



Comparison of different chemometric methods to extract chemical and physical information from Raman images of homogeneous and heterogeneous semi-solid pharmaceutical formulations

Hery Mitsutake^a, Simone R. Castro^b, Eneida de Paula^b, Ronei J. Poppi^a, Douglas N. Rutledge^c, Márcia C. Breitkreitz^{a,*}

^a Department of Analytical Chemistry, Chemistry Institute, University of Campinas (UNICAMP), Campinas, SP, Brazil

^b Department of Biochemistry and Tissue Biology, Biology Institute, University of Campinas (UNICAMP), Campinas, SP, Brazil

^c UMR Genial, AgroParisTech, INRA, Université Paris-Saclay, 91300 Massy, France

ARTICLE INFO

Keywords:

PCA
MCR-ALS
ICA
Semi-solid pharmaceutical formulations
Raman mapping

ABSTRACT

In formulations of nanostructured lipid carriers, lipid solid dispersions and self-emulsifying drug delivery systems, it is common that a solid or semi-solid lipid excipient is mixed with a liquid solvent or liquid lipid. Even when the excipients are visually miscible upon melting, they might have microscopic non-homogeneities which could lead to instability over time and future phase separation. Raman mapping associated with chemometric methods can be useful to evaluate spatial distribution of compounds, however it has not been extensively applied to the formulations mentioned above. The aim of this work was to compare the outcomes of three different chemometric methods – principal components analysis (PCA), multivariate curve resolution with alternating least squares (MCR-ALS) and independent components analysis (ICA) – to study two systems of very different degrees of microscopic miscibility: cetyl palmitate + Transcutol[®] (heterogeneous) and polyethylene glycol 6000 (PEG 6000) + Tween 80[®] (homogeneous). These two samples were chosen due to large differences in spatial distribution of the compounds over the pixels which could require different approaches for data treatment. The three methods were compared regarding recovered concentrations (or scores), signals (or loadings) and the need for matrix augmentation to obtain reliable results. Results showed that PCA loadings were the mathematical differences of the spectra of pure compounds for both samples, and therefore only ‘contrast images’ could be generated. MCR and ICA provided signals that could be related to the chemical components, however MCR presented rotational ambiguities even for the very heterogeneous sample, a situation in which ICA performed better as a blind search method. For the homogeneous sample, both methods showed rank deficiency and therefore the use of a matrix augmentation was necessary. ICA and PCA allowed identifying physical modifications in the homogeneous semi-solid PEG 6000/Tween 80[®] sample over the time, probably due to the folding/unfolding of the crystalline chains of PEG 6000. Therefore, this work discusses the ability of the three chemometrics methods to extract information from Raman spectra in order to characterize the chemical, spatial and even physical aspects of semi-solid pharmaceutical formulations, which could be of much use for stability studies of different drug delivery systems.

1. Introduction

Modern pharmaceutical development has to deal with the fact that many of the recently discovered molecules with pharmacological activities are hydrophobic and therefore present low water solubility, which is responsible for their poor and erratic bioavailability. For these compounds to become commercially marketed products, special formulation strategies need to be developed such as solid-dispersions using

hydrophilic carriers (Serajuddin, 1999), lipid-based formulations (LBF) (Pouton and Porter, 2008; Williams et al., 2013; Feeney et al., 2016), solid-lipid nanoparticles (SLN) and nanostructured lipid carriers (NLC), (Müller et al., 2000; Müller et al., 2011), to mention just a few.

In many of these formulations a solid or semi-solid excipient is melted and mixed with a liquid solvent, surfactant or co-surfactant in one step of the process. The manufacturing processes that use meltable excipients are nowadays preferred over solvent-based processes, due to

* Corresponding author at: Rua Josué de Castro, s/n Cidade Universitária, Campinas, SP 13084970, Brazil.

E-mail address: marciacristinab@iqm.unicamp.br (M.C. Breitkreitz).

<https://doi.org/10.1016/j.ijpharm.2018.09.058>

Received 23 June 2018; Received in revised form 31 August 2018; Accepted 23 September 2018

Available online 25 September 2018

0378-5173/ © 2018 Elsevier B.V. All rights reserved.

the savings of time and costs, besides being ecologically more attractive. In the case of SLN and NLC formulations, the solid/semi-solid excipient is hydrophobic while in case of solid dispersions and self-emulsifying drug-delivery systems (SEDDS) formulations it is normally hydrophilic. The final mixture can be either solid, semi-solid or liquid depending on the relative amounts of the excipients. Visual inspection of miscibility during the mixing of the molten semi-solid and liquid excipients is normally carried out. However this procedure might not be sufficient to assure real miscibility and stability. Breitzkreitz et al. showed that SEDDS formulations of atorvastatin calcium prepared with Gelucire[®] 44/14 and different solvents/co-surfactants that were visually homogeneous, presented very different miscibility behavior by Raman mapping (Breitzkreitz et al., 2013). Even if excipients are miscible upon melting, after cooling they could present micro-heterogeneities. Such inhomogeneities (as agglomerates or channels of the liquid excipient inside the waxy solid matrix) could lead to instability over time and eventually phase separation. In this sense, it is important to detect this problem in the early steps of pharmaceutical development to avoid waste of time, work and resources.

Cetyl palmitate (hexadecyl hexadecanoate, CP) is one of most commonly used solid lipids for the preparation of NLC formulations (Anantachaisilp et al., 2010; Chantaburanan et al., 2017). In this kind of formulation, the mixing of the solid and liquid excipient leads to enhanced solubility of the active principle ingredient (API) by formation of an imperfect matrix structure (Saupe et al., 2006). The NLC is the second generation of solid lipid nanoparticles. In comparison to SLN, the improved drug upload capacity and physicochemical stability of NLC has been attributed to the incorporation of a liquid component into the inner part of the waxy solid matrix (Beloqui et al., 2016).

While many papers describe the use of differential scanning calorimetry (DSC), X-ray powder diffraction (XRD), atomic force microscopy and zeta potential measures (Saupe et al., 2006; Carbone et al., 2014; Karn-Orachai et al., 2014), only a few have used Raman imaging to evaluate the miscibility of excipients (Anantachaisilp et al., 2010; Karn-Orachai et al., 2016; Suto et al., 2016). Recently we showed the miscibility of CP-based NLCs with Dhaykol 6040[®] and Capryol[®] using Raman mapping and classical least squares (CLS) where it was possible to observe the advantages of the use of multivariate methods (da Silva et al., 2017). However, CLS only works well in very limited situations (when Lambert-Beer's law is strictly followed and there are no interactions among the components). Therefore, exploring other chemometric tools should lead to a deeper understanding of this type of formulations.

Polyethylene glycol (PEG) is a hydrophilic polymer largely used in pharmaceutical development to enhance the solubility of poorly water-soluble drugs (Gullapalli and Mazzitelli, 2015; Van Duong and Van den Mooter, 2016). Depending on its molecular weight, PEG can vary from a clear liquid (200–600 g/mol) to a white solid wax (> 1000 g/mol). Above 20,000 g/mol they are called polyethylene oxides (PEOs). PEG has a safety profile for human use and its availability at low cost makes it a very interesting excipient for pharmaceutical development. Although this polymer enhances the wettability of poorly water-soluble drugs, it is not capable of preventing precipitation upon dilution in aqueous medium by itself and, for this reason, it is common to associate it with another component, normally a lipid or a surfactant. Solid polyethylene glycol has a semi-crystalline structure, forming lamellae with chains either fully extended or folded – once or twice – depending on the crystallization conditions (Verheyen et al., 2002). Unga et al. studied binary mixtures of PEG with twelve liquid lipids to understand their influence on PEG crystallization by DSC and XRD. According to their results, the properties of incorporated lipids govern the crystallization behavior of PEG, changing the ratio of folded/unfolded chains of this excipient (Unga et al., 2010).

PEG is frequently associated with polysorbate 80 (Tween[®]), a non-ionic surfactant. It is suggested that when PEG and Tween 80[®] are mixed, the surfactant is miscible only in the amorphous part of PEG

(Morris et al., 1992). Also, it has been described that the miscibility of the two compounds increases with the molecular weight of PEG (Tejwani et al., 2000). The PEG/Tween 80[®] system has demonstrated increased bioavailability for many drugs (Dannenfels et al., 2004; Joshi et al., 2004) generating stable formulations. Nevertheless, a solid dispersion of saquinavir mesylate using PEG 4000 and Tween 80[®] was reported to be unstable over time. The authors attributed this instability to modifications in the degree of crystalline order/disorder of the carrier over the time (Caon et al., 2013).

Raman mapping generates chemical and spatial information, and it has been used for the evaluation of several different delivery systems (Gordon and McGoverin 2011; Smith et al., 2015). However, very few studies focus on the use of this technique for the development of semi-solid formulations for oral delivery (Breitzkreitz et al., 2013; Sacré et al., 2015). The use of chemometric methods associated with this technique ensures that relevant information is extracted from the samples. Chemometric methods that do not require a calibration sample set in order to extract information are of particular interest, such as principal components analysis (PCA) and curve resolution methods, for example MCR-ALS (multivariate curve resolution with alternating least squares) and independent components analysis (ICA). While the first two are more common in literature, ICA is a recent method, not widely exploited in pharmaceutical development up to the moment. Using “ICA and pharmaceutical” as topics in Web of Science, 32 publications were found and only 2 using Raman imaging (Boiret et al., 2014; Lin et al., 2012). To the best of our knowledge, this is the first time that ICA is used on semi-solid formulations. It is important to highlight that even though these methods do not require a calibration sample set, data from a very homogeneous sample can present rank deficiency issues due to the lack of variability in the pixels (Lohumi et al., 2017; Tauler et al., 1995), therefore it becomes necessary to include samples with different concentration of the components, generating what is called an augmented matrix.

The aim of this work is to compare the outcomes of three different chemometric methods: principal components analysis (PCA), multivariate curve resolution with alternating least squares (MCR-ALS) and independent components analysis (ICA) to study two systems of very different degrees of microscopic miscibility: cetyl palmitate + Transcutol[®] (heterogeneous) and PEG 6000 + Tween 80[®] (homogeneous). We discuss the ability of these chemometric methods to describe the miscibility and physical stability of two very different semi-solid formulations, based on Raman mapping data. It is described how the degree of heterogeneity (pixel-to-pixel variability) affects each method and their ability to detect physical changes that could lead to instabilities even in very homogeneous semi-solid formulations. A brief description of each method is provided in the Section 2.4.

2. Materials and methods

2.1. Materials

Cetyl Palmitate was purchased by Dhaymers Química Fina (Brazil), Transcutol[®] was kindly donated by Gattefossé (France), polyethylene glycol 6000 and Tween 80[®] were purchased from Synth (Brazil). The structure of each excipient is shown in Fig. S1.

2.2. Sample preparation and Raman mapping

The samples were prepared by heating 10 °C above the melting point of CP (54 °C) and PEG 6000 (58–63 °C) and the liquid excipient (Transcutol[®] or Tween 80[®]) was added under stirring until a visually homogeneous mixture was obtained. The sample CP/Transcutol[®] was prepared to have the bulk concentration of 70% w/w of CP and 30% w/w of Transcutol[®]. Nine samples of PEG 6000 and Tween 80[®] were prepared, varying the proportions from 10 to 90% (w/w) (Table S1). The samples were cooled to room temperature in an aluminum cell and

an area of $2.0 \times 2.0 \text{ mm}$ (4 mm^2) was mapped using a Raman Station 400 (Perkin Elmer, CT, USA).

The exposure time was 3 s/pixel and each spectrum was the average of 2 exposures. The step size was $50 \mu\text{m}$ and the spectral range $600\text{--}3200 \text{ cm}^{-1}$ with a resolution of 4 cm^{-1} . Each sample generated a cube of data with dimensions of $40 \times 40 \times 651$, where 40 was the number of pixels at x and y axis and 651 the number of spectral variables. Raman images of all 9 samples of PEG 6000/Tween 80[®] were obtained at the initial time and also after 6 and 12 months.

2.3. Data processing

Spikes from Raman spectra were excluded using an algorithm developed by Sabin and co-workers (Sabin et al., 2012). The spectral range of $1804\text{--}724 \text{ cm}^{-1}$ was selected for CP/Transcutol[®] and $1724\text{--}844 \text{ cm}^{-1}$ for PEG 6000/Tween 80[®]. Dead pixels were replaced by median of neighbors (Vidal and Amigo, 2012). The data cube was unfolded to a 2D ($\text{NM} \times \lambda$) matrix, where M is number of pixels at x axis, N is the number of pixels at y axis and λ is the number of spectral variables. The spectra were smoothed using Savitzky-Golay method (width of 5, second polynomial order), baseline correction by asymmetric least squares (AsLS) and normalization by unit vector. For PCA, the data were mean centered.

2.4. Chemometric analysis

To compare the outcomes of the three chemometric methods, the data treatment was divided into three situations. Situation 1) the models were built on the matrix obtained only by unfolding the cube, without any other information. The goal was to simulate a completely blind search. Situation 2) the reference pure spectra were added to the same matrix (Fig. S2a). These two cases simulate situations when only one sample is available for analysis, and represent the most practical scenario for pharmaceutical development. Situation 3) this one was carried out only for PEG/Tween samples. A model for PEG 6000 and Tween 80[®] using the 9 samples with different concentrations and the reference pure spectra (augmented matrix) was built (Fig. S2b). Also, to analyze modifications over time, all 9 samples from the initial time, as well as 6 and 12 months samples were organized into a single augmented matrix (Fig. S2c). The data were analyzed by principal components analysis (PCA), multivariate curve resolution with alternating least squares (MCR-ALS) and independent components analysis (ICA). All models were built using Matlab version 8.3 (Mathworks Inc., Natick MA, USA). PCA data analyses were performed using the PLS_toolbox[®] version 7.3.1 (Eigenvector Research Inc., Wenatchee WA, USA). For MCR and ICA the MCR analyses it was used the MCR_toolbox version 2 and JADE algorithm (Rutledge and Bouveresse, 2013, 2015), respectively.

PCA is an exploratory method widely used in hyperspectral imaging (Amigo et al., 2008; Gendrin et al., 2008; Kandpal et al., 2015). In this method, the unfolded X matrix (matrix containing the spectra) is decomposed according to Eq. (1):

$$X = TP^T + E \quad (1)$$

where T is the scores matrix ($\text{MN} \times F$), P is the loadings matrix ($\lambda \times F$) and E is the residual matrix ($\text{MN} \times \lambda$) for F principal components (PCs). M and N represent the spatial dimensions of the cube. The objective of PCA is to reduce the dimensionality of data by finding linear combinations of the original variables, where correlated information is gathered in the same component (Bro and Smilde, 2014). The scores in each PC were re-folded to have the original spatial organization and one image was built for each significant PC.

In PCA, since the loadings are combinations of the original variables, their interpretation is not straightforward and they may lack physical or chemical meaning. On the other hand, curve resolution algorithms can provide chemically meaningful solutions, and therefore a

more straightforward interpretation of the results is possible. MCR-ALS is the most popular algorithm used for multivariate curve resolution, and it is based on Eq. (2):

$$X = CS^T + E \quad (2)$$

where C is the concentration profile ($\text{MN} \times G$), S the spectral profile ($\lambda \times G$) and E is residual matrix ($\text{MN} \times \lambda$) for G components. C and S are iteratively calculated until convergence is achieved (de Juan et al., 2014). In this work, the number of components was chosen based on singular value decomposition (SVD). This method shows the disadvantages of ambiguities, both rotational and intensity, for the estimated C and S matrices. Some constraints can be used to overcome these ambiguities, such as non-negativity of spectra and concentrations profiles, and both were used in this work. For a fair comparison with ICA, the SIMPLISMA (SIMPLE-to-use Interactive Self-modeling Mixture Analysis) algorithm was employed to provide initial estimates of the pure spectra in Situation 1.

ICA is a method of Blind Source Separation (BSS) developed in the field of signal processing in telecommunications (Jutten and Herault, 1991) to extract pure underlying signals of a mixed signals set according to Eq. (3):

$$X = AS_s \quad (3)$$

where S_s is a matrix of independent source signals ($k \times \lambda$) and A is a matrix of mixing coefficients or proportions of the pure signals with dimensions of ($n \times k$), k is the number of independent components (ICs) (Rutledge and Bouveresse, 2013, 2015). ICA assumes that the original source signals and their proportions in the analyzed mixtures are unknown, and it aims at extracting them by using the criterion of maximum independence among the source signals.

The source signals, or ICs, are analogous to PCA loadings and the corresponding proportions (also called “mixing coefficients”) are analogous to PCA scores. Source signals are assumed to have a specific structure, which means that their intensity distribution is not random, and therefore, does not give a Gaussian histogram. On the other hand, because of the Central Limit Theorem, mixtures of non-Gaussian source signals will tend to be more Gaussian. Therefore, ICA aims to find the source signals by maximization of their non-Gaussianity (Mishra et al., 2016; Stone, 2002; Wang et al., 2008). ICA is a more recent method and it hasn't been largely exploited in the pharmaceutical area yet. Two criteria were used to choose the optimal number of ICs: correlation with excipient spectra or ICA_by_blocks (Bouveresse et al., 2012).

3. Results and discussion

3.1. Raman spectra

Raman spectra of the two systems are shown in Fig. 1. Cetyl palmitate presents main peaks at 1728, 1460, 1440, 1420, 1296, 1132, 1100, 1064, 1016, 924 and 892 cm^{-1} . Transcutol[®] shows Raman peaks on 1460, 1276, 1244, 1140, 1072, 888, 840 and 808 cm^{-1} . Bands between 800 and 900 cm^{-1} refer mainly to $-\text{CH}_3$ rocking. Peaks at 1016 and 1072 cm^{-1} refer to $-\text{CH}_3$ bonds. Peaks around 1440–1450 represent mainly $-\text{CH}_2$ bonds while the $-\text{C}=\text{O}$ bond in esters is found as a weak signal at 1728 cm^{-1} for CP. PEG 6000 exhibits peaks on 1480, 1448, 1396, 1364, 1280, 1236, 1144, 1124, 1064, 948, 932 and 860 cm^{-1} , and Tween 80[®] has peaks on 1652, 1464, 1444, 1296, 1284, 1248, 1136, 1064, 1040 cm^{-1} . Another band at 1064 cm^{-1} , related to $-\text{CH}_3$ bonds, is found in CP, PEG 6000 and Tween 80[®] (Patnaik, 2004).

3.2. Cetyl palmitate and Transcutol[®]

Chemometric models were built without (Situation 1) and with (Situation 2) the reference spectra. PCA did not show any significant differences due to the huge heterogeneity of this sample, as shown by the Scores plot (Fig. 2a) where a wide dispersion of values among the

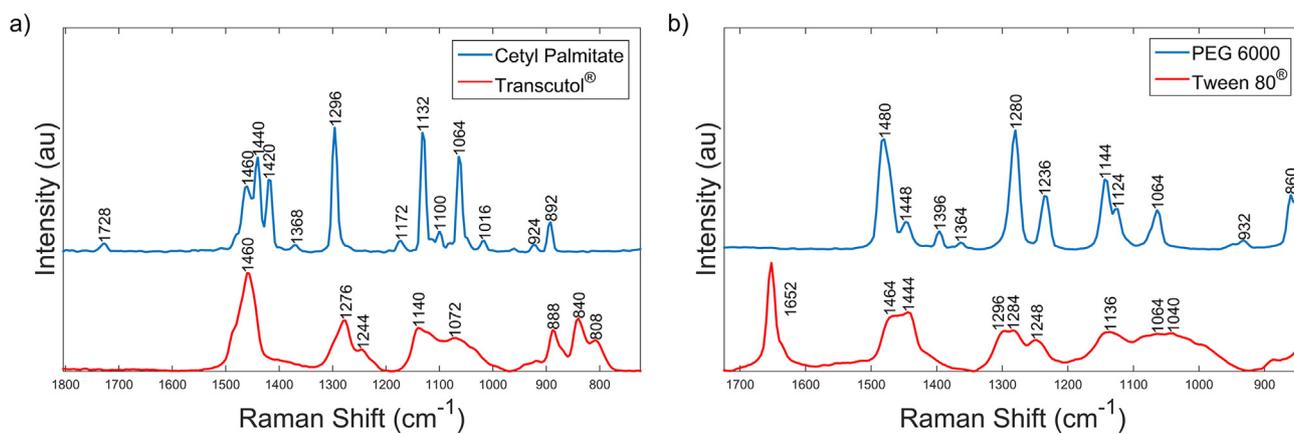


Fig. 1. Raman spectra of a) cetyl palmitate and Transcutol[®] and b) PEG 6000 and Tween 80[®].

pixels is observed. Since the principal components are calculated as the directions of maximum variability of the data set, which in this case are the differences of concentration of excipients, the loadings are the differences between the pure spectra. Fig. 2b compares the loadings in

the first PC (82.88% of variance) with the calculated difference between the pure spectra. As previously mentioned, PCA loadings refer to mixed signals. So, they do not give information on the pure compounds but provide a ‘contrast’ image instead. By re-folding PC1 scores, the

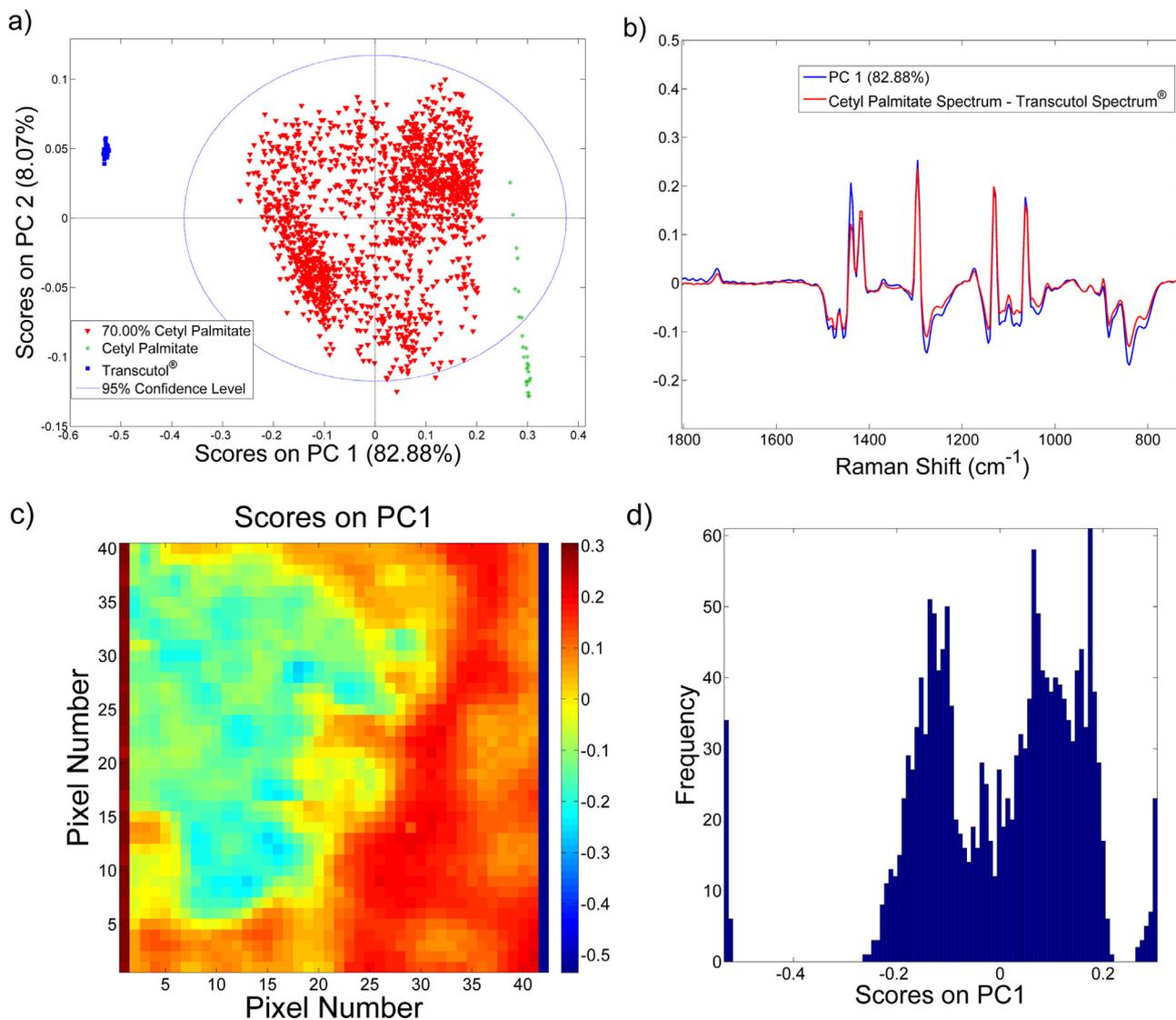


Fig. 2. a) Score plot of image pixels and reference spectra; b) Loadings on PC1 and the mathematical difference of the two reference spectra; c) contrast image of PCA scores and d) histogram of the scores.

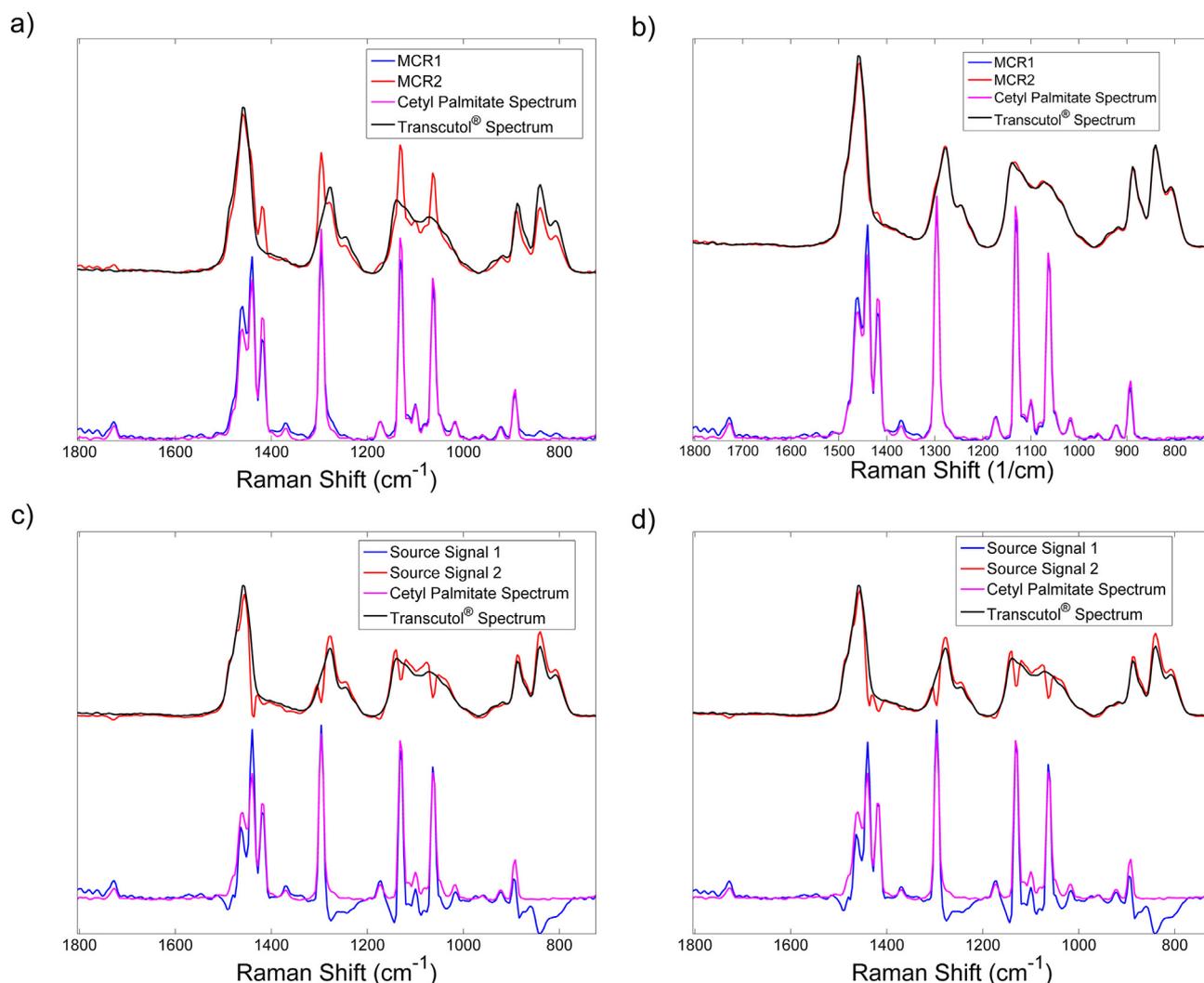


Fig. 3. Spectral profiles (S) recovered for CP/Transcutol[®] sample by MCR-ALS: a) with and b) without pure excipient spectra. Source signals recovered by ICA: c) with and d) without pure excipient spectra.

chemical map (Fig. 2c) is obtained. It can be seen that Transcutol[®] was accumulated in one part of image and revealed poor miscibility with CP. When this matrix is refolded, the scores of reference spectra are the first (CP) and last (Transcutol[®]) lines of the chemical image with the highest and lowest scores values, which gives an intense red and blue color for these lines.

However, even for this heterogeneous sample there are no pixels with only pure excipient spectra, as can be seen in Fig. 2a. This was a problem for MCR-ALS using SIMPLISMA, as shown in Fig. 3a. When the model was built using only one sample spectra (Situation 1), the recovered spectral profile of Transcutol[®] presented mixed peaks of CP (see for example 1430–1410, 1300–1290 and 1150–1050 cm^{-1}). This is called rotational ambiguity and to overcome this issue, the variability of the matrix should be increased even more. The easiest way of doing that is by adding the pure reference spectra (Situation 2, Fig. 3b). For ICA, the presence of pure excipient spectra did not influence the recovery of the source signals (Fig. 3c and d), i.e. in both situations, ICA recovered well CP and Transcutol[®] spectra without rotational ambiguity. Therefore ICA was found to be a good option as a blind search method. BSS implies that any a priori information is used to build a model. In other words, on the contrary of MCR-ALS which needs initial estimates, ICA can be performed without them. One potential use of ICA, thus, is in analysis of unknown composition of samples (Boiret et al., 2014). Nevertheless, it should be pointed out that in the case of

MCR-ALS, if reference spectra had been used as initial estimates, the rotational ambiguity problem should be minimized. Since this was not an option for ICA, the comparison in this case would not have been straightforward and therefore it was not discussed.

Chemical images generated by MCR-ALS and ICA, including the reference pure excipient spectra were similar, and therefore Fig. 4 shows only the ICA images. The proportions were normalized using the reference spectra as 0 and 100. For example, IC1 refers to cetyl palmitate thus, pixels with pure spectra of this excipient have values of 100 and pixels with pure spectra of Transcutol[®] have values of 0 (complete absence of CP). Also, the inverse was done for IC2: pixels of Transcutol[®] have values of 100 and pixels of excipient CP have values of zero. Using this normalization, proportions obtained from different formulations have the same scale of variation. These normalized proportions are very useful because they allow to compare samples with different excipients and predict the more homogenous formulation amongst them. Fig. 4a and b show CP and Transcutol[®] chemical images, respectively; and Fig. 4c and d the corresponding histograms.

This sample was observed to be heterogeneous even at the initial time by Raman mapping and chemometrics. The NLC formulation using these excipients was prepared to encapsulate the tetracaine drug, using Poloxamer as a surfactant and instabilities of particle size and polydispersity were observed after 60 days, which finally resulted in phase separation. This result highlights the power of this technique to foresee

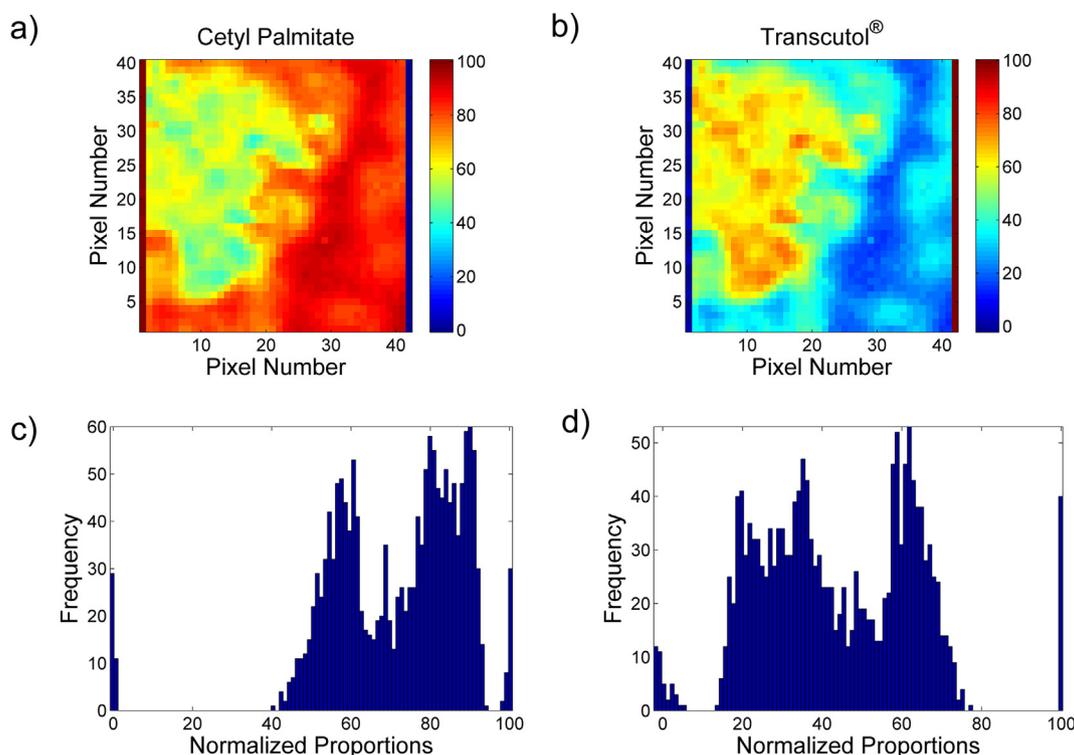


Fig. 4. Chemical images obtained by ICA: a) CP and b) Transcutol[®]; histograms for c) CP and d) Transcutol[®].

stability issues during pharmaceutical development.

3.3. PEG 6000 and Tween 80[®]

3.3.1. Initial time results

PCA analysis showed that this sample is very homogeneous in the whole concentration range. Without any reference spectra, the variance captured by each Principal Component was very low, due to the lack of a preferential direction in space. SVD indicated only one component for MCR-ALS model and the recovered spectrum was the average spectrum of the image. Similar results would be found if two components were chosen. With ICA, the source signals were mixtures of the reference spectra and did not refer to any of the excipients (Fig. S3). Therefore, the use of alternatives that increase data variability is mandatory in this case, and the simplest solution was augmenting the matrix with the reference spectra, which were available in this case. This approach provided a direction of preferential variability for PCA. Results of the sample containing 30.0% (w/w) of Tween 80[®] and 70.0% of PEG 6000 are shown in Fig. 5 as an example. The scores plot of Fig. 5a shows that the variability among the pixels was small along the first PC, a very different situation than the one found for the heterogeneous sample (Fig. 2a). The scores of reference spectra are shown in green and blue in Fig. 5a where it can be seen that all the image pixels are concentrated between the scores of the two excipients. This implies that this sample does not have any pixels with only one of excipients, and therefore the two compounds are well mixed. Also, similarly to the heterogeneous case, the first PC loadings refer to the difference between the two excipient spectra (Fig. 5b), therefore, once again, the scores map (Fig. 5c) is a ‘contrast image’. Due to the homogeneity this sample, this chemical map seems monochromatic, and reference excipient scores are the first and last line of the image. The high homogeneity of this sample is confirmed by its very narrow histogram shown in Fig. 5d.

The augmenting of the matrix just with the reference spectra allowed to recover the spectral profile (MCR) and source signals (ICA), as can be seen in Fig. 6a and b. The maps were very similar for the two methods. Fig. 6c to f show chemical maps and histograms obtained by

ICA. Comparing with the heterogeneous case (Fig. 4), the differences between the two systems are evident, with histograms thinner and lower dispersion of proportions values.

It is noteworthy that the scales of scores are different in each case: PCA scores varied between -0.6 and 0.2 , centered in zero (Fig. 5). MCR-ALS gives the C matrix as a solution of Eq. (2) and varies between 0 and 1. In our data, these values were multiplied by 100 to give percentages. ICA proportions were normalized between 0 and 100 using reference spectra as extreme values. It is important to highlight that images of homogeneous/heterogeneous samples should be compared using the same scale.

Then the approach of augmenting the matrix with samples of different concentrations was evaluated (Situation 3, Fig. S2b). Using 2 components in MCR-ALS and ICA (as suggested by SVD and correlation with reference spectra, respectively), the recovered spectral profiles/source signals were very similar to the ones obtained using only one sample and the reference spectra (shown in Fig. 6a and b). These similar results indicated that augmenting only with the reference spectra was enough to create the necessary variability to solve the rank deficiency problem. This is an important aspect for semi-solid formulations because, even in cases of heterogeneous samples (e.g. CP/Transcutol[®]), it is possible that there are no ‘pure’ pixels. Nevertheless, the availability of different concentrations in the case of PEG/Tween[®] system made it possible to plot the predicted vs experimental concentrations, as shown in Fig. S4 for MCR and ICA. Both presented good agreement of the experimental average concentration with the extracted scores/proportions. However, it can be seen that there is a deviation from linearity for PEG 6000. The same result was obtained by PCA and MCR scores, therefore it was not related to the chemometric method used. One possible explanation for this is the physical aspect of samples: in one extreme there are PEG 6000 and CP, solid excipients; in the other there are Tween 80[®] and Transcutol[®], liquid excipients, and in between there are several semi-solid formulations, varying from hard to soft wax. Therefore, the scattering can be very different all over the concentration range and cause small deviations from linearity. Indeed, the same curvature was observed for the CP/Transcutol[®] sample – even though

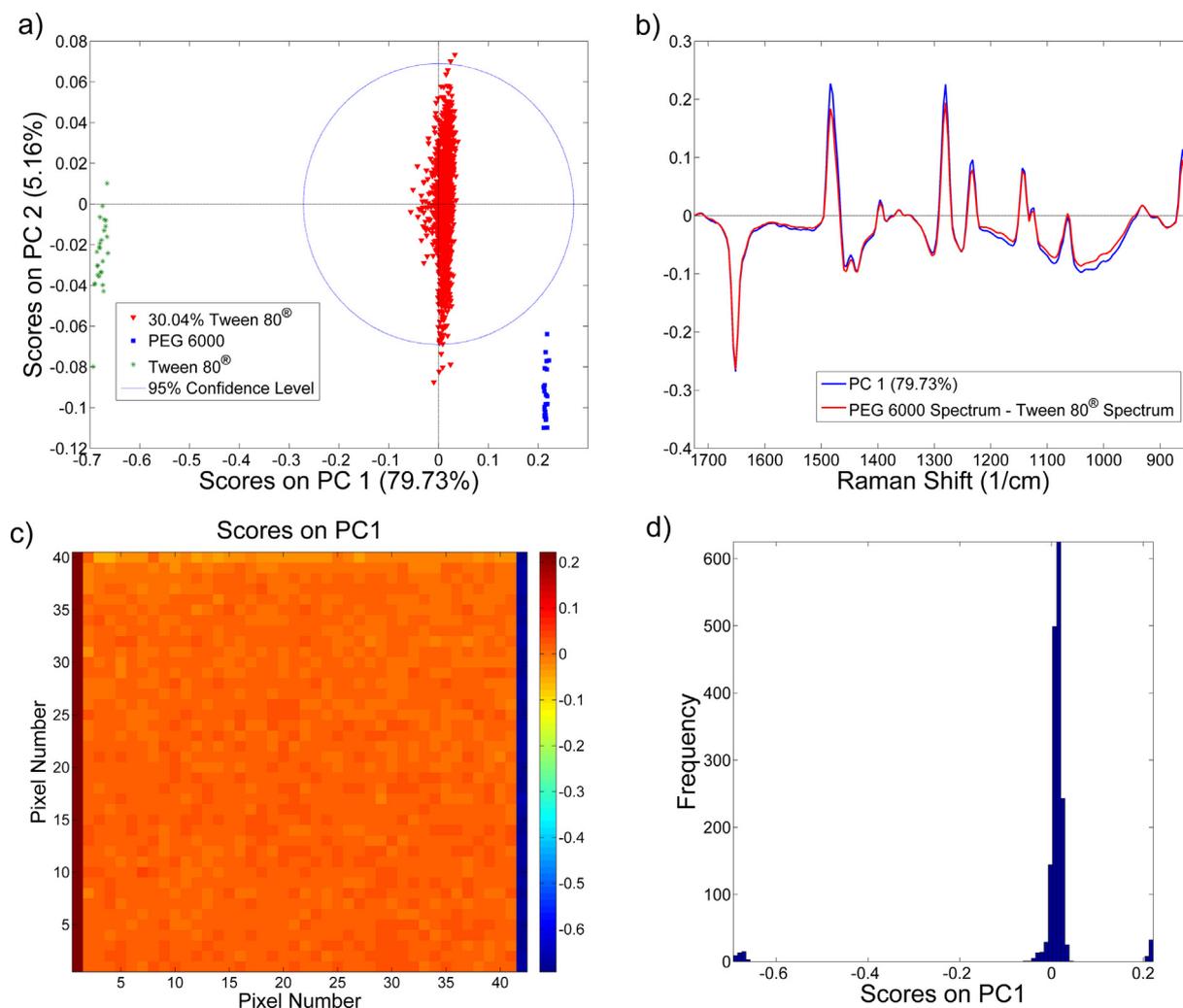


Fig. 5. PCA results of sample containing 30.04% (w/w) of Tween 80° and 70% of PEG 6000, including the reference spectra to augment the image matrix: a) Score plot on PC1 vs PC2, b) Comparison of loadings on PC1 with the difference between the two excipient spectra, c) chemical map of scores on PC1, d) histogram of the scores.

there was a single sample in this case, the wide range of concentration over the pixels made it possible to study the linearity.

3.3.2. Stability studies

The Raman images of the 9 mixtures kept at room temperature were obtained in fresh samples and after 6 months and 1 year, totaling 27 data cubes. First, an ICA model was built for each time (initial, 6 months and 1 year) using a matrix augmented with reference spectra. By the ICA-by-blocks method, the ideal number of ICs for the different mixtures increased over the time, implying that new sources of variation were appearing in the data with time. Therefore, to describe the ageing effects and the behavior of the samples over time, the spectra of all 27 samples were organized into a single augmented matrix, and an ICA model using 3ICs was built.

Fig. 7a to c show the source signals obtained. IC1 refers to PEG 6000, IC2 refers to Tween 80° and IC3 presents mixed signals of the two excipients (Fig. 7c). Fig. 7d and e display the experimental concentrations of each excipients vs the proportions of IC1 and 2 for the samples of the initial time, 6 months and 1 year. As can be seen, IC 1 and 2 signals can be straightforwardly assigned to the variations in the concentrations of PEG 6000 and Tween 80°, respectively. However, over the time a broader variation among the samples over the time is observed for PEG (Fig. 7d) compared to Tween 80° (Fig. 7e). IC3 in the initial time was related to the concentration of PEG 6000 in the samples

(blue circles of Fig. 7f), however, this signal has changed over time (red and green circles of Fig. 7f) until it lost the correlation with concentration after 1 year. This implies that modifications occurred in these samples over time so that IC3 may be describing a physical effect: modifications of PEG 6000, in the presence of Tween 80°. The presence of a liquid compound in the formulation increases the mobility of polymer chains, which could explain the broader variability of samples at higher concentrations of Tween 80° (Fig. 7d). Caon and co-workers (Caon et al., 2013) stabilized PEG 4000/Tween 80° formulations by adding polyvinylpyrrolidone (PVP) which reduced the chain mobility of PEG in the solid dispersions. Unga and co-workers (Unga et al., 2010) using DSC have shown that the presence of lipids in solid dispersions influences the folding and unfolding of PEG 4000. The unfolding rate of PEG seems to be related to the stability of formulations and some lipids can accelerate or decrease this process. In general, they observed that small hydrophilic lipids increased the folding of PEG on cooling whereas large non-polar lipids retarded the PEG unfolding.

These modifications of PEG chains were on a molecular scale and did not provoke phase separation, as shown by chemical images, which remained homogeneous over time. Chemical maps of each IC over the time for samples containing 20% of PEG 6000 and 80% (w/w) of Tween 80° are shown in Fig. 8. It can be seen that the distributions of PEG 6000 and Tween 80° is homogeneous for one year in this sample (Fig. 8a–f). However, IC3 signals clearly change over the time

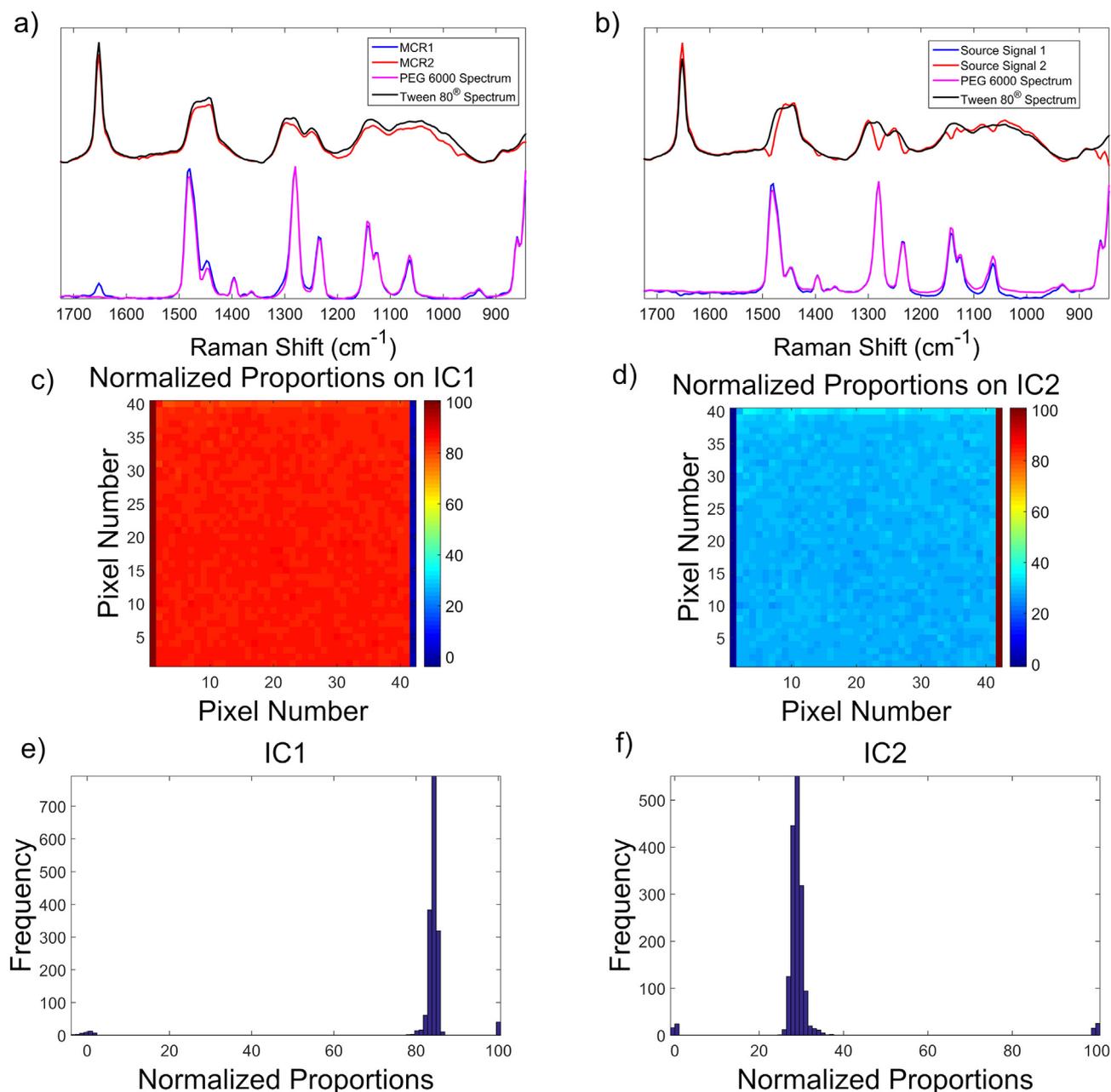


Fig. 6. a) Spectral profile obtained by MCR-ALS and b) source signals obtained by ICA using 2 components at the initial time. Chemical images for: c) IC1 and d) IC2, histograms obtained by ICA for PEG 6000/Tween 80[®] sample on e) IC1 and f) IC2.

(Fig. 8g–i), indicating that IC3 represents a kinetic process that ceased after some time, as can also be seen in Fig. 7f (more negative values and without any difference between the samples after one year). Also, DSC analysis was done on a sample at the initial time and after two years (Fig. S5). The melting point of the sample was modified (59.29 to 55.75 °C), indicating that PEG structure was indeed changed over time.

This variation over the time was also observed by PCA. Whereas PC1 describes the variability in the concentrations (Fig. S6a shows that PC1 scores increase with the concentration of Tween 80[®]), PC2 describes another phenomenon that is taking place in the samples, as can be seen in Fig. S6b with samples after one year presenting higher scores than the earlier times. Since PCA provides loadings instead of signals, the interpretation of PC2 in Fig. S6c is not straightforward. Finally, MCR-ALS only recovered 2 signals related to the chemical components, therefore this method could not provide any information on the structural changes of the samples over the time. This highlights an

interesting use of ICA: the separation of physical phenomena from chemical information, as demonstrated in this work.

4. Conclusions

Observing the results for the heterogeneous sample (CP/Transcutol[®]), ICA outperformed MCR as a blind resolution method. However for the very homogeneous sample (PEG 6000/Tween 80[®]) both suffered from the problem of rank deficiency. It was shown that augmenting the matrix with just the pure reference spectra or several calibration samples provided the same results. Therefore the simplest procedure of using only the reference spectra should be preferred. The plot of different concentrations showed small deviations from linearity, mainly for PEG 6000, probably due to the changes of scattering over the concentration range. In both cases, PCA recovered the differences of the two pure spectra in the loadings, indicating that loadings cannot be

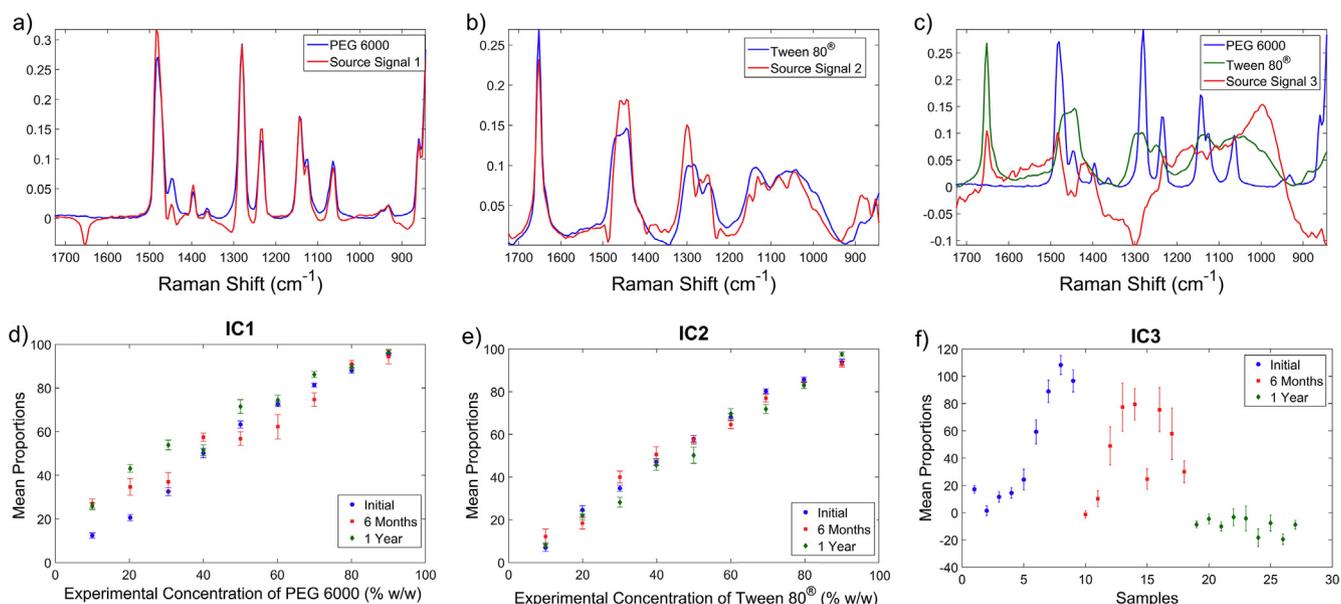


Fig. 7. Source signals on: a) IC1, b) IC2 and c) IC3. Proportions vs. concentrations on d) IC1, e) IC2 and f) Proportions on IC3 for samples of PEG 6000/Tween 80[®] over the time.

interpreted as pure signals, as sometimes found in literature. Therefore PCA method provides ‘contrast’ images only.

The major contribution of this work was to show how Raman chemical images and chemometrics can be valuable to evaluate chemical miscibility of components in semi-solid formulations and predict their stability and physical transformations. In the case of CP/Transcutol[®] which chemical images indicated heterogeneity the formulation indeed

presented phase separation after 6 months. The mixture PEG 6000/Tween 80[®] was observed to be homogeneous in the beginning, and even after one year no phase separation was observed. However, PCA and ICA indicated that a physical transformation was occurring within the sample over the time. ICA could separate this information in a third IC, which clearly showed to be a kinetic phenomenon. IC1 and 2 were related to the concentration of the two compounds over the complete

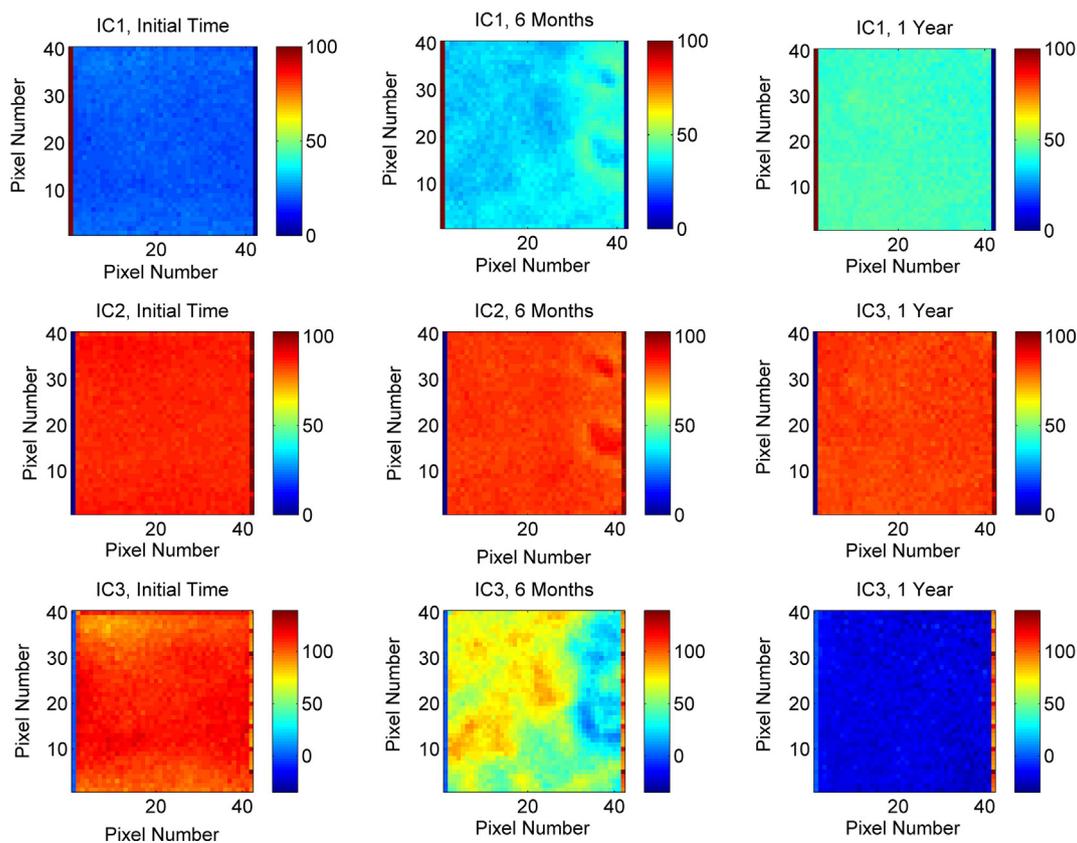


Fig. 8. Chemical maps of the sample containing 20% of PEG 6000 and 80% (w/w) of Tween 80[®]. a) IC1 at the initial time, b) IC1 after 6 months, IC1 after 1 year, c) IC2 at initial time, e) IC 2 after 6 months, f) IC2 after 1 year, g) IC3 at initial time, h) IC3 after 6 months and i) IC3 after 1 year.

time range. Our results suggested that the presence of Tween 80[®] has great influence the physical stability of PEG 6000, which could be deleterious for a pharmaceutical formulation, as it has been described by some authors that have investigated it by DSC. MCR-ALS could not detect these physical changes that were occurring in the sample.

Further research on of lipid pharmaceutical formulations is being conducted in our laboratory, aiming at understanding miscibility, physical transformations and chemical interactions for the development of stable formulations for poorly water soluble drugs.

Declaration of interest

The authors declare they have no conflict of interest.

Acknowledgments

The authors thank FAPESP (2016/05636-9), CNPq (H.M. scholarship) and INCT-Bioanalitica (FAPESP/INCT 14/50867-3; CNPq/INCT 465389/2014-7) for financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijpharm.2018.09.058>.

References

- Amigo, J.M., Cruz, J., Bautista, M., MasPOCH, S., Coello, J., Blanco, M., 2008. Study of pharmaceutical samples by NIR chemical-image and multivariate analysis. *Trend. Anal. Chem.* 27, 696–713. <https://doi.org/10.1016/j.trac.2008.05.010>.
- Anantachaisilp, S., Smith, S.M., Treetong, A., Pratontep, S., Puttipipatkachorn, S., Ruktanonchai, U.R., 2010. Chemical and structural investigation of lipid nanoparticles: drug-lipid interaction and molecular distribution. *Nanotechnology* 21, 125102. <https://doi.org/10.1088/0957-4484/21/12/125102>.
- Beloqui, A., Solinís, M.A., Rodríguez-Gascón, A., Almeida, A.J., Prêat, V., 2016. Nanostructured lipid carriers: promising drug delivery systems for future clinics. *Nanomedicine* 12, 143–161. <https://doi.org/10.1016/j.nano.2015.09.004>.
- Boiret, M., Rutledge, D.N., Gorretta, N., Ginot, Y.-M., Roger, J.-M., 2014. Application of independent component analysis on Raman images of a pharmaceutical drug product: pure spectra determination and spatial distribution of constituents. *J. Pharm. Biomed. Anal.* 90, 78–84. <https://doi.org/10.1016/j.jpba.2013.11.025>.
- Bouveresse, D.J.R., Moya-González, A., Ammari, F., Rutledge, D.N., 2012. Two novel methods for the determination of the number of components in independent components analysis models. *Chemometr. Intell. Lab.* 112, 24–32. <https://doi.org/10.1016/j.chemolab.2011.12.005>.
- Breitkreitz, M.C., Sabin, G.P., Polla, G., Poppi, R.J., 2013. Characterization of semi-solid self-emulsifying drug delivery systems (SEDDS) of atorvastatin calcium by Raman image spectroscopy and chemometrics. *J. Pharm. Biomed. Anal.* 73, 3–12. <https://doi.org/10.1016/j.jpba.2012.03.054>.
- Bro, R., Smilde, A.K., 2014. Principal component analysis. *Anal. Methods.* 6, 2812–2831. <https://doi.org/10.1039/C3AY41907J>.
- Caon, T., König, R.A., da Cruz, A.C., Cardoso, S.G., Campos, C.E., Cuffini, S.L., Koester, L.S., Simões, C.M., 2013. Development and physicochemical characterization of saquinavir mesylate solid dispersions using Gelucire 44/14 or PEG 4000 as carrier. *Arch. Pharm. Res.* 36, 1113–1125. <https://doi.org/10.1007/s12272-013-0142-2>.
- Carbone, C., Campisi, A., Manno, D., Serra, A., Spatuzza, M., Musumeci, T., Bonfanti, R., Puglisi, G., 2014. The critical role of didodecyltrimethylammonium bromide on physico-chemical, technological and biological properties of NLC. *Colloid. Surface. B.* 121, 1–10. <https://doi.org/10.1016/j.colsurfb.2014.05.024>.
- Chantaburanan, T., Teeranachaidekul, V., Chantasant, D., Jintapattanakit, A., Junyaprasert, V.B., 2017. Effect of binary solid lipid matrix of wax and triglyceride on lipid crystallinity, drug-lipid interaction and drug release of ibuprofen-loaded solid lipid nanoparticles (SLN) for dermal delivery. *J. Colloid. Interf. Sci.* 504, 247–256. <https://doi.org/10.1016/j.jcis.2017.05.038>.
- Da Silva, G.H., Ribeiro, L.N.M., Mitsutake, H., Guilherme, V.A., Castro, S.R., Poppi, R.J., Breitkreitz, M.C., de Paula, E., 2017. Optimised NLC: a nanotechnological approach to improve the anaesthetic effect of bupivacaine. *Int. J. Pharm.* 529, 253–263. <https://doi.org/10.1016/j.ijpharm.2017.06.066>.
- Dannenfelser, R.M., He, H., Joshi, Y., Bateman, S., Serajuddin, A.T., 2004. Development of clinical dosage forms for a poorly water soluble drug I: application of polyethylene glycol-polysorbate 80 solid dispersion carrier system. *J. Pharm. Sci.* 93, 1165–1175. <https://doi.org/10.1002/jps.20044>.
- De Juan, A., Jaumot, J., Tauler, R., 2014. Multivariate curve resolution (MCR) solving the mixture analysis problem. *Anal. Methods* 6, 4964–4976. <https://doi.org/10.1039/C4AY00571F>.
- Feeney, O.M., Crum, M.F., McEvoy, C.L., Trevaskis, N.L., Williams, H.D., Pouton, C.W., Charman, W.N., Bergstrom, C.A.S., Porter, C.J.H., 2016. 50 years of oral lipid-based formulations: provenance, progress and future perspectives. *Adv. Drug Delivery Rev.* 101, 167–194. <https://doi.org/10.1016/j.addr.2016.04.007>.
- Gendrin, C., Roggo, Y., Collet, C., 2008. Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review. *J. Pharm. Biomed. Anal.* 48, 533–553. <https://doi.org/10.1016/j.jpba.2008.08.014>.
- Gordon, K.C., McGoverin, C.M., 2011. Raman mapping of pharmaceuticals. *Int. J. Pharm.* 417, 151–162. <https://doi.org/10.1016/j.ijpharm.2010.12.030>.
- Gullapalli, R.P., Mazzitelli, C.L., 2015. Polyethylene glycols in oral and parenteral formulations – a critical review. *Int. J. Pharm.* 496, 219–239. <https://doi.org/10.1016/j.ijpharm.2015.11.015>.
- Joshi, H.N., Tejwani, R.W., Davidovich, M., Sahasrabudhe, V.P., Jemal, M., Bathala, M.S., Varia, S.A., Serajuddin, A.T.M., 2004. Bioavailability enhancement of a poorly water-soluble drug by solid dispersion in polyethylene glycol-polysorbate 80 mixture. *Int. J. Pharm.* 269, 251–258. <https://doi.org/10.1016/j.ijpharm.2003.09.002>.
- Jutten, C., Herault, J., 1991. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal. Process.* 1991 (24), 1–10. [https://doi.org/10.1016/0165-1684\(91\)90079-X](https://doi.org/10.1016/0165-1684(91)90079-X).
- Kandpal, L.M., Park, E., Tewari, J., Cho, B.K., 2015. Spectroscopic techniques for non-destructive quality inspection of pharmaceutical products: a review. *J. Biosyst. Eng.* 40, 394–408. <https://doi.org/10.5307/JBE.2015.40.4.394>.
- Karn-Orachai, K., Smith, S.M., Phunpee, S., Treethong, A., Puttipipatkachorn, S., Pratontep, S., Ruktanonchai, U.R., 2014. The effect of surfactant composition on the chemical and structural properties of nanostructured lipid carriers. *J. Microencapsul.* 31, 609–618. <https://doi.org/10.3109/02652048.2014.911374>.
- Karn-Orachai, K., Smith, S.M., Saesoo, S., Treethong, A., Puttipipatkachorn, S., Pratontep, S., Ruktanonchai, U.R., 2016. Surfactant effect on the physicochemical characteristics of γ -oryanol-containing solid lipid nanoparticles. *Colloid. Surface. A* 488, 118–128. <https://doi.org/10.1016/j.colsurfa.2015.10.011>.
- Lin, H., Marjanović, O., Lennox, B., Šašić, S., Clegg, I.M., 2012. Multivariate statistical analysis of Raman images of a pharmaceutical tablet. *Appl. Spectrosc.* 66, 272–281. <https://doi.org/10.1366/11-06238>.
- Lohumi, S., Kim, M.S., Qin, J., Cho, B.K., 2017. Raman imaging from microscopy to macroscopy: quality and safety control of biological materials. *Trend. Anal. Chem.* 93, 183–198. <https://doi.org/10.1016/j.trac.2017.06.002>.
- Mishra, P., Cordella, C.B.Y., Rutledge, D.N., Barreiro, P., Roger, J.M., Diezma, B., 2016. Application of independent component analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration. *J. Food. Eng.* 168, 7–15. <https://doi.org/10.1016/j.jfoodeng.2015.07.008>.
- Morris, K.R., Knipp, G.T., Serajuddin, A.T.M., 1992. Structural properties of polyethylene glycol-polysorbate 80 mixture, a solid dispersion vehicle. *J. Pharm. Sci.* 81, 1185–1188. <https://doi.org/10.1002/jps.2600811212>.
- Müller, R.H., Mader, K., Gohla, S., 2000. Solid lipid nanoparticles (SLN) for controlled drug delivery: a review of the state of the art. *Eur. J. Pharm. Biopharm.* 50, 161–177. [https://doi.org/10.1016/S0939-6411\(00\)00087-4](https://doi.org/10.1016/S0939-6411(00)00087-4).
- Müller, R.H., Shegokar, R., Keck, C.M., 2011. 20 years of lipid nanoparticles (SLN & NLC): present state of development & industrial applications. *Curr. Drug Discov. Technol.* 8, 207–227. <https://doi.org/10.2174/157016311796799062>.
- Patnaik, P., 2004. *Infrared and Raman spectroscopy. Dean's Analytical Chemistry Handbook, second ed. McGraw-Hill Handbooks.*
- Pouton, C.W., Porter, C.J., 2008. Formulation of lipid-based delivery systems for oral administration: materials, methods and strategies. *Adv. Drug. Delivery Rev.* 60, 625–637. <https://doi.org/10.1016/j.addr.2007.10.010>.
- Rutledge, D.N., Bouveresse, D.J.R., 2013. Independent components analysis with the JADE algorithm. *Trend. Anal. Chem.* 50, 22–32. <https://doi.org/10.1016/j.trac.2013.03.013>.
- Rutledge, D.N., Bouveresse, D.J.R., 2015. Corrigendum to 'Independent Components Analysis with the JADE algorithm'. *Trend. Anal. Chem.* 67, 220. <https://doi.org/10.1016/j.trac.2015.02.001>.
- Sabin, G.P., de Souza, A.M., Breitkreitz, M.C., Poppi, R.J., 2012. Desenvolvimento de um algoritmo para identificação e correção de spikes em espectroscopia Raman de imagem. *Quim. Nova* 35, 612–615. <https://doi.org/10.1590/S0100-40422012000300030>.
- Serajuddin, A.T.M., 1999. Solid dispersion of poorly water-soluble drugs: early promises, subsequent problems, and recent breakthroughs. *J. Pharm. Sci.* 88, 1058–1066. <https://doi.org/10.1021/js9804031>.
- Saupe, A., Gordon, K.C., Rades, T., 2006. Structural investigations on nanoemulsions, solid lipid nanoparticles and nanostructured lipid carriers by cryo-field emission scanning electron microscopy and Raman spectroscopy. *Int. J. Pharm.* 314, 56–62. <https://doi.org/10.1016/j.ijpharm.2006.01.022>.
- Sacré, P.Y., Netchacovitch, L., De Bleye, C., Chavez, P.F., Servais, C., Klinkenberg, R., Streel, B., Hubert, P., Ziemons, E., 2015. Thorough characterization of a self-emulsifying drug delivery system with Raman hyperspectral imaging: a case study. *Int. J. Pharm.* 484, 85–94. <https://doi.org/10.1016/j.ijpharm.2015.02.052>.
- Smith, G.P.S., McGoverin, C.M., Fraser, S.J., Gordon, K.C., 2015. Raman imaging of drug delivery systems. *Adv. Drug. Deliver. Rev.* 89, 21–41. <https://doi.org/10.1016/j.addr.2015.01.005>.
- Stone, J.V., 2002. Independent component analysis: an introduction. *Trend. Cogn. Sci.* 6, 59–64. [https://doi.org/10.1016/S1364-6613\(00\)01813-1](https://doi.org/10.1016/S1364-6613(00)01813-1).
- Suto, B., Berko, S., Kozma, G., Kukovec, A., Budai-Szucs, M., Eros, G., Kemeny, L., Sztjokov-Ivanov, A., Gaspar, R., Csanyi, E., 2016. Development of ibuprofen-loaded nanostructured lipid carrier-based gels: characterization and investigation of in vitro and in vivo penetration through the skin. *Int. J. Nanomed.* 11, 1201–1212. <https://doi.org/10.2147/IJN.S99198>.
- Tauler, R., Smilde, A., Kowalski, B., 1995. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometr.* 9, 31–58. <https://doi.org/10.1002/cem.1180090105>.

- Tejwani, R.W., Joshi, H.N., Varia, S.A., Serajuddin, A.T.M., 2000. Study of phase behavior of poly(ethylene glycol)–polysorbate 80 and poly(ethylene glycol)–polysorbate 80–water mixtures. *J. Pharm. Sci.* 89, 946–950. [https://doi.org/10.1002/1520-6017\(200007\)89:7<946::AID-JPS12>3.0.CO;2-2](https://doi.org/10.1002/1520-6017(200007)89:7<946::AID-JPS12>3.0.CO;2-2).
- Unga, J., Matsson, P., Mahlin, D., 2010. Understanding polymer-lipid solid dispersions—the properties of incorporated lipids govern the crystallisation behaviour of PEG. *Int. J. Pharm.* 386, 61–70. <https://doi.org/10.1016/j.ijpharm.2009.10.049>.
- Van Duong, T., Van den Mooter, G., 2016. The role of the carrier in the formulation of pharmaceutical solid dispersions. Part I: crystalline and semi-crystalline carriers. *Expert Opin. Drug. Del.* 13, 1583–1594. <https://doi.org/10.1080/17425247.2016.1198768>.
- Verheyen, S., Bleton, N., Kinget, R., Van den Mooter, G., 2002. Mechanism of increased dissolution of diazepam and temazepam form polyethylene glycol 6000 solid dispersions. *Int. J. Pharm.* 249, 45–58. [https://doi.org/10.1016/S0378-5173\(02\)00532-X](https://doi.org/10.1016/S0378-5173(02)00532-X).
- Vidal, M., Amigo, J.M., 2012. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometr. Intell. Lab.* 117, 138–148. <https://doi.org/10.1016/j.chemolab.2012.05.009>.
- Wang, G., Ding, Q., Hou, Z., 2008. Independent component analysis and its applications in signal processing for analytical chemistry. *Trend. Anal. Chem.* 27, 368–376. <https://doi.org/10.1016/j.trac.2008.01.009>.
- Williams, H.D., Trevaskis, N.L., Charman, S.A., Shanker, R.M., Charman, W.N., Pouton, C.W., Porter, C.J.H., 2013. Strategies to address low drug solubility in discovery and development. *Pharmacol. Rev.* 65, 315–499. <https://doi.org/10.1124/pr.112.005660>.