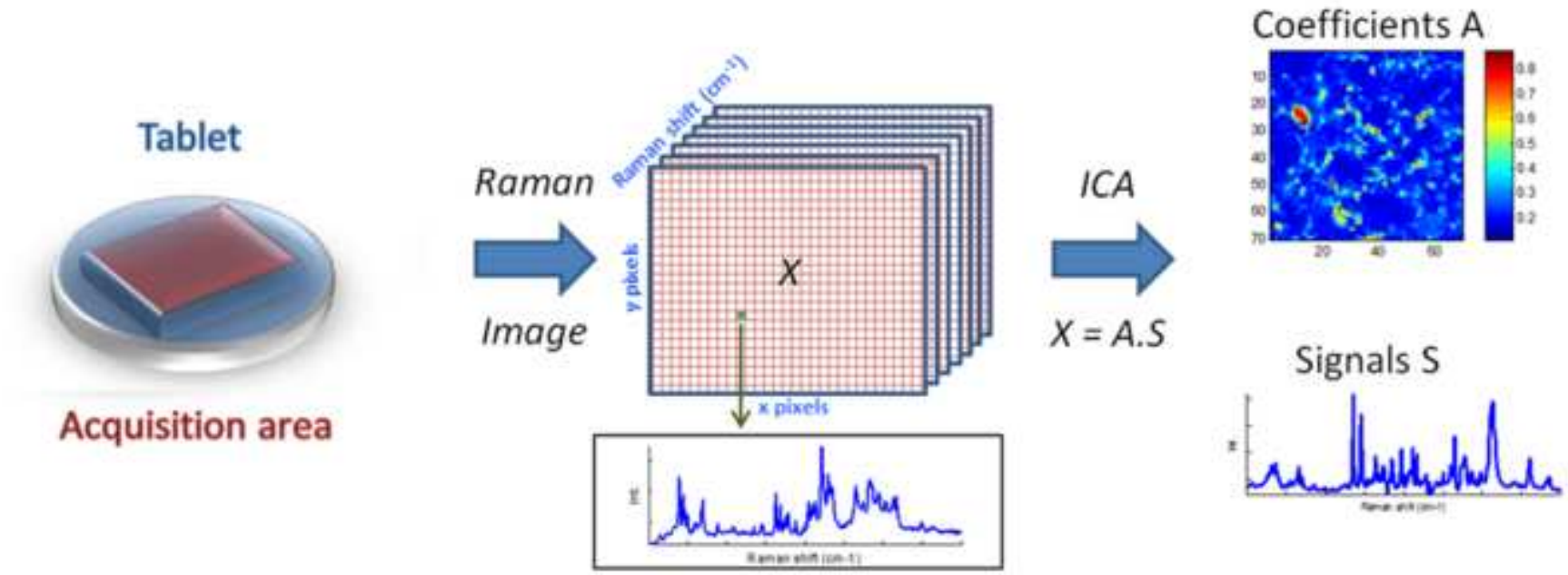
Application of independent component analysis on raman images of a pharmaceutical drug product: pure spectra determination and spatial distribution of constituents

M. Boiret, D.N. Rutledge, N. Gorretta, Y.M. Ginot, J.M. Roger

***Graphical Abstract**



Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Highlights:

- Raman imaging is an innovative analytical tool for pharmaceutical development.
- Independent component analysis was used to detect pure spectra of a formulation.
- The distribution of active and excipients was examined without prior knowledge.
- Innovative tools to select the number of components were explored.
- Limitations of the technique and future improvements were discussed.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Application of Independent Component Analysis on Raman Images of a pharmaceutical drug product: Pure spectra determination and spatial distribution of constituents

Mathieu Boiret[a*], Douglas N. Rutledge[b], Nathalie Gorretta[c], Yves-Michel Ginot[a], Jean-Michel Roger[c]

[a] Technologie Servier, 27 rue Eugène Vignat, 45000, Orléans, France
[b] AgroParisTech, UMR 1145 Ingénierie Procédés Aliments, rue Claude Bernard, F-75005 Paris, France
[c] Irstea, UMR ITAP 361, Avenue Jean-François Breton, 34033, Montpellier, France

* Corresponding author:
Tel.: +33 238 238 175
E-mail addresses: mathieu.boiret@fr.netgrs.com

1. Introduction

In recent years, chemical imaging has become an emerging technique that integrates conventional imaging and spectroscopy to combine spatial and spectral information from a sample [1]. The use of vibrational spectroscopies such as near infrared or Raman is particularly appreciated within the pharmaceutical research and development environment. Indeed, vibrational spectroscopy technologies on solid pharmaceutical samples have many advantages such as the rapidity of analysis, the non-destruction of the sample and the possibility to perform an analysis without using solvents. The spatial information provides useful information on product processing, for formulation development or to control the quality of an existing drug product. Indeed, the distribution of actives or excipients within a specific formulation becomes an important quality control parameter.

Several applications of Raman spectroscopy have been published and the potential of this technique is widely accepted [2]. The use of Raman spectroscopy for the detection of trace cristallinity [3] and the determination of active content within pharmaceutical capsules [4], are of great interest for the development and the quality control of a formulation. Moreover, hyperspectral imaging shows considerable promise for providing information in diverse fields such as remote sensing [5] for interpretation of experimental spectroscopic images from the geographical region of Cuprite, foods and agriculture [6] for analysis of cucumber leaves and pharmaceuticals [7,8] for analysis of solid dosage forms or the detection of polymorphic forms in tablets.

Coupling spectroscopy and imaging generates a huge amount of data. Most of the time, the image cube is unfolded into a data matrix and to extract the maximum of information, it is necessary to use multivariate data analysis methods and spectral decomposition techniques [9]. Standard chemometric tools such as principal component analysis [10], cluster analysis [11], classical least squares [12] and multivariate curve resolution [13] have previously been described in the literature on Raman datasets.

Independent component analysis (ICA) is a blind source separation algorithm [14] particularly appreciated for the decomposition of spectroscopic data. Its ability for spectral decomposition of UV-VIS spectra has already been evaluated [15]. Wang et al. [16] also highlighted that ICA can be used as a blind source separation technique to extract pure component information from various measured analytical signals such as mass spectra, mid-Infrared spectroscopy spectra or chromatograms. In this article, ICA was applied on a promising technique for pharmaceutical drug product analysis: the Raman spectroscopy. In ICA, each row of the data matrix is considered to be a sum of pure source

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

signals, neither the source signals, nor their proportions being known. ICA aims to extract these pure sources, underlying the observed signals, as well as their concentration in each mixture. Source signals are assumed to have a definite structure, and so their intensity does not have a Gaussian distribution. On the other hand, although the distributions of independent signals are not Gaussian, their sum tends towards a Gaussian distribution. ICA aims to extract the pure source signals by maximization of their non-Gaussianity [17].

In this paper, a commercial pharmaceutical tablet was analysed by Raman chemical imaging. The objective was to extract interpretable pure signals using ICA, in order to examine the distribution of active principal ingredients (API) and major excipients. ICA approach can be used without pure spectra knowledge. The direct data analysis of the image is a huge advantage comparing with the usual Chemometric algorithms. This approach can become a useful tool for quality control of a pharmaceutical drug product or to analyse a product with an unknown composition. As a method based on decomposition of the original data matrix, the number of independent components is a critical step of this algorithm. Usually the number of independent components to extract is determined based on prior knowledge concerning the formulation [18]. In order to select the best number of independent components, innovative tools previously developed and published were used in this study. Each calculated source signal was compared with the pure spectra of the constituents and the distribution of the compound in the tablet determined. Being a critical parameter of the ICA model, the number of ICs was intentionally modified, simulating under-decomposition or over-decomposition, in order to test the effect on results.

2. Materials and methods

2.1. Samples

A commercial coated tablet of Bipreterax® was used for the study. Bipreterax® is used for arterial hypertension treatment and is commercialised by “Les Laboratoires Servier”. It is also known as Perindopril (active principal ingredient 2 or API 2) / Indapamide (active principal ingredient 1 or API 1) association and contains respectively 4mg of API 2 and 1.25mg of API 1 in the commercial drugs. Actives are known to have several solid state forms, but only one of them is present in this formulation. Major core excipients are lactose monohydrate, microcrystalline cellulose (Avicel) and magnesium stearate. In order to analyse the tablet core, the coating was removed by eroding the sample with a Leica EM Rapid system (Leica, Wetzlar, Germany). A visual examination of the tablet did not provide any information concerning the distribution of the different compounds within the tablet.

2.2. Raman imaging system

The image was collected using a RS400 PerkinElmer system (Perkin Elmer, Waltham, MA) and the Spectrum Image version 6.1 software. The microscope was coupled to the spectrometer and spectra were acquired through it with a spatial resolution of 10µm in a Raman diffuse reflection mode. Wavenumber range was 3200–100 cm^{-1} with a resolution of 2 cm^{-1}. Spectra were acquired at a single point on the sample, then the sample was moved and another spectrum was taken. This process was repeated until spectra of points covering the region of interest were obtained.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

A 785nm laser with a power of 400mW was used. Two scans of two seconds were accumulated for each spectrum. An image of 70 pixels per 70 pixels corresponding to 4900 spectra was acquired for a surface of 700µm by 700µm.

2.3. Pre-processing

Data were pre-processed in order to remove non-chemical biases from the spectra (scattering effect due to non-homogeneity of the surface, interference from external light source, spikes due to cosmic rays, random noise). First of all, data were spike-corrected in order to reduce the effect of cosmic rays. Next, the spectral range was reduced in order to focus only on the region of interest, corresponding to a Raman shift from 1800cm^{-1} to 200 cm^{-1}. Reduced spectra were pre-processed by standard normal variates correction (SNV) [19] in order to reduce the effect of baseline variations and uninformative variations in global spectral intensity.

2.4. Independent Component Analysis (ICA)

ICA is one of the most powerful techniques in blind source separation [20]. It has been developed to extract the pure underlying signals from a set of mixed signals in unknown proportions. Considering a noise-free ICA model, a matrix X (n x m) is decomposed as a linear generative model by the following expression:

$$X = A.S \quad (1)$$

Where S is a (k x m) matrix of k independent source signals called the independent components and A is a (n x k) mixing matrix of coefficients or proportions of the pure signals in each mixed signal of X.

The objective of ICA is to estimate a set of vectors that are as independent as possible, and the mixed signals in X can then be expressed as linear combinations of these independent components (ICs). It attempts to recover the original signals by estimating a linear transformation, using a criterion which reflects the statistical independence among the sources.

To solve the previous equation (Eq. (1)), an unmixing matrix W based on the observation of X needs to be calculated. The output U, constituted by the independent component u_1,u_2, … u_n should be as independent as possible. For a noise-free ICA model, W should be the inverse of A, and U should be equal to S, according to the following equation:

$$U = WX = W(AS) = S \quad (2)$$

The mixing matrix A can then be calculated as:

$$A = XS^T (SS^T)^{-1} \quad (3)$$

Lots of algorithms are available to perform ICA calculations such as FastICA [21] or Radical [22]. In this paper, the JADE (Joint Approximate Diagonalization of Eigenmatrices) algorithm was used [23]. Compared with other methods based on parameter optimization, the Jade algorithm performs matrix diagonalisations, and therefore does not involve an optimization procedure [24].

The ICA_by_blocks algorithm [25] was used to determine the optimal number of signals to extract. This method starts by splitting the initial data matrix X into B blocks of samples (with approximately equal numbers of rows). Note that the samples in each block have to be representative of the whole dataset. ICA models are then computed with an increasing number of ICs for each block. To ensure

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

the same signs of the ICs of the different models, the signs of the vector A (and therefore the corresponding S) are adjusted so that the most intense value in each vector of A is positive. ICs corresponding to true source signals should be found in all representative subsets of samples, or row blocks, of the full data matrix. These ICs should be strongly correlated.

2.5. Data analysis

Data analysis was performed by using Matlab R2012a software. The Matlab code of the JADE algorithm was downloaded from the web site in ref. [26].

3. Results & discussion

3.1. Selection of number of independent components

Determination of the number of ICs for ICA decomposition is a critical step of the data analysis. Indeed, calculating too few ICs results in non-pure signals, whereas calculating too many ICs can decompose pure signals into several contributions. The ICA_by_blocks method was applied by splitting the dataset row-wise into two blocks and by performing ICAs on each block. Sample selection to create the two subsets was done by using a "venetian blind" procedure. ICA models were calculated for both blocks with from 1 IC to 20 ICs. ICs were compared in each block by calculating the correlation coefficients between all pairs of signals from both blocks for a given model. The highest-dimensional model for which ICs obtained in a block were similar to ICs obtained in another block indicates the optimal number of ICs to extract from the data under study. Figure 1 shows that the lowest correlation between signals significantly decreases after 9 ICs, which was therefore considered as the optimal number of component for the decomposition of the dataset. The initial drop after 4 ICs and then after 7 ICs is assumed to be due to the fact that the ICs are not extracted from the two data blocks in exactly the same order.

Since the sample contains five compounds and supposing that the five spectra are independent and that the acquired mixture spectra are linear combinations of the pure spectra, five ICs should have been sufficient. In this example, in contrast with the theoretical decomposition, four more components were used to build ICA models. Physical effects such as particle size variation or fluorescence of a compound could explain this “over-decomposition” of the dataset.

3.2. Distribution of API

An ICA model based on the JADE decomposition with 9 ICs was calculated on the unfolded, SNV pre-processed data cube. The matrices of the proportions, **A**, for each signal, **S**, were then folded back in order to obtain a representation of the spatial distribution of each independent component. In figure 2, different textures of images can be observed. Indeed, IC1, IC6 and IC9 show very specific inhomogeneous distributions with agglomerates. Considering the different scales of score images, IC2, IC3, IC4, and IC5 have similar textures (or distributions) such as IC7 and IC8 which are the same as that in IC1. It can also be seen that the distributions observed in these two sets of images are complementary, indicating that these two sets of Independent components occupy complementary regions in the tablet. In order to associate an independent component with a chemical compound, the calculated signals were examined.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Figure 3 shows the 9 signals calculated by ICA. Signals from IC1, IC2, IC3, IC4, IC5, IC6 and IC9 look like well-defined Raman spectra with no baseline shift due to fluorescence effects whereas the signals in IC7 or IC8 contain noise and baseline variations which could be explained by a fluorescence effect. In theory, and supposing the independence of each spectrum within the formulation, 5 ICs should have been sufficient for the matrix decomposition. However, 9 ICs were determined to be present, possibly due to physical effects, or interactions between constituents. Considering the simplicity of the preprocessing method applied on the Raman spectra (spike correction, selection of a specific range and SNV), the quality of the calculated signals was sufficient and perfectly suitable for analytical interpretation.

The spectra for the known constituents in the tablets are plotted in figure 4. Even though the spectra of all compounds are very different, lots of Raman bands are overlapped. A mixture spectrum is a combination of these spectra, given the presence of each constituent in any specific pixel of the image. In order to interpret the ICA results, the correlation coefficients between the ICA signals and the pre-processed spectra of the compounds were calculated. Results can be found in table 1.

The comparison between the calculated signals and the true spectrum of each compound shows that only two ICs are directly linked to the drug product constituents. For each component, the highest correlation was highlighted with bold characters in table 1.

No high correlations were found for Magnesium stearate. Two very high correlations were highlighted between the pure spectra and the calculated signals (respectively 0.92 between IC9 and the active principal ingredient 1 and 0.96 between IC6 and the active principal ingredient 2). As is shown in figure 5 and figure 6, the calculated signals (IC9 and IC6) are in effect very similar to the pure spectra of API 1 and API 2. The refolded images of the corresponding proportions, **A**, therefore reflect the distribution of these two compounds. As can be seen in figure 2, the distribution of active principal ingredients is not perfectly homogeneous and agglomerates are observed.

IC1 is mainly correlated with the spectrum of Avicel. Specific bands due to the chemical bond vibrations are observed in this component (especially between 1250cm^{-1} and 1000cm^{-1}, spectral range linked to CC ring bond stretches and CO stretches). IC7 and IC8 are mainly correlated with the spectrum of avicel (0.38 for IC7 and 0.61 for IC8) but the correlation with lactose (0.36 for IC7 and 0.45 for IC8), magnesium stearate (0.32 for IC7 and 0.23 for IC8) and API1 (0.21 for IC7 and 0.18 for IC8) cannot be considered as non-significant. IC7 and IC8 signals are not well defined Raman spectra and contain principally noise or baseline variations which can explain these high correlations with several different products. As can be seen in figure 2, IC7 and IC8 have similar spatial distributions which are the same as that in IC1. Avicel is a microcrystalline cellulose powder which is known as a product providing a fluorescence effect with Raman, which could explain the contribution of IC7 and IC8.

As is shown in figure 7, IC2, IC3, IC4 and IC5 are linked to the lactose spectrum. Lots of lactose Raman bands are identified in these IC signals (for example band at 460cm^{-1} in signals 2, 3 and 5 due to various CCO and OCO bending modes, or band at 1088 cm^{-1} linked to the stretching vibration of the COC bridge). These 4 components gave their highest correlations with the lactose spectrum. However, these correlations were low (from 0.23 to 0.47) reflecting the decomposition of the pure spectrum into 4 components. The signal decomposition was particularly significant in the low Raman shift spectral range. In this spectral region, coupled CC and CO vibrations rather than single functional

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

group are mainly observed. By observing the refolded image of coefficients, note that the distribution of this product was very similar (considering the different image scales).

The observed decomposition of the lactose information into separate Independent Components could be due to two phenomena. The first one is the physical effect. Indeed, lactose is known to have important particle size variations which can modify the light scattering and as a consequence the Raman spectra. These slight modifications could behave as independent phenomena and thus result in separate ICs. Moreover, the different combinations of vibrations could be interpreted by ICA decomposition as an independent variation. The second hypothesis is that it is an artefact due to the ICA decomposition itself. Indeed, as the formulation contains 5 compounds, the model may have mathematically over-decomposed the dataset by using 9 ICs.

In order to explore the ability of ICA to extract a pure signal from lactose, an ICA model was calculated with 5 ICs, which was the known number of constituents used to manufacture the tablet. By comparing the 5 ICs with the pure spectra, a high correlation was found with lactose (R = 0.90), one with API1 (R = 0.95), one with API2 (R = 0.94), while a weak correlation was found with avicel (R = 0.39) and one signal contained noise and mixed pure contributions. The lactose contribution was therefore not divided among several components, as was observed when using 9 ICs. As is shown in figure 8, the calculated signal IC3 was very similar to the pure lactose spectrum. With 5 ICs, the decomposition of the original matrix was mainly due to chemical variations whereas the decomposition using 9 ICs included physical effects.

By observing ICA coefficients and signals, it can be seen that no information from the magnesium stearate was observed. The non-detection of this compound, frequently used as a lubricant in a pharmaceutical formulation, could be mainly due to its low concentration in the tablet (0.5 w/w%). Indeed, several hypotheses can be advanced to explain this lack of detection: the physical formulation of the product, the sensitivity of the spectroscopy or the failure of the ICA algorithm. As the analysed area does not represent the whole surface of the tablet and because of its low content, it is possible that the acquired spectra did not contain any magnesium stearate information. Moreover, the Raman contribution of the magnesium stearate could be hidden by the contribution of the other constituents. In order to test the ability of ICA to detect and extract the information related to magnesium stearate, new models with more components and other preprocessing methods were tested (details not shown). By using a Savitzky-Golay preprocessing [27] and a model with 15 ICs, one signal (figure 9) was highly correlated (r = 0.87) with the pure spectrum of magnesium stearate and the distribution of the product can then be studied (figure 10). However, the quality of other signals significantly decreased. Pure spectra were divided among several components and the analytical meaning of each signal was not intuitive.

4. Conclusions

ICA was successfully applied on a Raman image of a commercial tablet. A representative image of the tablet was acquired and the spectrum of each pixel, which can be associated to a mixture of the different pure compounds, was pre-processed and analysed using the JADE algorithm to calculate signals and proportions with a specified number of components. This parameter was estimated by using the ICA_by_blocks method. This technique shows very good results to choose the most appropriate number of ICs on a real Raman dataset. It avoids arbitrary selection of this critical criterion.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

This method gave good results to provide pure spectra of the active substances. Contribution of avicel was spread among 3 ICs. The first one was very similar to the pure spectrum of avicel whereas the two others were mainly fluorescence signals. Being a microcrystalline cellulose, avicel is known to be prone to fluorescence effects. The contribution of lactose was shared over 4 ICs which may be due to an over-decomposition of the original dataset or to physical contributions. In order to improve the pure lactose signal quality and based on knowledge of the product formulation, an ICA model was calculated using fewer ICs. The lactose contribution was then no longer divided among several signals but, the physical effects were no longer observed.

This should be contrasted with the fact that using an insufficient number of ICs leads to the non-detection of a low content compound, magnesium stearate. It has been shown here that using a very large number of components and another preprocessing method resulted in a well-defined ICA signal linked to magnesium stearate. It was then possible to examine the distribution of this low content product within the tablet. However, due to the over-decomposition of the dataset, other pure signals were divided among several components, which made the identification of each contribution within the tablet more difficult.

The ICA_by_blocks method was therefore a compromise between under- and over- decomposition. Even if the contribution of lactose or avicel were divided among several components, the spatial information obtained could be very useful for formulation development or to improve the quality control of pharmaceutical samples. New approaches, based on data fusion from ICA calculations to gather information from the same constituent, are under development and will be detailed in a future work.

References

[1] C. Gendrin, Y. Roggo, C. Collet, Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review, J. Pharm. Biomed. Anal. 48 (2008) 533-553.

[2] J. Rantanen, Process analytical applications of Raman spectroscopy, J. Pharm. Pharmacol. 59 (2007) 171-177.

[3] E. Widjaja, P. Kanaujia, G. Lau, W. K. Ng, M. Garland, C. Saal, A. Hanefeld, M. Fischbach, M. Maio, R.B.H. Tan, Detection of trace cristallinity in an amorphous system using Raman microscopy and chemometrics analysis, Eur. J. Pharm. Sci. 42 (2011) 45-54.

[4] P. Matousek, A.W. Parker, Non-invasive probing of pharmaceutical capsules using transmission Raman spectroscopy, J. Raman Spectrosc. 38 (2007) 563–567.

[5] X. Zhang, R. Tauler, Application of Multivariate Curve Resolution Alternative Least Squares (MCR-ALS) to remote sensing hyperspectral imaging, Anal. Chim. Acta. 762 (2013) 25-38.

[6] Z. Xiabo, Z. Jiewen, M. Holmes, M. Hanpin, S. Jiyong, Y. Xiaopin, L. Yanxiao, Independent component analysis in information extraction from visible/near-infrared hyperspectral imaging data of cucumber leaves, Chemometr. Intell. Lab. Syst. 104 (2010) 265-270.

[7] Y. Roggo, A. Edmond, P. Chalus, M. Ulmschneider, Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms, Anal. Chim. Acta. 535 (2005) 79-87.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

[8] S. Šašić, S. Mehrens, Raman chemical mapping of low-content active pharmaceutical ingredient formulations. III. Statistically optimized sampling and detection of polymorphic forms in tablets on stability, Anal. Chem. 84 (2012) 1019-1025.

[9] M.J. Goetz , G.L. Coté, R. Erckens, W. March, M. Motamedi, Application of a multivariate technique to Raman spectra for quantification of body chemicals, IEEE Trans Biomed Eng. 42 (1995) 728-731.

[10] H. Grahn, P. Geladi, Techniques and Applications of Hyperspectral Image Analysis, John Wiley & son Ltd., New York, 2007.

[11] M.B. Lopes, J.-C. Wolff, Investigation into classification/sourcing of suspect counterfeit Heptodin™ tablets by near infrared chemical imaging, Anal. Chim. Acta. 633 (2009) 149–155.

[12] L. Zhang, Multivariate data analysis for Raman imaging of a model pharmaceutical tablet, Anal. Chim. Acta. 545 (2005) 262-278.

[13] B. Vajna, A. Farkas, H. Pataki, Z. Zsigmond, T. Igricz, G. Marosi, Testing the performance of pure spectrum resolution from Raman hyperspectral images of differently manufactured pharmaceutical tablets, Anal. Chim. Acta. 712 (2012) 45-55.

[14] J.V. Stone, independent component analysis – A tutorial introduction, A Bradford Book, London, England, 2004.

[15] Y.B. Monakhova, S.A. Astakhov, A. Kraskov, S.P. Mushtakova, Independent components in spectroscopic analysis of complex mixtures, Chemometr. Intell. Lab. Syst. 103 (2010) 108-115.

[16] G. Wang, Q. Ding, Z. Hou, Independent component analysis and its applications in signal processing for analytical chemistry, TrAC - Trends Analyt. Chem. 27 (2008) 368-376.

[17] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (2000) 411-430.

[18] H. Lin, O. Marjanović, B. Lennox, S. Šašić, I.M. Clegg, Multivariate statistical analysis of Raman images of a pharmaceutical tablet, Appl. Spectrosc. 66 (2012) 272-281.

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

[20] L. De Lathauwer, B. De Moor and J. Vandewalle, An introduction to independent component analysis, J. Chemom. 14 (2000) 123-149.

[21] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Comput., 9 (1997) 1483-1492.

[22] E.G. Learned-Miller, J.W. Fisher III, ICA using spacings estimates of entropy, J. Mach. Learn. Res. 4 (2003) 1271-1295.

[23] J.F. Cardoso, High-order contrasts for independent component analysis, Neural Comput. 11 (1999) 157-192.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

[24] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, Independent Component Analysis with the JADE algorithm, TrAC - Trends Analyt. Chem. 50 (2013) 22-32.

[25] D. Jouan-Rimbaud Bouveresse, A. Moya-González, F. Ammari, D.N. Rutledge, Two novel methods for the determination of the number of components in independent component analysis models, Chemometr. Intell. Lab. Syst. 112 (2012) 24-32.

[26] http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m

[27] A. Savitzky et M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Chemometr. Anal. Chem. 36 (1964) 1627-1639.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Figure captions

Figure 1 - Lowest correlation between signals obtained using ICA_by_blocks

The lowest correlation obtained using the ICA_by_blocks approach significantly decreases after 9 ICs, which was considered as the optimal number of component for the decomposition of the dataset

Figure 2 - Proportions coefficients (A) of each IC

Images correspond to the proportions coefficients (A) of a 9 ICs model. A red color corresponds to a high value whereas a blue color corresponds to a low value.

Figure 3 - Signals, S, of the ICA model

These signals correspond to the calculated signals (S) of a 9 ICs model.

Figure 4 - Pure spectra of the drug product constituents

Pure spectra of the five tablet constituents. In blue API 1, in green API 2, in black lactose, in red avicel and in magenta the magnesium stearate. Relative intensities were used as the spectra were split for a better observation.

Figure 5 - Calculated signal of independent component 9 superposed on the spectrum of API 1

Comparison between API 1 spectrum and IC9 signal. The correlation between the two signals is equal to 0.92.

Figure 6 - Calculated signal of independent component 6 superposed on the spectrum of API 2

Comparison between API 2 spectrum and IC6 signal. The correlation between the two signals is equal to 0.96.

Figure 7 - Calculated signal of independent component 2, 3, 4, 5 plotted with the spectrum of Lactose

Comparison between lactose spectrum and IC2, IC3, IC4 and IC5 signals. The correlations between the signals are respectively equal to 0.44, 0.23, 0.25 and 0.47. The pure spectrum of lactose and the four calculated independent components are displayed. The pure spectrum was decomposed into four components.

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

365 Figure 8 - IC3 signal from a 5 components ICA model superposed on the spectrum of lactose.

366 Comparison between lactose spectrum and IC3 signal from a 5 components ICA model. The
367 correlation between the two signals is equal to 0.90.

368

369 Figure 9 - IC12 superposed on the magnesium stearate spectrum from a 15 component ICA model

370 Comparison between magnesium stearate spectrum and IC12 signal from a 15 components ICA
371 model. The correlation between the two signals is equal to 0.87.

372

373 Figure 10 - Distribution of IC12 (magnesium stearate) from a 15 component ICA model

374 Distribution of IC12 from a 15 component ICA model. This component is highly correlated to
375 magnesium stearate.

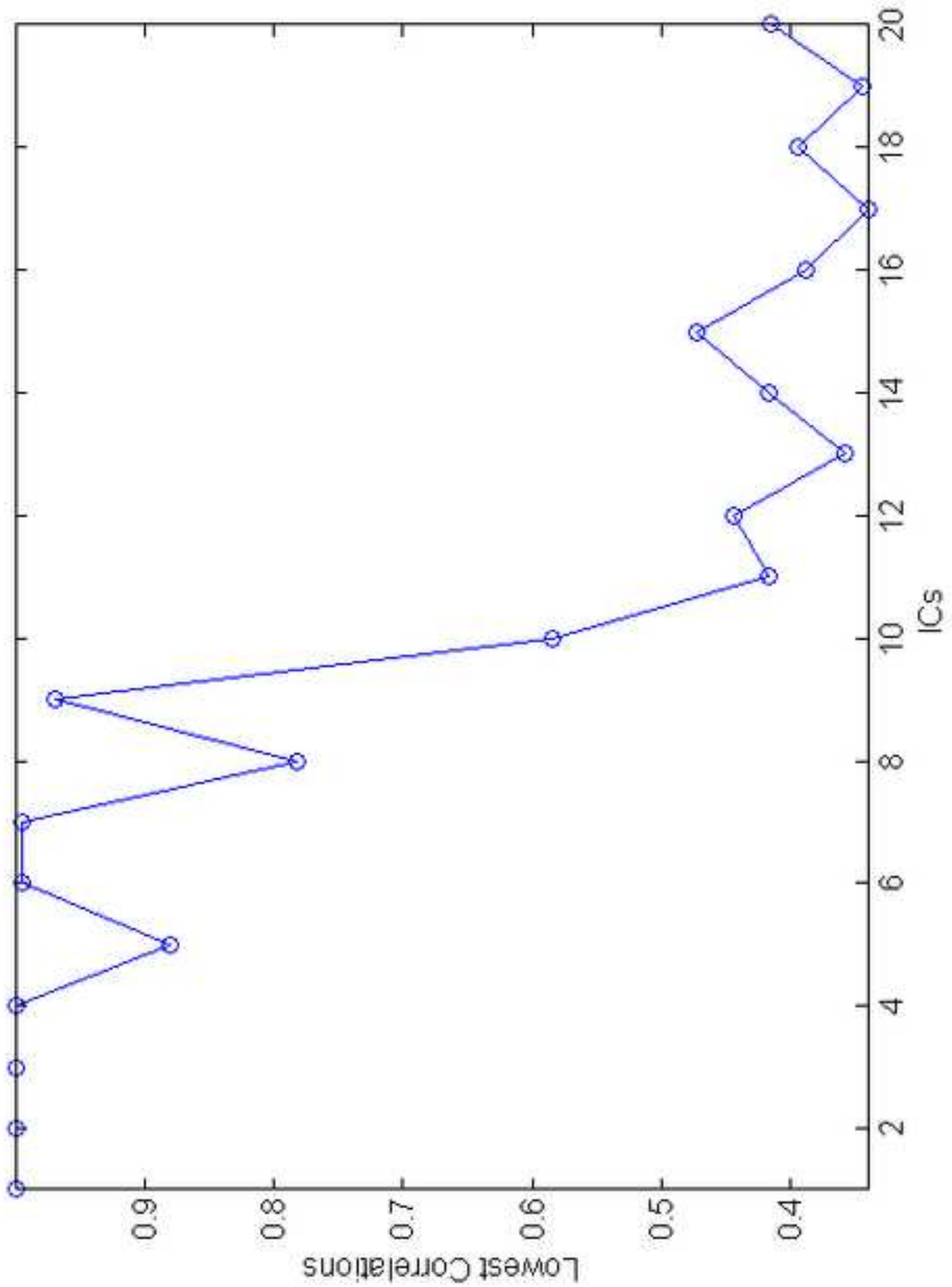
Table(s)

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

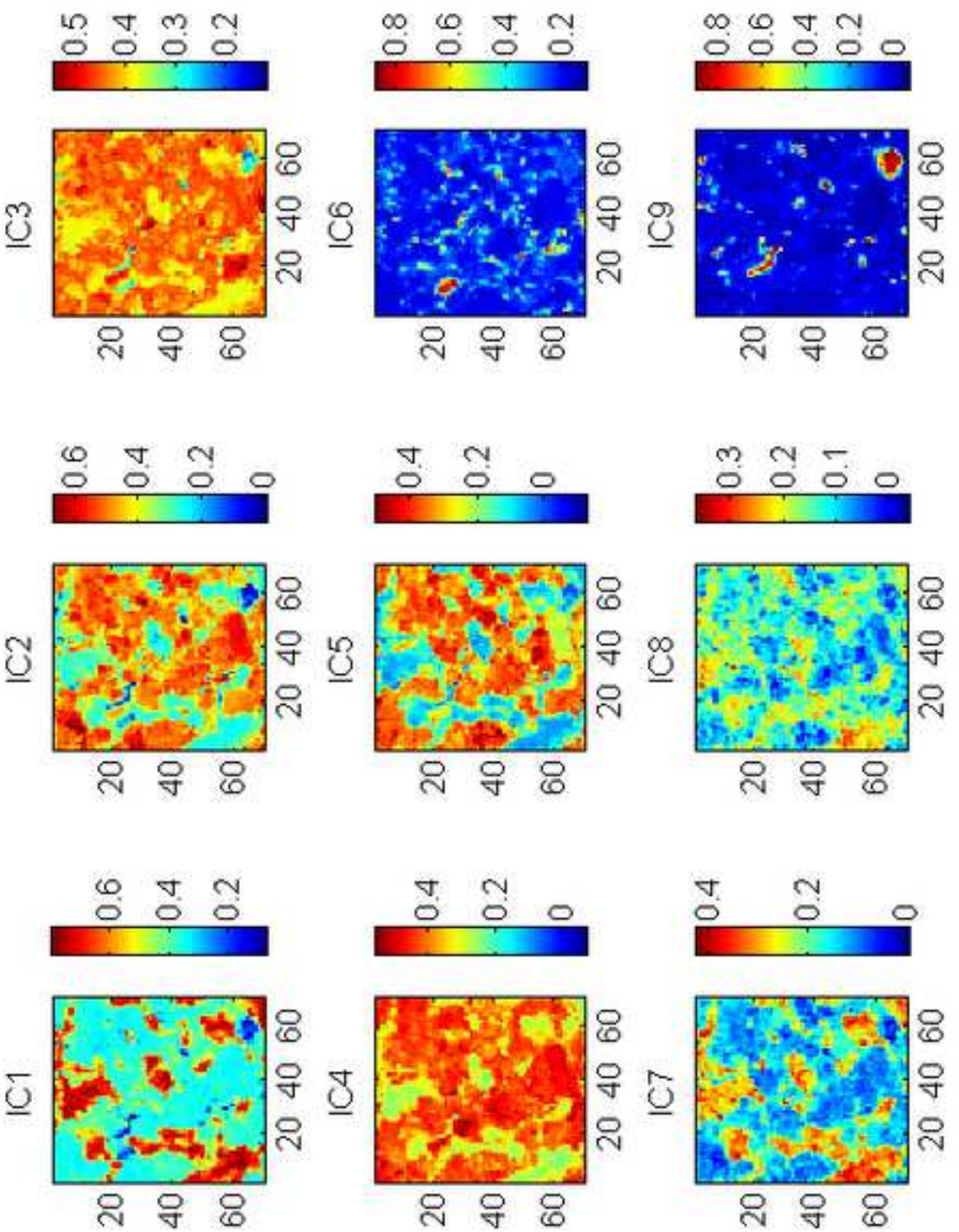
Pure spectrum	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	IC9
API1	0.01	-0.09	0.07	0.14	-0.04	0.13	0.21	0.18	**0.92**
API2	0.06	-0.01	0.11	0.08	0.03	**0.96**	0.08	0.10	-0.06
Lactose	0.25	**0.44**	**0.23**	**0.25**	**0.47**	0.00	0.36	0.45	-0.17
Avicel	**0.49**	0.15	0.06	0.02	0.20	-0.07	**0.38**	**0.61**	-0.20
Magnesium Stearate	0.20	0.00	0.01	0.04	0.04	0.41	0.32	0.23	-0.12

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Figure(s)
Click here to download high resolution image

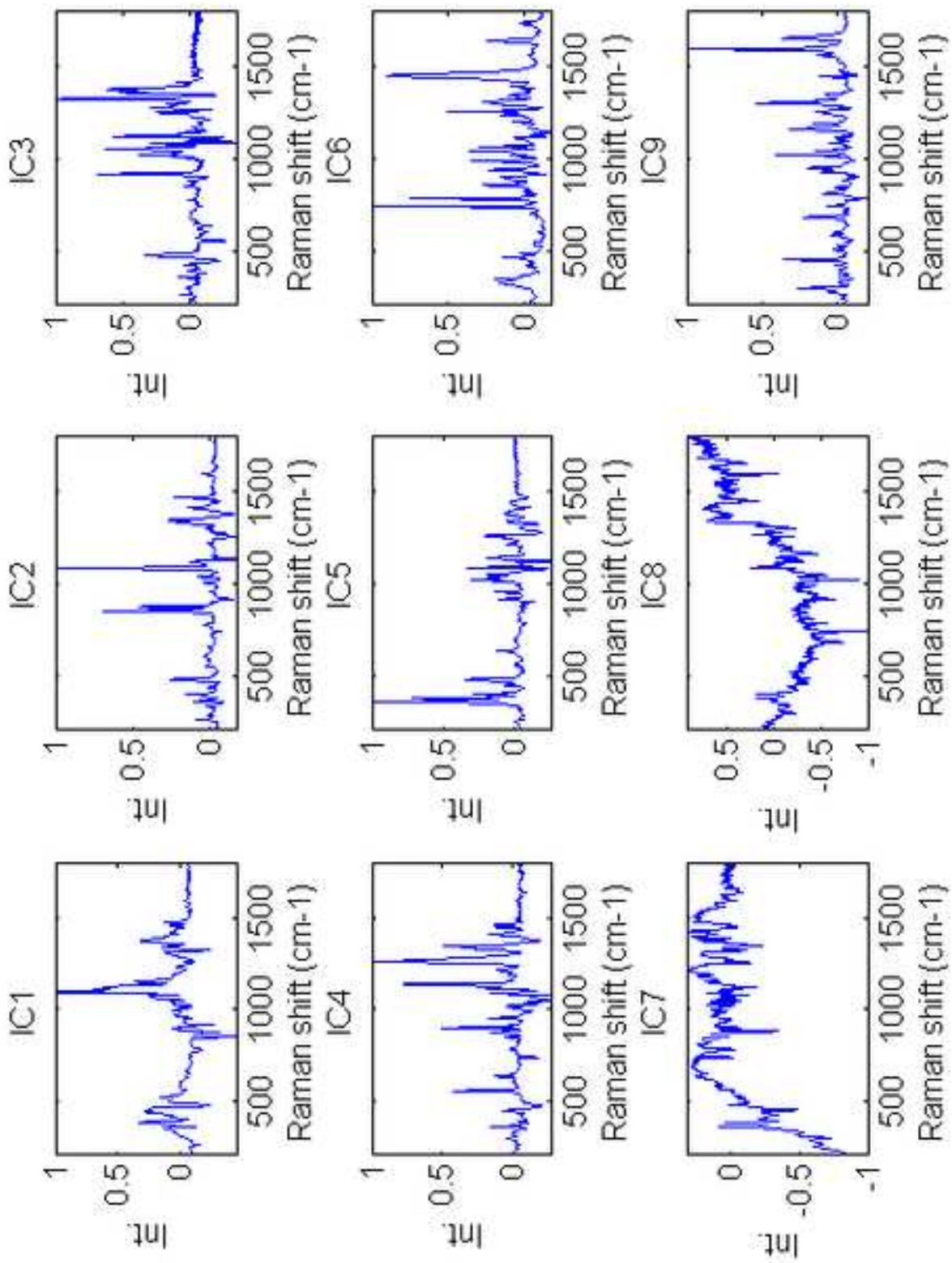


Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

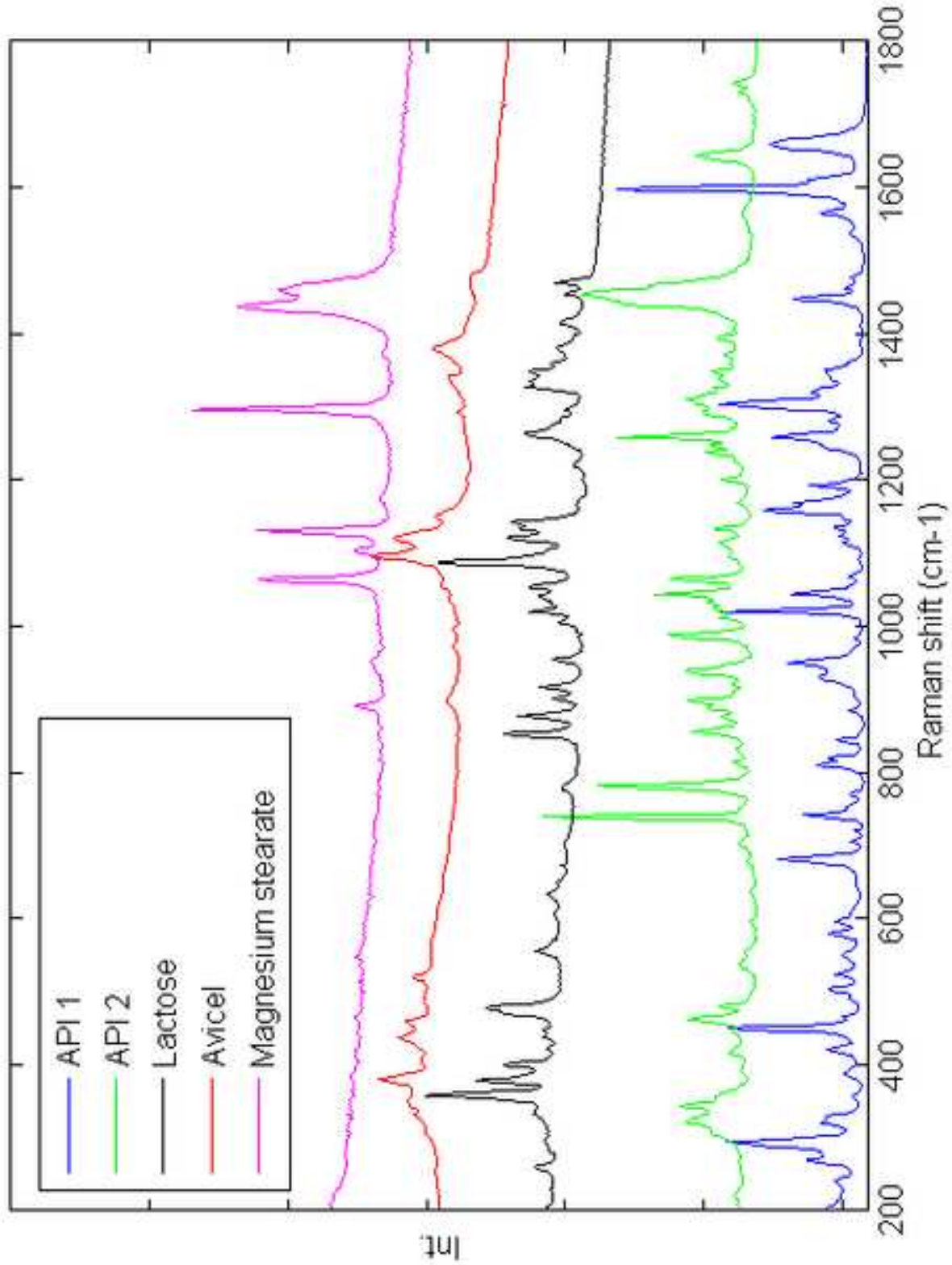


Figure(s)
Click here to download high resolution image

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

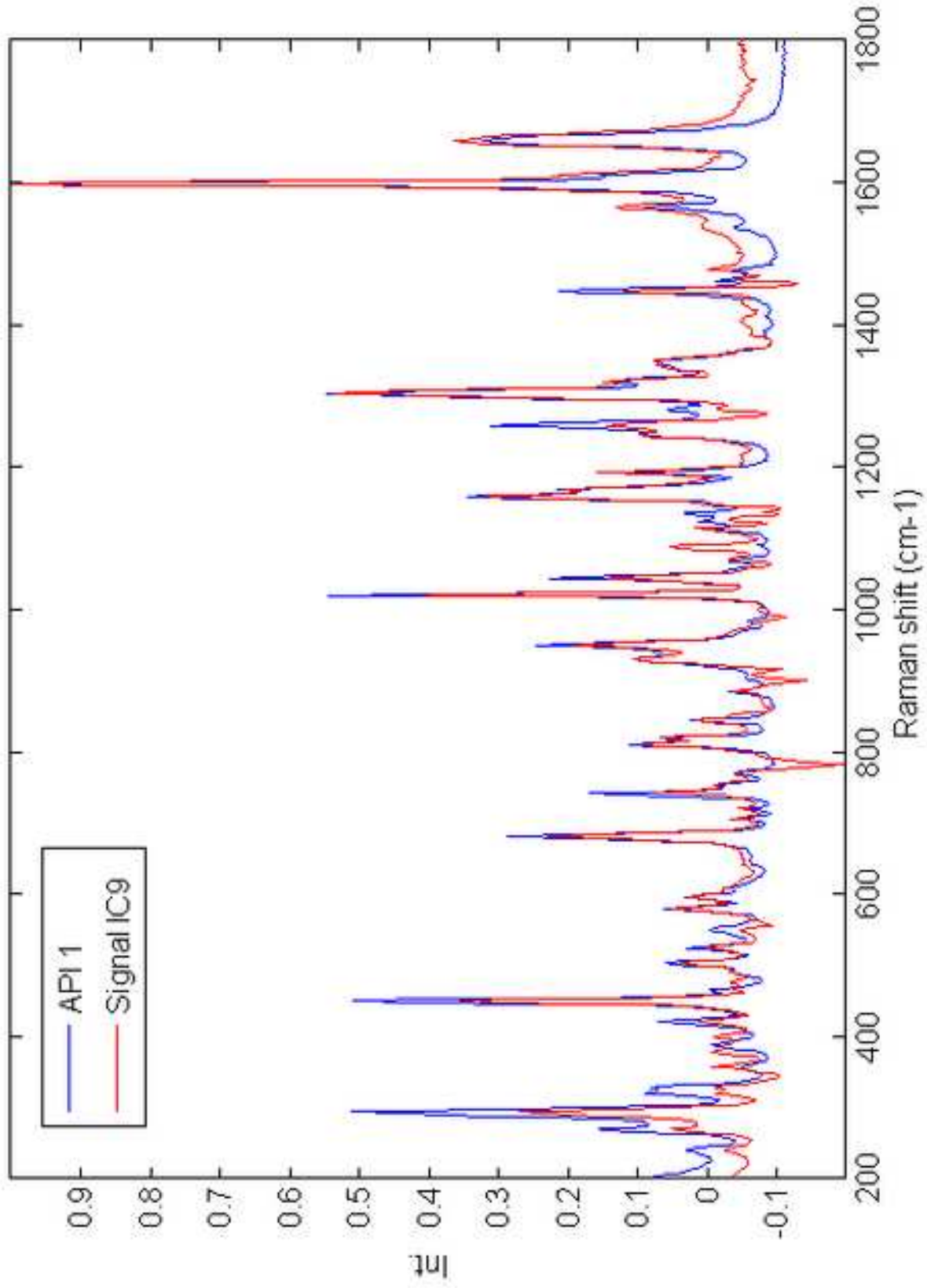


Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025



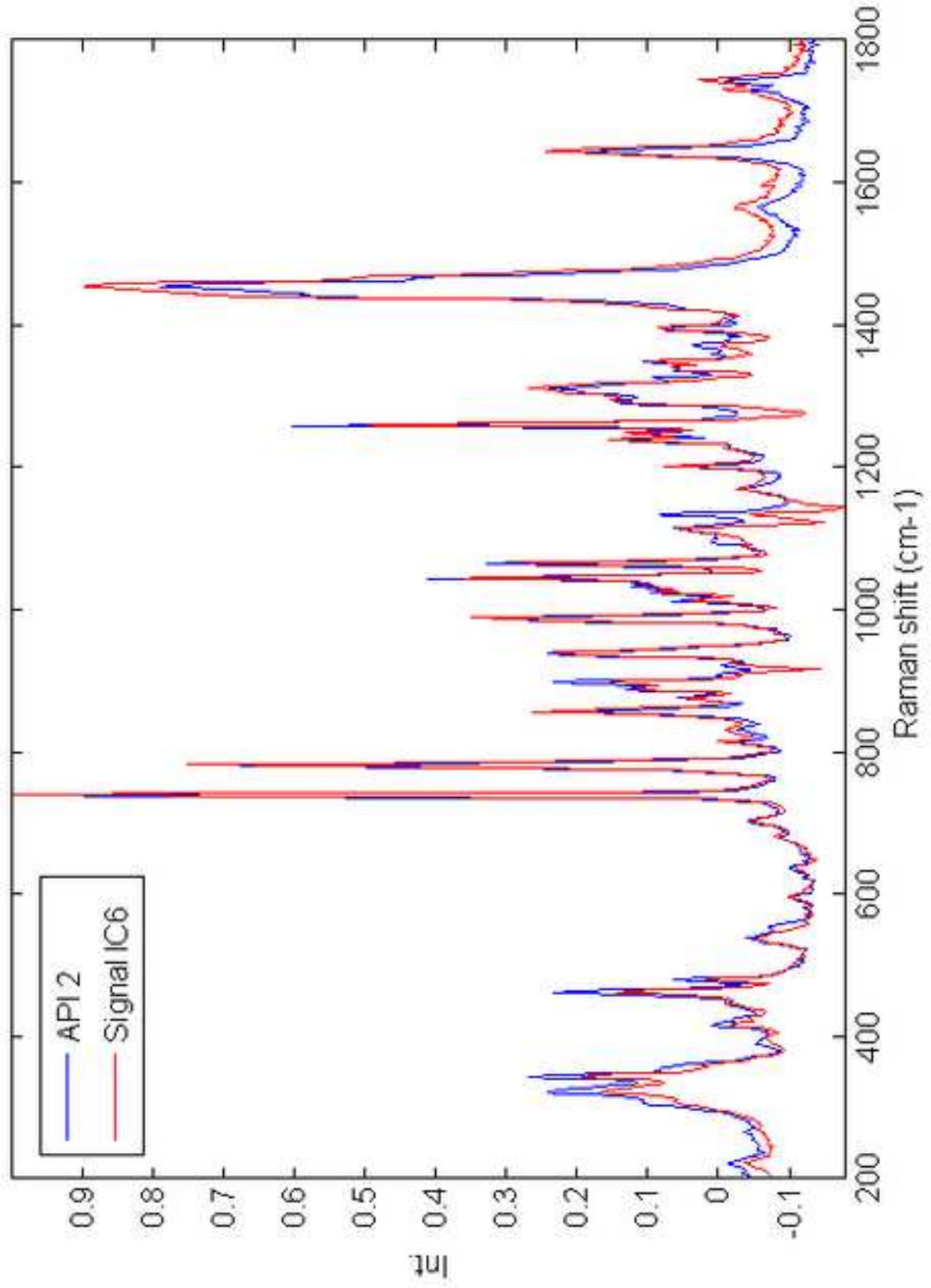
Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Figure(s)
Click here to download high resolution image

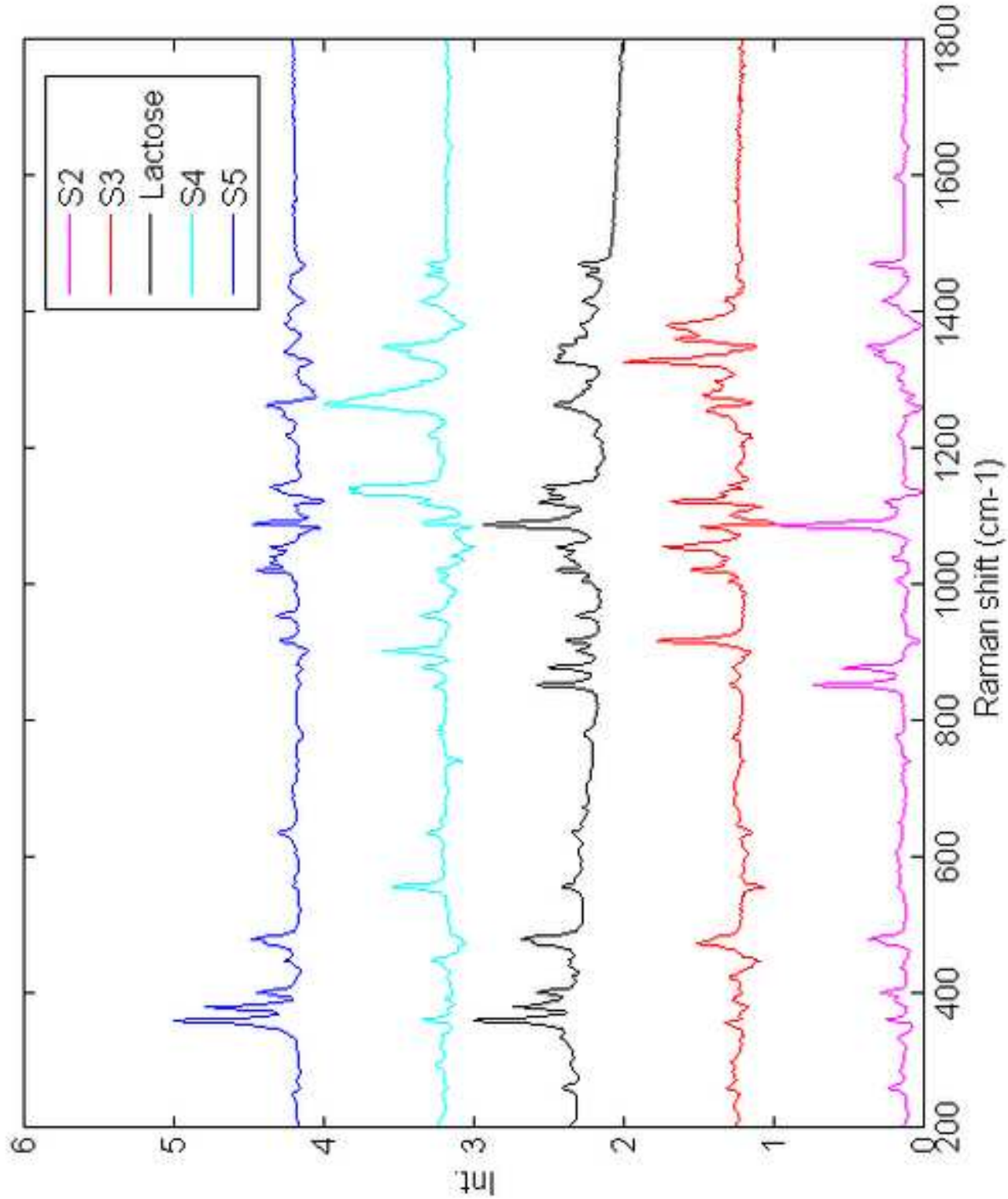


Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

Figure(s)
Click here to download high resolution image

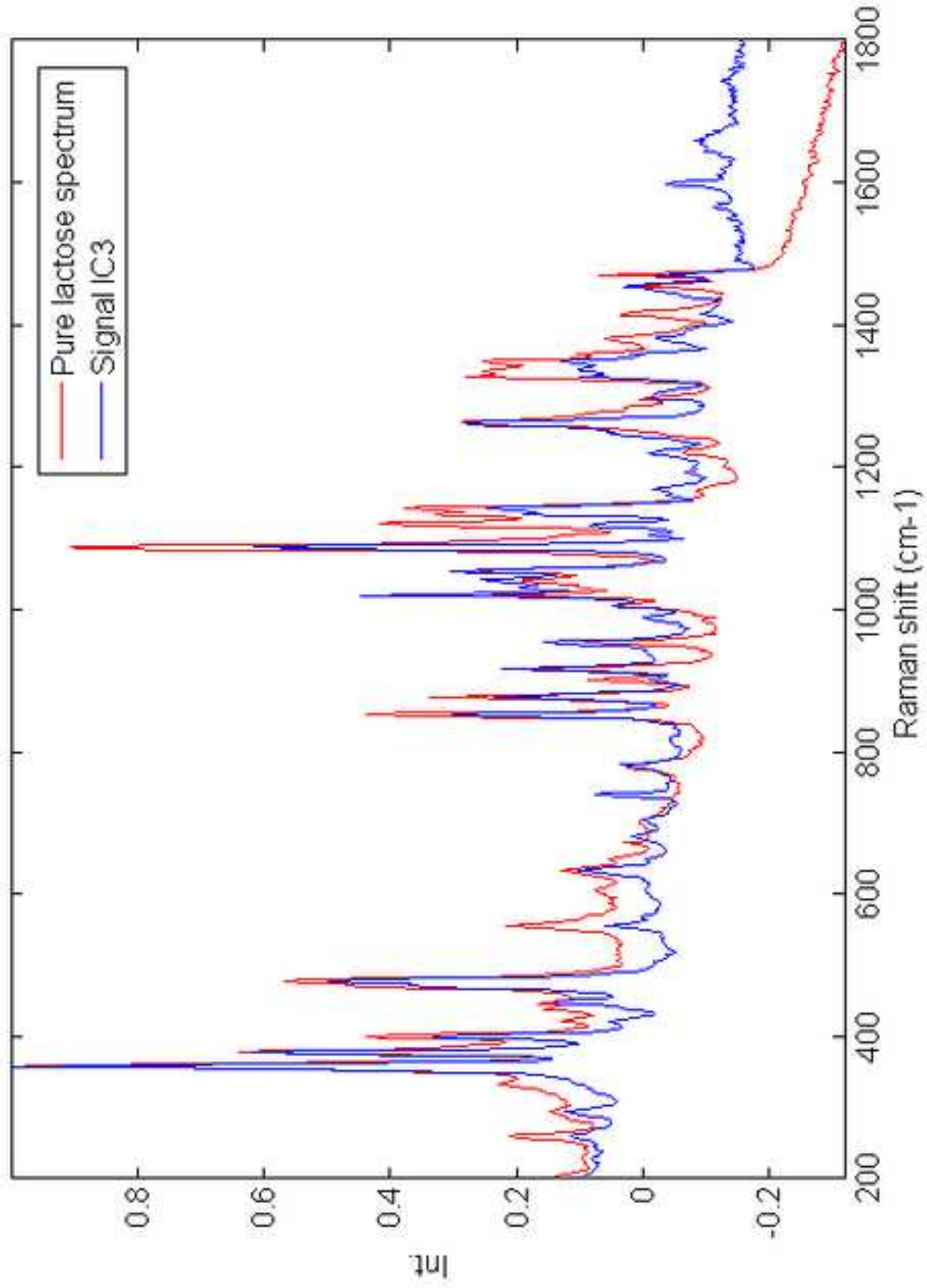


Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025



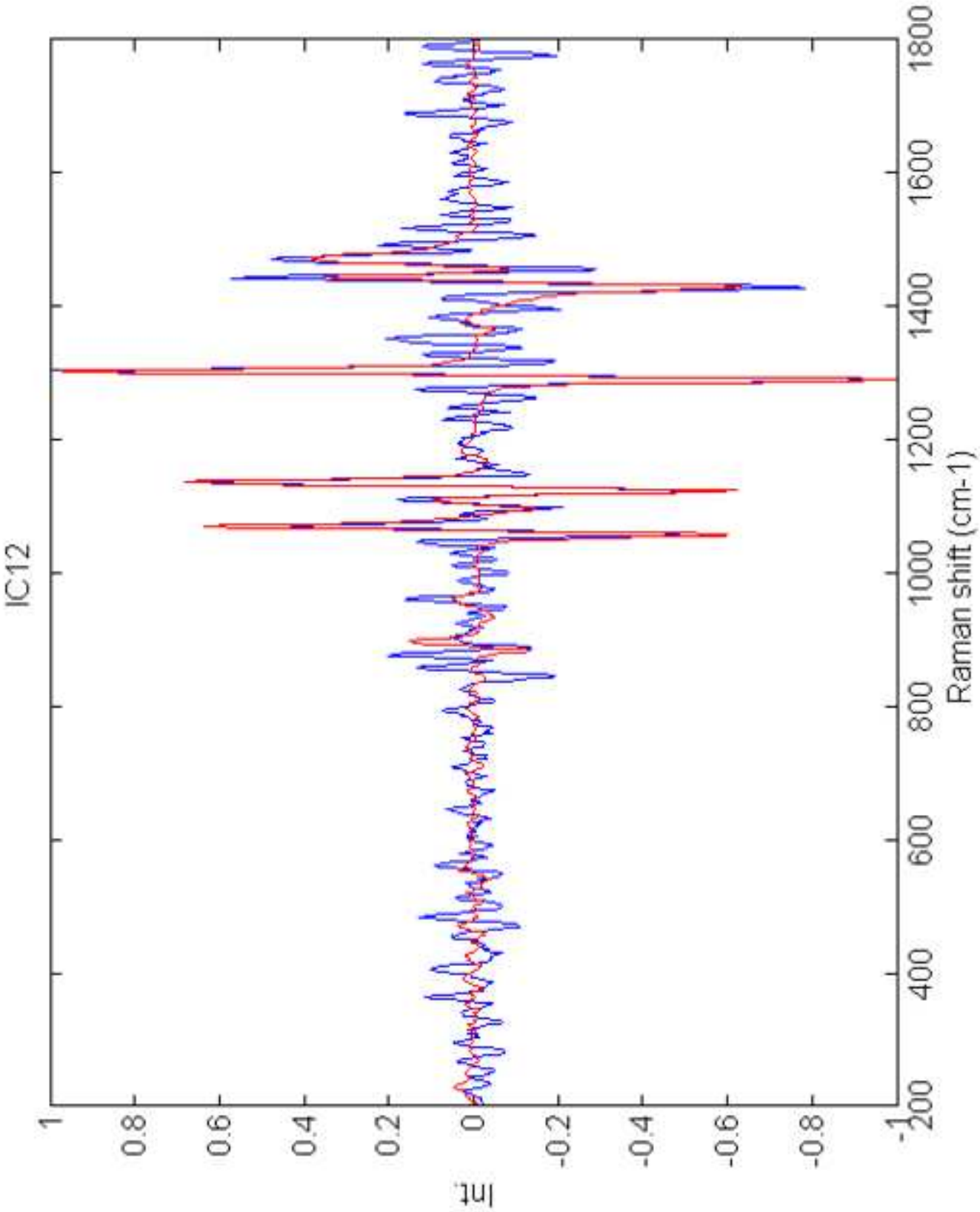
Figure(s)
Click here to download high resolution image

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025

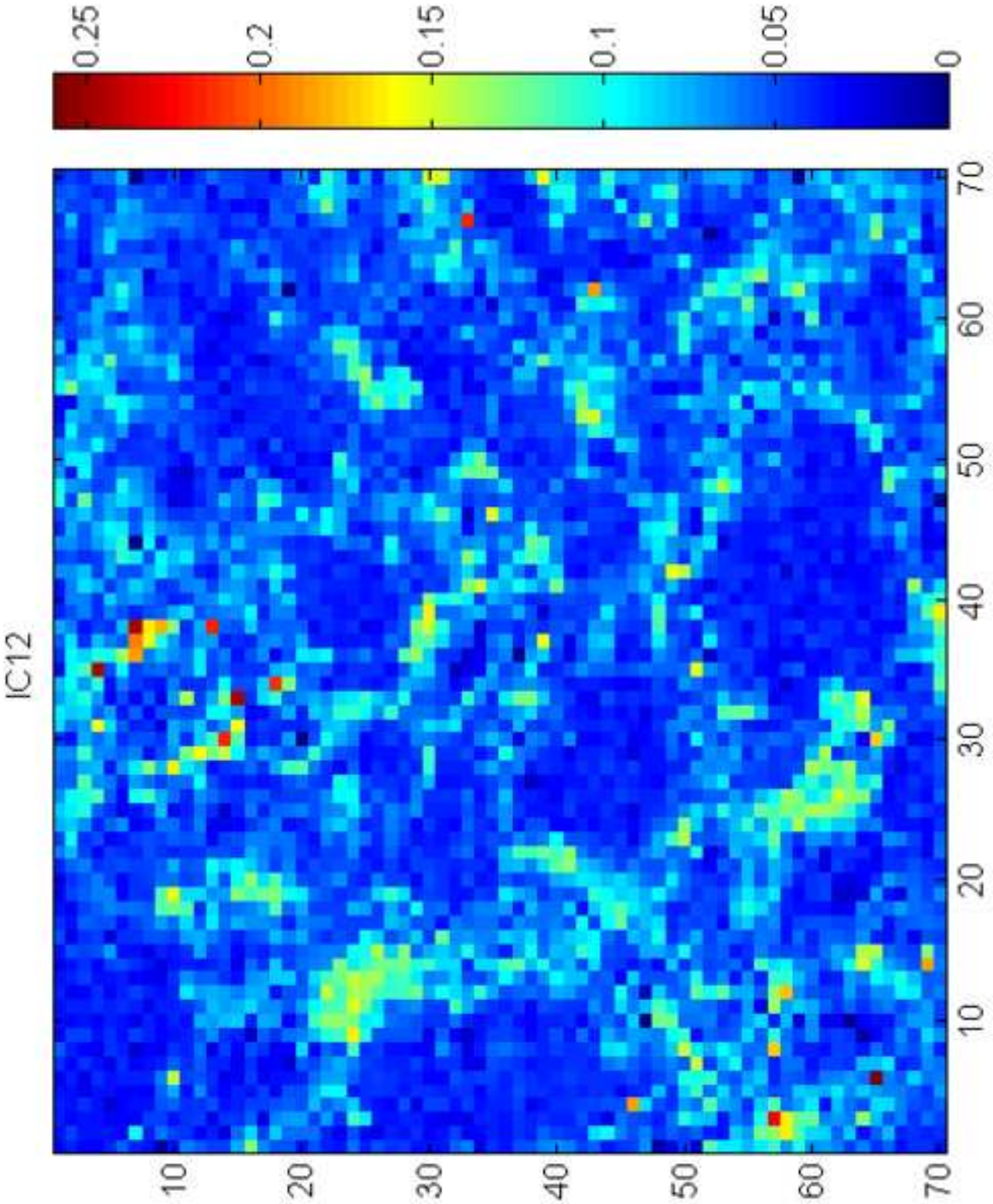


Figure(s)
Click here to download high resolution image

Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025



Author-produced version of the article published in Journal of pharmaceutical and biomedical analysis, 2014, 90, 78-84.
The original publication is available at http://www.sciencedirect.com/science/journal/07317085
DOI : 10.1016/j.jpba.2013.11.025



Figure(s)
Click here to download high resolution image