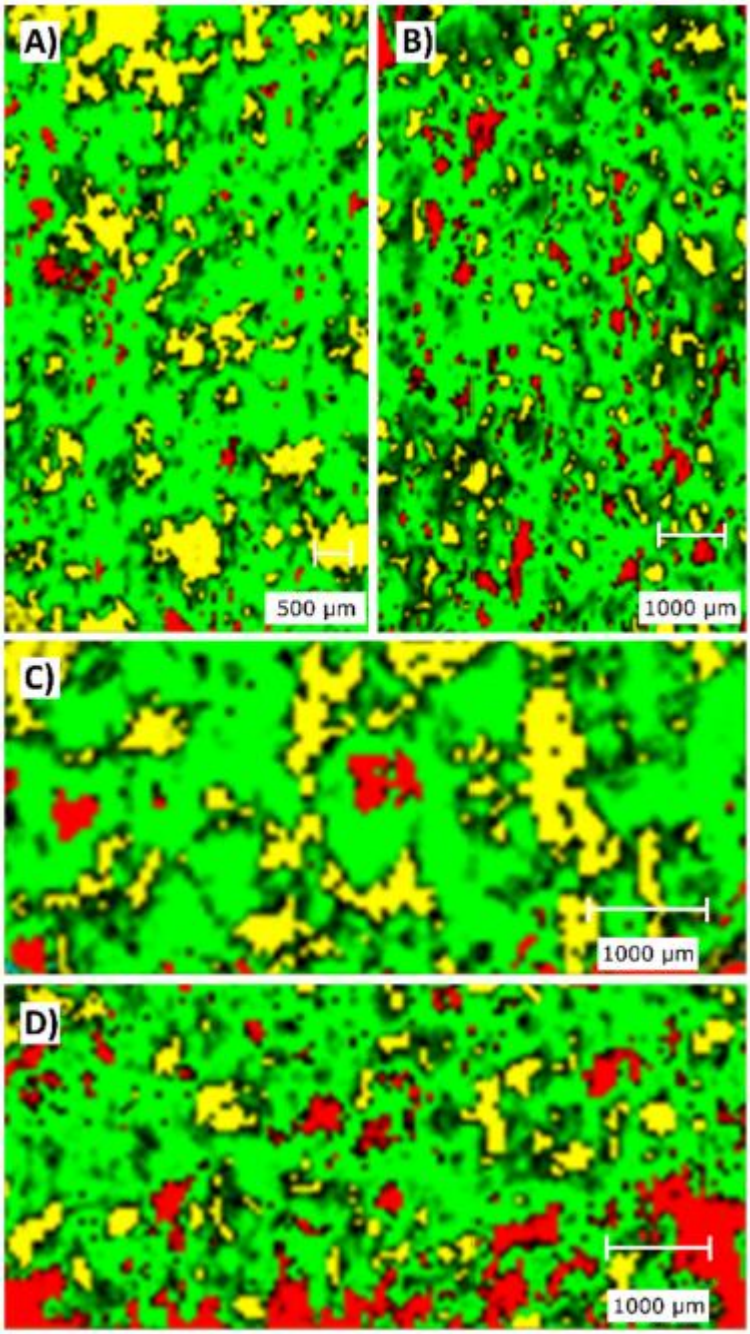

Fig. 3. Raman maps of formulation C; longitudinal cut with a scalpel - A) batch 1C, B) batch 2C and transversal cut with microtome - C) batch 1C D) batch 2C.

The typical output of SIMCA calculation is a table of distances to individual PCA classes, e.g. tablets formulation A, B and C. The distance values, for the first 10 spectra, are highlighted in shades of red, blue and green for tablets A, B and C, respectively, in **Table S2**. The distances between tablets of the same batch sp1 and sp2 are close to one, for example 1.2, 1.3, 1.4 or 1.5 (Table S2). This means that the samples are almost indistinguishable, because PCA models (tablet spectra) are similar. The exceptions are tablets 2B_sc_sp1 and sp2, which have a slightly higher mutual distance, namely 2.2. Histogram graphs of distances between PCA models (**Fig. 6**) were also generated for better orientation in SIMCA results. Since the distances, respectively the differences for example between the tablet formulations A and C are in the thousands, therefore, for clarity, the distances in the histograms were defined only in the range 0-100.

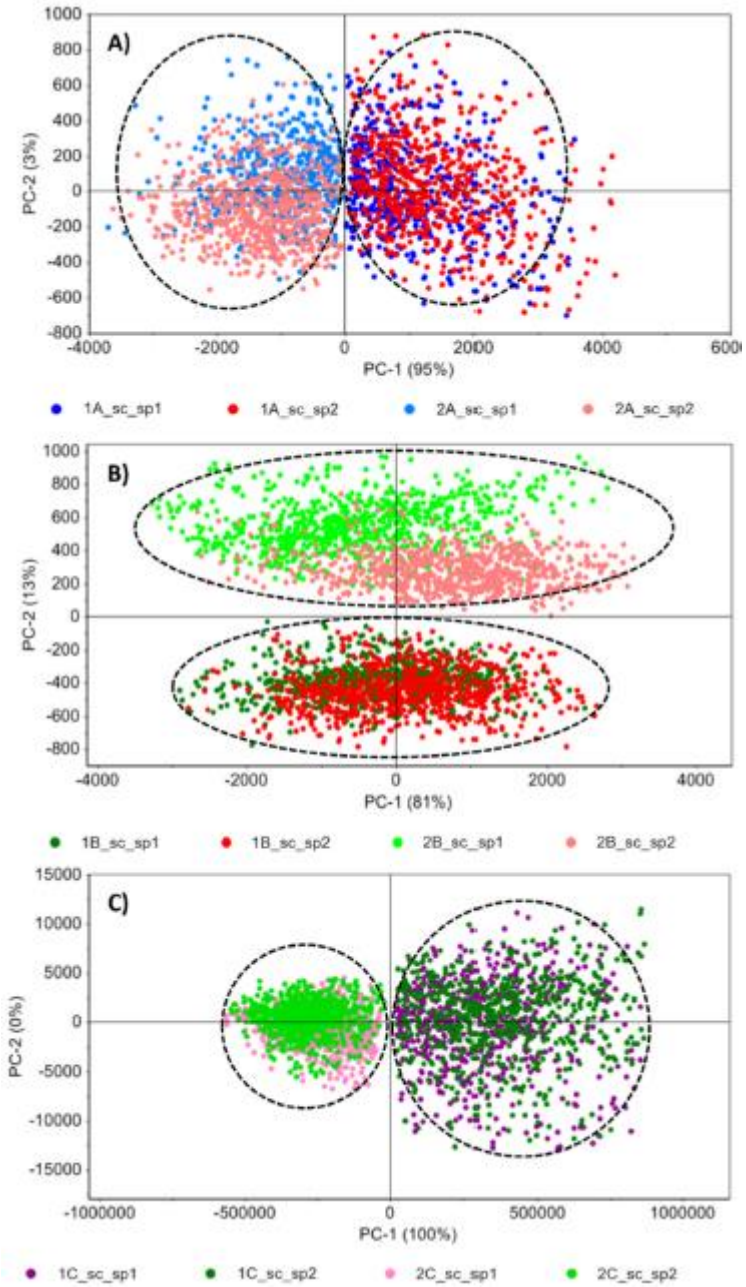

Fig. 4. Component score plots of PC-1 and PC-2 for tablet formulations A, B and C - scalpel cut (sc) longitudinal cross-section, two samples of each batch; A) batch 1A and 2A, B) batch 1B and 2B, C) batch 1C and 2C.

The title of each graph (Fig. 6) shows the PCA model which all tablet samples were compared with. As in the case of distances in the Table S2, a low value (small height of the histogram) indicates similarity to the model, and high values indicate a large distance, i.e. the largest differences. Fig. 6 shows the distances for all formulations and batches of tablets A, B and C, but only for their samples 1 (sp1). The results for samples 2 (sp2) are identical and are given in the supplementary data Fig. S3. The histogram SIMCA calculation also confirmed that two tablets of the same batch (sp1 and sp2) cannot be distinguished, because their distances were very small. The low distance values within one batch of tablets were expected because the individual tablets of a single batch should be identical.

From the perspective of individual A and B formulations, all spectra were classified correctly according to the batches. In the case of 1C and 2C batches, occasional classification to the correct and incorrect class occurred simultaneously within the formulation C.

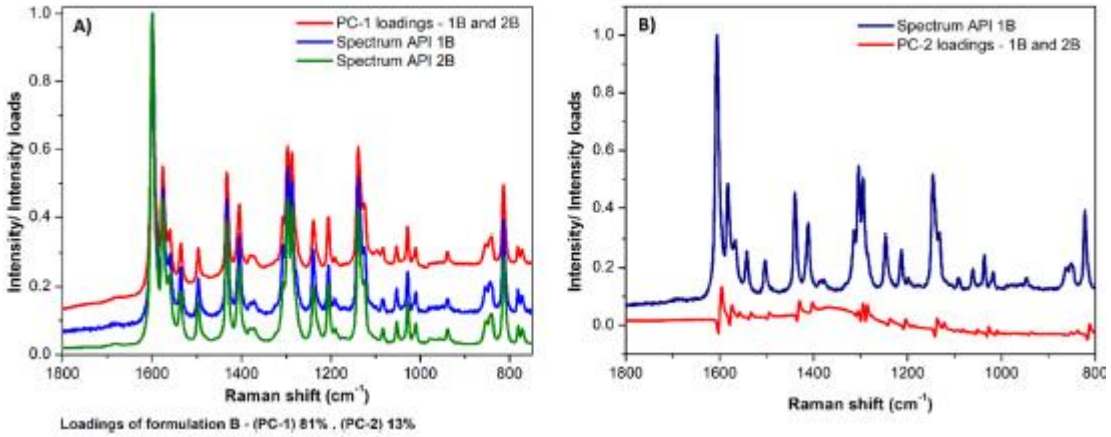

Fig. 5. Comparison of A) PC-1 loadings of the batches 1B and 2B, B) PC-2 loadings of the batches 1B and 2B with Raman spectra of corresponding API.

The ambiguity is due to the significant fluorescence in the spectra of 1C and 2C batches. In terms of distances between individual formulation models, the smallest difference in distances is in formulation A and C. These low distances are caused by formulation A (**Table S2** A) in that the tablets have the same composition and they differ in the manufacturing technology of API. The formulation C (**Table S2** C, **Fig. 6** 1C_sc_sp1, **Fig. 6** 2C_sc_sp1) on the contrary, they differ in quantitative composition but the fluorescence in the spectra affects the classification and thus the distance between the tested batch groups. The substantially higher distances (19.3 - 26.8) between batches B indicates that the input particle size influenced preparation and thus the tablet composition. When comparing all A, B and C formulations mutually, the classification to each category was correct. The individual formulation models differ from each other by the high PCA distance values, see **Table S2**.

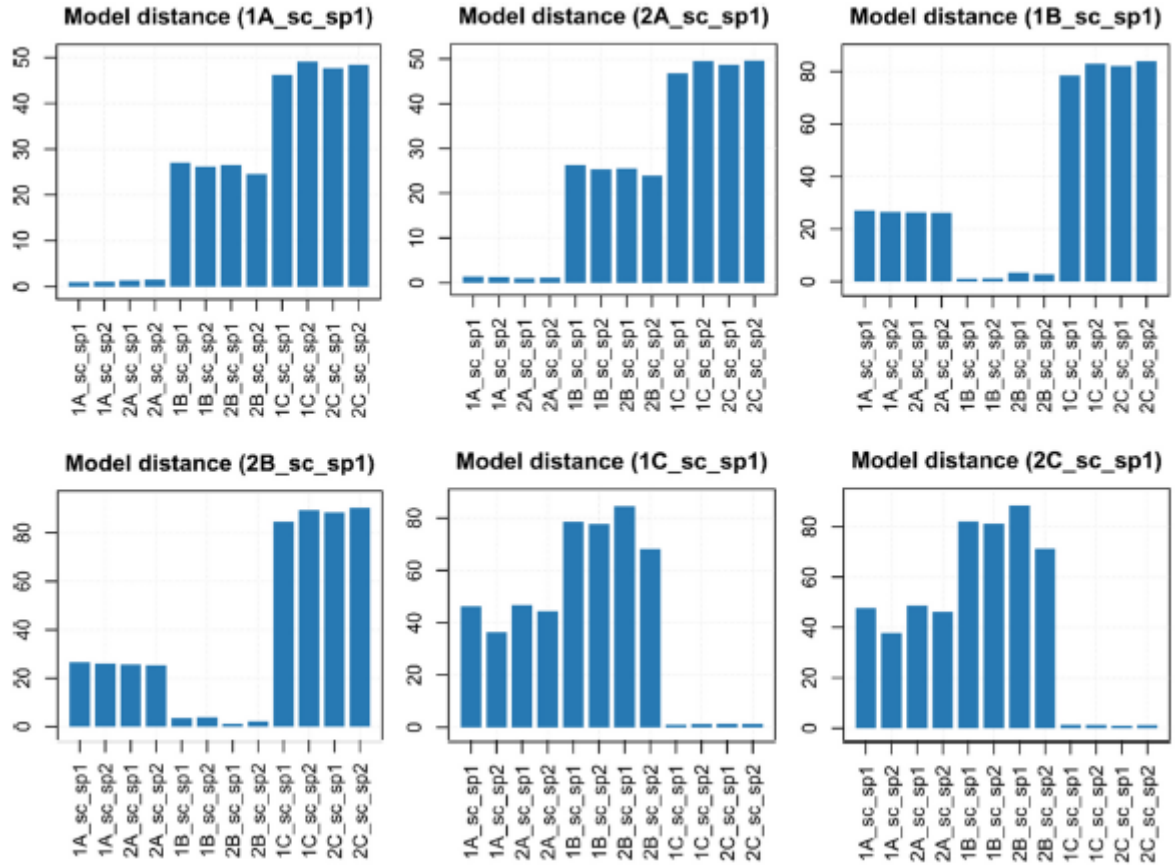

Fig. 6. Histogram SIMCA distances between PCA models of tablets with formulation A, B and C for sample 1 - scalpel cut.

3.2.3. Linear discriminant analysis

LDA, like the SIMCA method, belongs to classification methods that use the so-called primary classification, for sorting spectra into predefined classes. The discriminant function values were calculated for 400 of the 800 average spectra per map of each tested tablet separately, see chapter 2.5. Data evaluation and chemometric methods. Based on the discriminant function values, the spectra were classified to primary classes (prediction classes). The classification spectra, the other 400 spectra per map, were classified based on the most similarity to the primary classes (smallest distance).

In our case, a linear method was used to calculate discriminators. Like in the case of the SIMCA classification, all spectra of A and B formulations were correctly assigned to classes. Occasionally, 1C batch was classified as 2C batch, due to significant fluorescence in spectra. The LDA method points out the similarity or homogeneity of individual tablets within the same batch. Therefore, the samples sp1 and sp2 were classified incorrectly in the same batch very often. The obtained results of the first 10 classified spectra from each formulation and tested batches are shown in **Table S5**. The first column of the table shows the names of the tested spectra according to the formulation and the batch which they belong to. In the next columns calculated distances with the given classes are listed. The last column shows the name of the class to which the spectrum was assigned according to the calculated LDA distances.

The values of discriminators between particular batches (1A and 2A, 1B and 2B, 1C and 2C) are lower compared to the discriminators values of tablets formulations (A, B and C). The formulation A had the same quantitative composition and a similar component distribution in tablets, therefore, batches 1A and 2A have only slightly differed, as with previous statistical calculations. The value of discriminators

1A and 2A is only approx. ± 1. Despite this fact, the LDA method has classified the batches 1A and 2A correctly. The higher differences in discriminator values are between batches B (approx. ± 10) and the largest between batches C (approx. ± 30). These tablets differ more in the distribution of the ingredients and, in the case of batch C in the quantitative composition.

Classification methods such as SIMCA and LDA have significantly contributed to the distinguishing of tablet formulations, even though each of the methods uses a different procedure to classify the spectra. SIMCA is based on PCA and its advantage is that it can identify samples that belong to multiple classes, i.e. overlapping classes. In contrast, the LDA search a linear combination that characterises a class or distinguishes classes/samples from each other. Very low distance between spectra found by SIMCA and LDA indicates that tablets 1A and 2A are similar, which agree with the fact that they were of the same composition but different manufacturing procedure. Another determined parameter was the particle size of the input API of formulation B. Particle size distribution analysis is classically performed from disintegrated tablets (e.g. microscopically or by laser diffraction). However, the different distribution of API with different particle size has been clearly observed in the Raman maps of B. The spectra of batches 1B and 2B were well distinguished from each other by statistical analysis. The results of SIMCA and LDA methods of both batches B have revealed the largest differences in distances compared to other batches, which indicates that the input particle size of API influenced the preparation of tablets. The formulation of C differed in the quantitative composition and in the total weight of the tablets. The spectra of batches 1C and 2C were loaded with fluorescence. However, most of the spectra were classified by SIMCA and LDA to batch 1C or 2C correctly.

4. Conclusions

Our work provides comprehensive information on the possibilities of analysis by Raman microscopy with advanced chemometrics used to distinguish several pharmaceutical tablets with different parameters; represented the most solved problems in the field of reverse engineering in pharmacy. The Raman technique with advanced chemometric methods enables to distinguish formulations with a very close composition, a similar distribution of components and particle size distribution of API. Even very small differences may have a significant effect on the overall behaviour of tablets, for example, in disintegration and dissolution of tablet to release and dissolve API from a formulation. The classic example for the use of this combination of Raman technique with advanced chemometric processing is in the development of a new generic tablet, or in the detection of counterfeit [38] medicines. The next steps of our investigation involve a testing of the usability of the chosen technique not only on real tablets from development, but also on the tablets from the production and the market. The advantage of the use of Raman mapping technique for non-subjective data evaluation is that we work and study with the whole tablet formulation. The samples do not need to be modified significantly, separated, or disintegrated in order to obtain information about the tablet production parameters. The main novelty is a unique combination of objective evaluation of Raman maps for reliable distinguishing of similar tablets by several chemometric methods and the evidence that this approach is suitable for practical application in pharmaceutical R&D.

PCA analysis was used in the study as independent method that allows to categorise the sets of data based on the largest variability, such as the different parameters, and transform them into the principal components. The method allows to organise data in a multidimensional space without minimal loss of information. The classification methods SIMCA and LDA were used to verify the accuracy of the results, so that the data were grasped from different point of view. The individual technological parameters of formulations A, B and C were recognized from Raman maps of tablet cross-sections and from the

results of advanced chemometric methods. PCA has successfully distinguished all tested tablets produced with different parameters; namely by PC-1, PC-2 and PC-loading that described the highest variability corresponded to the API (e.g. form, particle size and amount of API). Moreover, the classification methods SIMCA and LDA have been able to identify tablet composition or distribution of components from calculation of distances between spectra of different tablet batches. Based on the experience from these analyses, the methods are applicable in reverse engineering.

References

[1] Generic Drug Facts=FDA, (n.d.). (**https://www.fda.gov/drugs/generic-drugs/ generic-drug-facts**) (Accessed 23 July 2021).

[2] S. Dunne, B. Shannon, C. Dunne, W. Cullen, A review of the differences and similarities between generic drugs and their originator counterparts, including economic benefits associated with usage of generic medicines, using Ireland as a case study, BMC Pharmacol. Toxicol. 14 (2013), **https://doi.org/10.1186/2050-6511-14-1**

[3] A.K. Bansal, V. Koradia, The Role of Reverse Engineering in the Development of Generic Formulations, Pharm. Technol. 29 2005 50-55. (**https://www. pharmtech.com/view/role-reverse-engineering-development-generic-formulations**) (Accessed 23 July 2021).

[4] B. Munjal, V. Koradia, S.H.S. Boddu, A. Bansal, Role of innovator product characterization in generic product development, in: A.S. Narang, S.H.S. Boddu (Eds.), Excipient Applications Formulation Design and Drug Delivery, 2015, pp. 521-538, **https://doi.oig/10.1007/978-3-319-20206-8_17**

[5] A. Charalabidis, M. Sfouni, C. Bergstrom, P. Macheras, The biopharmaceutics classification system (BCS) and the biopharmaceutics drug disposition classification system (BDDCS): beyond guidelines, Int. J. Pharm. 566 (2019) 264-281, **https://doi.org/10.1016/j.ijpharm.2019.05.041**

[6] D. Wu, B. Sarsfield, Particle size reduction: from microsizing to nanosizing, in: F. Giordanetto, R. Mannhold, H. Buschmann, J. Holenz (Eds.), Early Drug Dev. Bringing a Preclin. Candidate to Clin, 2018, pp. 271-304, **https://doi.org/10.1002/ 9783527801756.ch11**

[7] M. Windbergs, M. Haaser, C.M. McGoverin, K.C. Gordon, P. Kleinebudde, C.J. Strachan, Investigating the relationship between drug distribution in solid lipid matrices and dissolution behaviour using raman spectroscopy and mapping**Maike Windbergs and Miriam Haaser contributed equally to this work, J. Pharm. Sci. 99 (2010) 1464-1475, **https://doi.org/10.1002/jps.21894**

[8] S. Šašic, P. Ojakovo, M. Warman, T. Sanghvi, Raman chemical mapping of magnesium stearate delivered by a punch-face lubrication system on the surface of placebo and active tablets, Appl. Spectrosc. 67 (2013) 1073-1079, **https://doi.org/ 10.1366/13-07012**

[9] Y. Yamamoto, M.Y. Fujii, T. Fukami, T. Koide, Evaluation of the three-dimensional distribution of droplets in a droplet dispersion-type ointment using confocal Raman microscopy, J. Drug Deliv. Sci. Technol. 51 (2019) 639-642, **https://doi. org/10.1016/j.jddst.2019.04.003**

[10] S. Šašic, T. Prusnick, Fast Raman chemical imaging of tablets with non-flat surfaces, Int. J. Pharm. 565 (2019) 143-150, **https://doi.org/10.1016/j.ijpharm.2019.05.004**

[11] PJ. Skrdla, D. Zhang, Disproportionation of a crystalline citrate salt of a developmental pharmaceutical compound: characterization of the kinetics using pH monitoring and online Raman spectroscopy plus quantitation of the crystalline free base form in binary physical mixtures, J. Pharm. Biomed. Anal. 90 (2014) 186-191, **https://doi.org/10.1016/jjpba.2013.12.001**

[12] A. Hédoux, Y. Guinet, M. Descamps, The contribution of Raman spectroscopy to the analysis of phase transformations in pharmaceutical compounds, Int. J. Pharm. 417 (2011) 17-31, **https://doi.org/10.1016/j.ijpharm.2011.01.031**

[13] J.R. Beattie, L.J. Barrett, J.F. Malone, J.J. McGarvey, M. Nieuwenhuyzen, V.L. Kett, Investigation into the subambient behavior of aqueous mannitol solutions using temperature-controlled Raman microscopy, Eur. J. Pharm. Biopharm. 67 (2007) 569-578, **https://doi.org/10.1016Zj.ejpb.2007.03.015**

[14] D.R. Willett, H. Yilmaz, A.M. Wokovich, J.D. Rodriguez, Low-frequency Raman mapping and multivariate image analysis for complex drug products, Am. Pharm. Rev. (2019), (**https://www.americanpharmaceuticalreview.com/Featured-Articles/360382-Low-Frequency-Raman-Mapping-and-Multivariate-Image-Analysis-for-Complex-Drug-Products/**) Accessed July 23, 2021.

[15] A. Docoslis, K.L. Huszarik, G.Z. Papageorgiou, D. Bikiaris, D. Stergiou, E. Georgarakis, Characterization of the distribution, polymorphism, and stability of nimodipine in its solid dispersions in polyethylene glycol by micro-Raman spectroscopy and powder X-ray diffraction, AAPS J. 9 (2007) E361-E370, **https:// doi.org/10.1208/aapsj0903043**

[16] A. Farkas, B. Nagy, G. Marosi, Quantitative evaluation of drug distribution in tablets of various structures via Raman mapping, Period. Polytech. Chem. Eng. 62 (2018) 1-7, **https://doi.org/10.3311/PPch.10143**

[17] P.-Y. Sacré, P. Lebrun, P.-F. Chavez, C. De Bleye, L. Netchacovitch, E. Rozet, R. Klinkenberg, B. Streel, P. Hubert, E. Ziemons, A new criterion to assess distributional homogeneity in hyperspectral images of solid pharmaceutical dosage forms, Anal. Chim. Acta 818 (2014) 7-14, **https://doi.org/10.1016/j.aca.2014.02.014**

[18] H. Mitsutake, S.R. Castro, E. de Paula, R.J. Poppi, D.N. Rutledge, M.C. Breitkreitz, Comparison of different chemometric methods to extract chemical and physical information from Raman images of homogeneous and heterogeneous semi-solid pharmaceutical formulations, Int. J. Pharm. 552 (2018) 119-129, **https://doi.org/ 10.1016/j.ijpharm.2018.09.058**

[19] M.C. Hennigan, A.G. Ryder, Quantitative polymorph contaminant analysis in tablets using Raman and near infra-red spectroscopies, J. Pharm. Biomed. Anal. 72 (2013) 163-171, **https://doi.ong/10.1016/jjpba.2012.10.002**

[20] K. Lee, M. Lankers, O. Valet, J. Cherris, Raman spectroscopy for identification of contaminant materials in pharmaceuticals, Spectroscopy 31 (2016) 8-17 (**http:// www.spectroscopyonline.com/raman-spectroscopy-identification-contaminant-materials-pharmaceuticals**) (Accessed 23 July 2021).

[21] B. Vajna, I. Farkas, A. Szabó, Z. Zsigmond, G. Marosi, Raman microscopic evaluation of technology dependent structural differences in tablets containing imipramine model drug, J. Pharm. Biomed. Anal. 51 (2010) 30-38, **https://doi. org/10.1016/j.jpba.2009.07.030**

[22] A. Afrose, E.T. White, T. Howes, G. George, A. Rashid, L. Rintoul, N. Islam, Preparation of ibuprofen microparticles by antisolvent precipitation crystallization technique: characterization, formulation, and In Vitro performance, J. Pharm. Sci. 107 (2018) 3060-3069, **https://doi.org/10.1016/j.xphs.2018.07.030**

[23] W.H. Doub, W.P. Adams, J.A. Spencer, L.F. Buhse, M.P. Nelson, PJ. Treado, Raman chemical imaging for ingredient-specific particle size characterization of aqueous suspension nasal spray formulations: a progress report, Pharm. Res. 24 (2007) 934-945, **https://doi.org/10.1007/s11095-006-9211-2**

[24] S. Šašic, Raman mapping of low-content API pharmaceutical formulations. I. Mapping of Alprazolam in Alprazolam/Xanax tablets, Pharm. Res. 24 (2007) 58-65, **https://doi.org/10.1007/s11095-006-9118-y**

[25] M.J. Henson, L. Zhang, Drug characterization in low dosage pharmaceutical tablets using Raman microscopic mapping, Appl. Spectrosc. 60 (2006) 1247-1255, **https://doi.org/10.1366/000370206778998987**

[26] T. Helešicová, T. Pekárek, P. Matějka, The influence of different acquisition settings and the focus adjustment on Raman spectral maps of pharmaceutical tablets, J. Drug Deliv. Sci. Technol. 47 (2018) 386-394, **https://doi.org/10.1016/j. jddst.2018.08.002**

[27] C. Gendrin, Y. Roggo, C. Collet, Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review, J. Pharm. Biomed. Anal. 48 (2008) 533-553, **https://doi.org/10.1016/jjpba.2008.08.014**

[28] H. Nie, Z. Liu, B.C. Marks, L.S. Taylor, S.R. Byrn, P.J. Marsac, Analytical approaches to investigate salt disproportionation in tablet matrices by Raman spectroscopy and Raman mapping, J. Pharm. Biomed. Anal. 118 (2016) 328-337, **https://doi. org/10.1016/j.jpba.2015.10.024**

[29] L. Netchacovitch, E. Dumont, J. Cailletaud, J. Thiry, C. De Bleye, P.-Y. Sacré, M. Boiret, B. Evrard, P. Hubert, E. Ziemons, Development of an analytical method for crystalline content determination in amorphous solid dispersions produced by hot-melt extrusion using transmission Raman spectroscopy: a feasibility study, Int. J. Pharm. 530 (2017) 249-255, **https://doi.org/10.1016/j.ijpharm.2017.07.052**

[30] T.R.M. De Beer, C. Bodson, B. Dejaegher, B. Walczak, P. Vercruysse, A. Burggraeve, A. Lemos, L. Delattre, Y. Vander Heyden, J.P. Remon, C. Vervaet, W.R.G. Baeyens, Raman spectroscopy as a process analytical technology (PAT) tool for the in-line monitoring and understanding of a powder blending process, J. Pharm. Biomed. Anal. 48 (2008) 772-779, **https://doi.org/10.1016/jjpba.2008.07.023**

[31] M.-L. O'Connell, A.G. Ryder, M.N. Leger, T. Howley, Qualitative analysis using Raman spectroscopy and chemometrics: a comprehensive model system for narcotics analysis, Appl. Spectrosc. 64 (2010) 1109-1121, **https://doi.org/10.1366/ 000370210792973541**

[32] M. Ishigaki, Y. Maeda, A. Taketani, B.B. Andriana, R. Ishihara, K. Wongravee, Y. Ozaki, H. Sato, Diagnosis of early-stage esophageal cancer by Raman spectroscopy and chemometric techniques, Analyst 141 (2016) 1027-1033, **https:// doi.org/10.1039/C5AN01323B**

[33] A. Daniel, A. Prakasarao, S. Ganesan, Near-infrared Raman spectroscopy for estimating biochemical changes associated with different pathological conditions of cervix, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 190 (2018) 409-416, **https://doi.org/10.1016/j.saa.2017.09.014**

[34] H. Lin, O. Marjanovic, B. Lennox, S. Šašic, I.M. Clegg, Multivariate statistical analysis of Raman images of a pharmaceutical tablet, Appl. Spectrosc. 66 (2012) 272-281, **https://doi.org/10.1366/11-06238**

[35] R. Roopwani, Z. Shi, I.S. Buckner, Application of principal component analysis (PCA) to evaluating the deformation behaviors of pharmaceutical powders, J. Pharm. Innov. 8 (2013) 121-130, **https://doi.org/10.1007/s12247-013-9153-2**

[36] N.L. Calvo, R.M. Maggio, T.S. Kaufman, Characterization of pharmaceutically relevant materials at the solid state employing chemometrics methods, J. Pharm. Biomed. Anal. 147 (2018) 538-564, **https://doi.org/10.1016/jjpba.2017.06.017**

[37] N. Kumar, A. Bansal, G.S. Sarma, R.K. Rawal, Chemometrics tools used in analytical chemistry: an overview, Talanta 123 (2014) 186-199, **https://doi.org/10. 1016/j.talanta.2014.02.003**

[38] D. Spálovská, T. Pekárek, M. Kuchař, V. Setnička, Comparison of genuine, generic and counterfeit Cialis tablets using vibrational spectroscopy and statistical methods, J. Pharm. Biomed. Anal. 206 (2021) 114383, **https://doi.org/10.1016/j. jpba.2021.114383**

[39] T. Čapková-Helešicová, T. Pekárek, M. Schongut, P. Matějka, New designed special cells for Raman mapping of the disintegration process of pharmaceutical tablets, J. Pharm. Biomed. Anal. 168 (2019) 113-123, **https://doi.org/10.1016/j. jpba.2019.02.019**