



HAL
open science

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

► To cite this version:

M. Boiret, D.N. Rutledge, N. Gorretta, Y.M. Ginot, J.M. Roger. Application of independent component analysis on raman images of a pharmaceutical drug product: pure spectra determination and spatial distribution of constituents. *Journal of Pharmaceutical and Biomedical Analysis*, 2014, 90, p. 78 - p. 84. 10.1016/j.jpba.2013.11.025 . hal-00948530

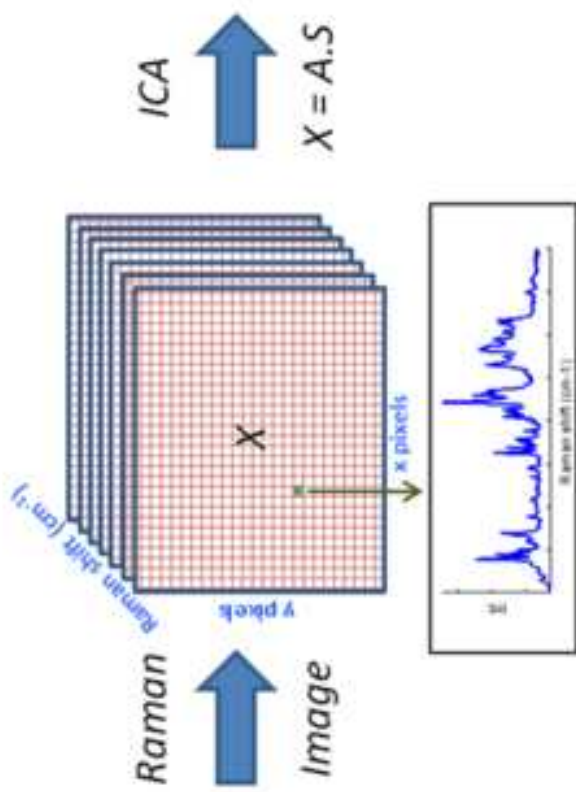
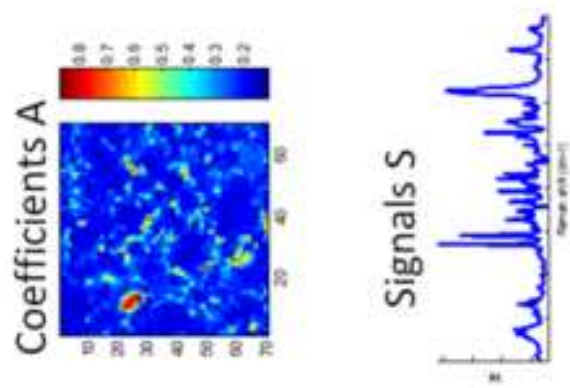
HAL Id: hal-00948530

<https://hal.science/hal-00948530v1>

Submitted on 18 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Tablet



Acquisition area

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.

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

3
4 Mathieu Boiret^{a*}, Douglas N. Rutledge^b, Nathalie Gorretta^c, Yves-Michel Ginot^a, Jean-Michel Roger^c

5
6 ^a Technologie Servier, 27 rue Eugène Vignat, 45000, Orléans, France

7 ^b AgroParisTech, UMR 1145 Ingénierie Procédés Aliments, rue Claude Bernard, F-75005 Paris,
8 France

9 ^c Irstea, UMR ITAP 361, Avenue Jean-François Breton, 34033, Montpellier, France

10
11 * Corresponding author:

12 Tel.: +33 238 238 175

13 E-mail addresses: mathieu.boiret@fr.netgrs.com

14
15 **1. Introduction**

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

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

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

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

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

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

65

66 2. Materials and methods

67 2.1. Samples

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

78 2.2. Raman imaging system

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

85

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

89 2.3. Pre-processing

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

97 2.4. Independent Component Analysis (ICA)

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

$$102 \quad X = A.S \quad (1)$$

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

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

109 To solve the previous equation (Eq. (1)), an unmixing matrix W based on the observation of X needs
110 to be calculated. The output U, constituted by the independent component u₁, u₂, ... u_n should be as
111 independent as possible. For a noise-free ICA model, W should be the inverse of A, and U should be
112 equal to S, according to the following equation:

$$113 \quad U = WX = W(AS) = S \quad (2)$$

114 The mixing matrix A can then be calculated as:

$$115 \quad A = XS^T (SS^T)^{-1} \quad (3)$$

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

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

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

128 2.5. Data analysis

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

131

132 3. Results & discussion

133 3.1. Selection of number of independent components

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

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

152 3.2. Distribution of API

153 An ICA model based on the JADE decomposition with 9 ICs was calculated on the unfolded, SNV
154 pre-processed data cube. The matrices of the proportions, A , for each signal, S , were then folded back
155 in order to obtain a representation of the spatial distribution of each independent component. In figure
156 2, different textures of images can be observed. Indeed, IC1, IC6 and IC9 show very specific
157 inhomogeneous distributions with agglomerates. Considering the different scales of score images, IC2,
158 IC3, IC4, and IC5 have similar textures (or distributions) such as IC7 and IC8 which are the same as
159 that in IC1. It can also be seen that the distributions observed in these two sets of images are
160 complementary, indicating that these two sets of Independent components occupy complementary
161 regions in the tablet. In order to associate an independent component with a chemical compound, the
162 calculated signals were examined.

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

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

177

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

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

188 IC1 is mainly correlated with the spectrum of Avicel. Specific bands due to the chemical bond
189 vibrations are observed in this component (especially between 1250cm^{-1} and 1000cm^{-1} , spectral range
190 linked to CC ring bond stretches and CO stretches). IC7 and IC8 are mainly correlated with the
191 spectrum of avicel (0.38 for IC7 and 0.61 for IC8) but the correlation with lactose (0.36 for IC7 and
192 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
193 IC8) cannot be considered as non-significant. IC7 and IC8 signals are not well defined Raman spectra
194 and contain principally noise or baseline variations which can explain these high correlations with
195 several different products. As can be seen in figure 2, IC7 and IC8 have similar spatial distributions
196 which are the same as that in IC1. Avicel is a microcrystalline cellulose powder which is known as a
197 product providing a fluorescence effect with Raman, which could explain the contribution of IC7 and
198 IC8.

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

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

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

216 In order to explore the ability of ICA to extract a pure signal from lactose, an ICA model was
217 calculated with 5 ICs, which was the known number of constituents used to manufacture the tablet. By
218 comparing the 5 ICs with the pure spectra, a high correlation was found with lactose ($R = 0.90$), one
219 with API1 ($R = 0.95$), one with API2 ($R = 0.94$), while a weak correlation was found with avicel ($R =$
220 0.39) and one signal contained noise and mixed pure contributions. The lactose contribution was
221 therefore not divided among several components, as was observed when using 9 ICs. As is shown in
222 figure 8, the calculated signal IC3 was very similar to the pure lactose spectrum. With 5 ICs, the
223 decomposition of the original matrix was mainly due to chemical variations whereas the
224 decomposition using 9 ICs included physical effects.

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

240

241 **4. Conclusions**

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

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

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

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

270

271 **References**

272 [1] C. Gendrin, Y. Roggo, C. Collet, Pharmaceutical applications of vibrational chemical imaging and
273 chemometrics: a review, *J. Pharm. Biomed. Anal.* 48 (2008) 533-553.

274 [2] J. Rantanen, Process analytical applications of Raman spectroscopy, *J. Pharm. Pharmacol.* 59
275 (2007) 171-177.

276 [3] E. Widjaja, P. Kanaujia, G. Lau, W. K. Ng, M. Garland, C. Saal, A. Hanefeld, M. Fischbach, M.
277 Maio, R.B.H. Tan, Detection of trace crystallinity in an amorphous system using Raman microscopy
278 and chemometrics analysis, *Eur. J. Pharm. Sci.* 42 (2011) 45-54.

279 [4] P. Matousek, A.W. Parker, Non-invasive probing of pharmaceutical capsules using transmission
280 Raman spectroscopy, *J. Raman Spectrosc.* 38 (2007) 563-567.

281 [5] X. Zhang, R. Tauler, Application of Multivariate Curve Resolution Alternative Least Squares
282 (MCR-ALS) to remote sensing hyperspectral imaging, *Anal. Chim. Acta.* 762 (2013) 25-38.

283 [6] Z. Xiabo, Z. Jiewen, M. Holmes, M. Hanpin, S. Jiyong, Y. Xiaopin, L. Yanxiao, Independent
284 component analysis in information extraction from visible/near-infrared hyperspectral imaging data of
285 cucumber leaves, *Chemometr. Intell. Lab. Syst.* 104 (2010) 265-270.

286 [7] Y. Roggo, A. Edmond, P. Chalus, M. Ulmschneider, Infrared hyperspectral imaging for qualitative
287 analysis of pharmaceutical solid forms, *Anal. Chim. Acta.* 535 (2005) 79-87.

- 288 [8] S. Šašić, S. Mehrens, Raman chemical mapping of low-content active pharmaceutical ingredient
289 formulations. III. Statistically optimized sampling and detection of polymorphic forms in tablets on
290 stability, *Anal. Chem.* 84 (2012) 1019-1025.
- 291 [9] M.J. Goetz , G.L. Coté, R. Erckens, W. March, M. Motamedi, Application of a multivariate
292 technique to Raman spectra for quantification of body chemicals, *IEEE Trans Biomed Eng.* 42 (1995)
293 728-731.
- 294 [10] H. Grahn, P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*, John Wiley
295 & son Ltd., New York, 2007.
- 296 [11] M.B. Lopes, J.-C. Wolff, Investigation into classification/sourcing of suspect counterfeit
297 HeptodinTM tablets by near infrared chemical imaging, *Anal. Chim. Acta.* 633 (2009) 149–155.
298
- 299 [12] L. Zhang, Multivariate data analysis for Raman imaging of a model pharmaceutical tablet, *Anal.*
300 *Chim. Acta.* 545 (2005) 262-278.
- 301 [13] B. Vajna, A. Farkas, H. Pataki, Z. Zsigmond, T. Igricz, G. Marosi, Testing the performance of
302 pure spectrum resolution from Raman hyperspectral images of differently manufactured
303 pharmaceutical tablets, *Anal. Chim. Acta.* 712 (2012) 45-55.
- 304 [14] J.V. Stone, *independent component analysis – A tutorial introduction*, A Bradford Book, London,
305 England, 2004.
- 306 [15] Y.B. Monakhova, S.A. Astakhov, A. Kraskov, S.P. Mushtakova, Independent components in
307 spectroscopic analysis of complex mixtures, *Chemometr. Intell. Lab. Syst.* 103 (2010) 108-115.
- 308 [16] G. Wang, Q. Ding, Z. Hou, Independent component analysis and its applications in signal
309 processing for analytical chemistry, *TrAC - Trends Analyt. Chem.* 27 (2008) 368-376.
- 310 [17] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural*
311 *Netw.* 13 (2000) 411-430.
- 312 [18] H. Lin, O. Marjanović, B. Lennox, S. Šašić, I.M. Clegg, Multivariate statistical analysis of Raman
313 images of a pharmaceutical tablet, *Appl. Spectrosc.* 66 (2012) 272-281.
- 314 [19] R.J. Barnes, M.S. Dhanoa, and Susan J. Lister, Standard normal variate transformation and de-
315 trending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.* 43 (1989) 772-777.
- 316 [20] L. De Lathauwer, B. De Moor and J. Vandewalle, An introduction to independent component
317 analysis, *J. Chemom.* 14 (2000) 123-149.
- 318 [21] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, *Neural*
319 *Comput.*, 9 (1997) 1483-1492.
- 320 [22] E.G. Learned-Miller, J.W. Fisher III, ICA using spacings estimates of entropy, *J. Mach. Learn.*
321 *Res.* 4 (2003) 1271-1295.
- 322 [23] J.F. Cardoso, High-order contrasts for independent component analysis, *Neural Comput.* 11
323 (1999) 157-192.

- 324 [24] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, Independent Component Analysis with the JADE
325 algorithm, *TrAC - Trends Analyt. Chem.* 50 (2013) 22-32.
- 326 [25] D. Jouan-Rimbaud Bouveresse, A. Moya-González, F. Ammari, D.N. Rutledge, Two novel
327 methods for the determination of the number of components in independent component analysis
328 models, *Chemometr. Intell. Lab. Syst.* 112 (2012) 24-32.
- 329 [26] <http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m>
- 330 [27] A. Savitzky et M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares
331 Procedures, *Chemometr. Anal. Chem.* 36 (1964) 1627-1639.
- 332

333 **Figure captions**

334

335 Figure 1 - Lowest correlation between signals obtained using ICA_by_blocks

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

338

339 Figure 2 - Proportions coefficients (A) of each IC

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

342

343 Figure 3 - Signals, S, of the ICA model

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

345

346 Figure 4 - Pure spectra of the drug product constituents

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

350

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

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

354

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

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

358

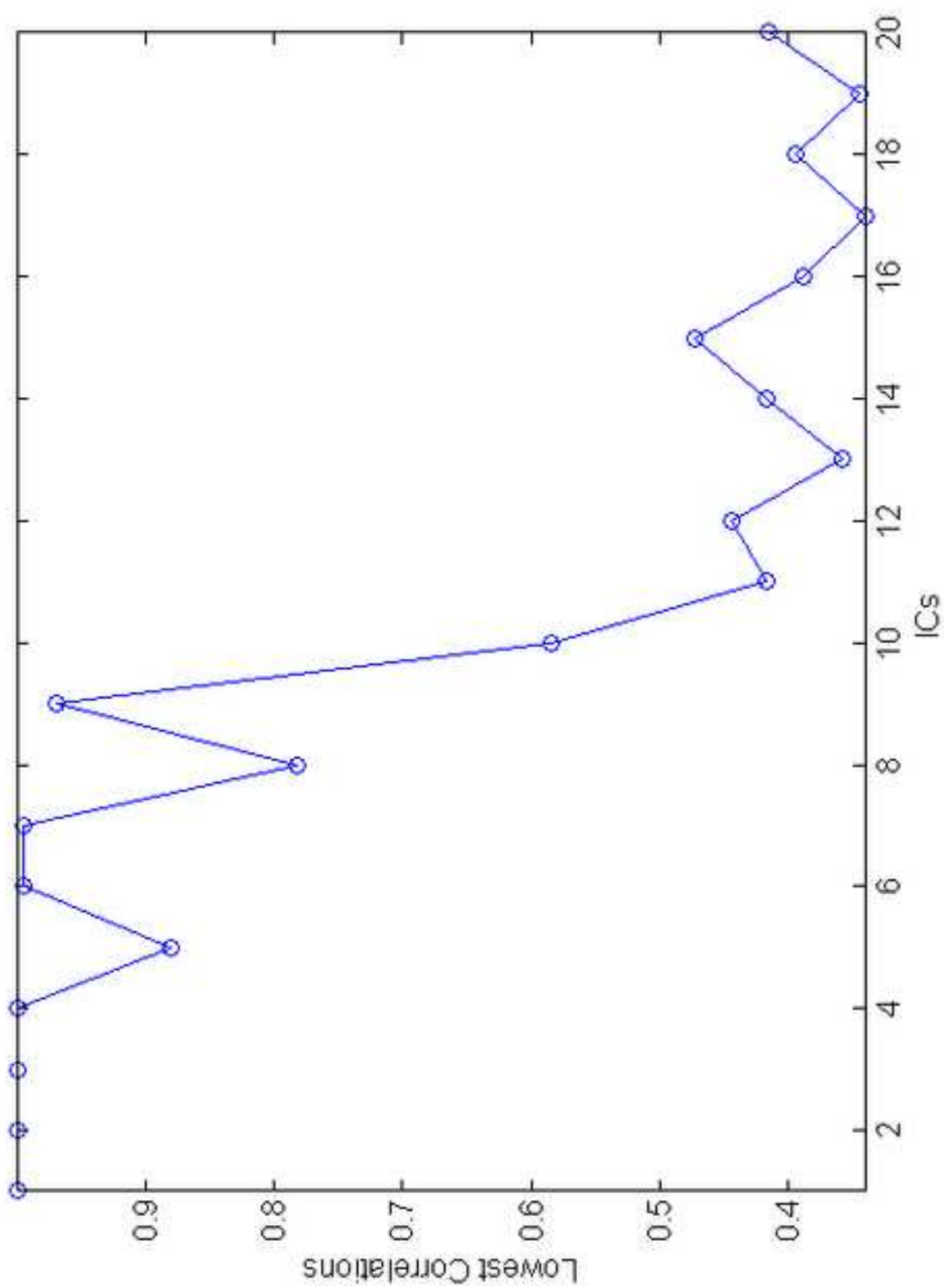
359 Figure 7 - Calculated signal of independent component 2, 3, 4, 5 plotted with the spectrum of Lactose

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

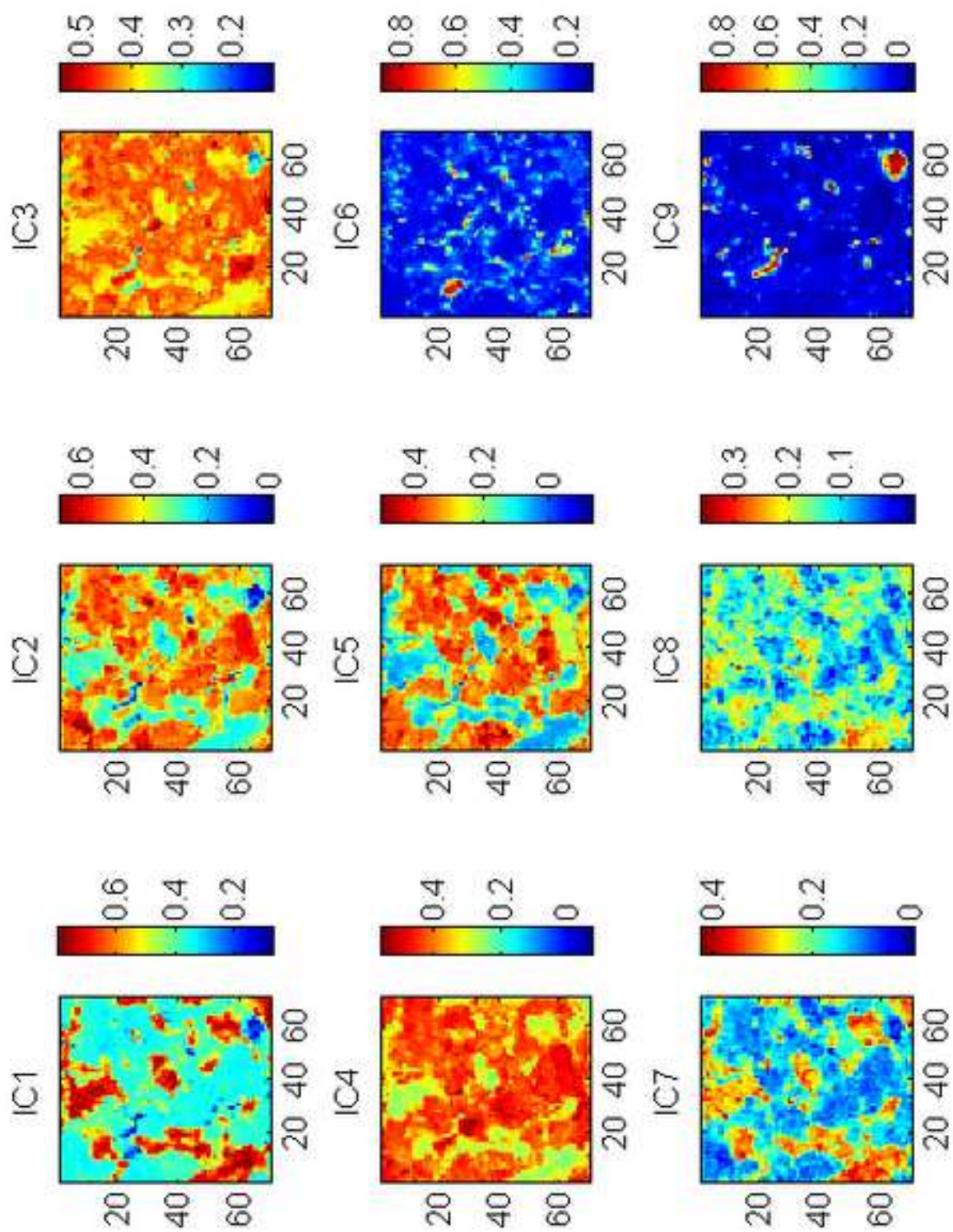
364

- 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.

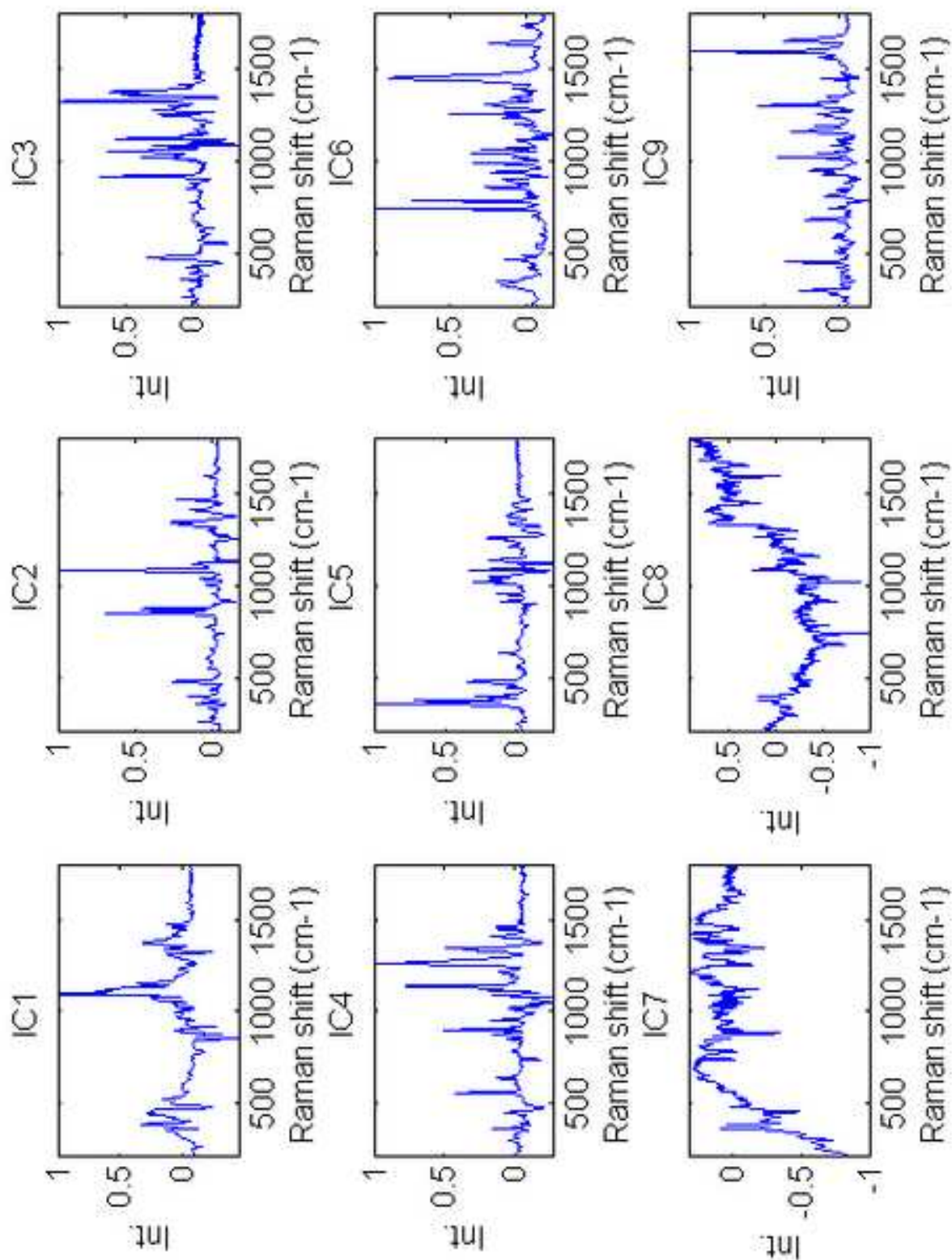
Pure spectrum	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	IC9
API1	0.01	-0.09	0.07	0.14	-0.04	0.13	0.21	0.18	0.92
API2	0.06	-0.01	0.11	0.08	0.03	0.96	0.08	0.10	-0.06
Lactose	0.25	0.44	0.23	0.25	0.47	0.00	0.36	0.45	-0.17
Avicel	0.49	0.15	0.06	0.02	0.20	-0.07	0.38	0.61	-0.20
Magnesium Stearate	0.20	0.00	0.01	0.04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

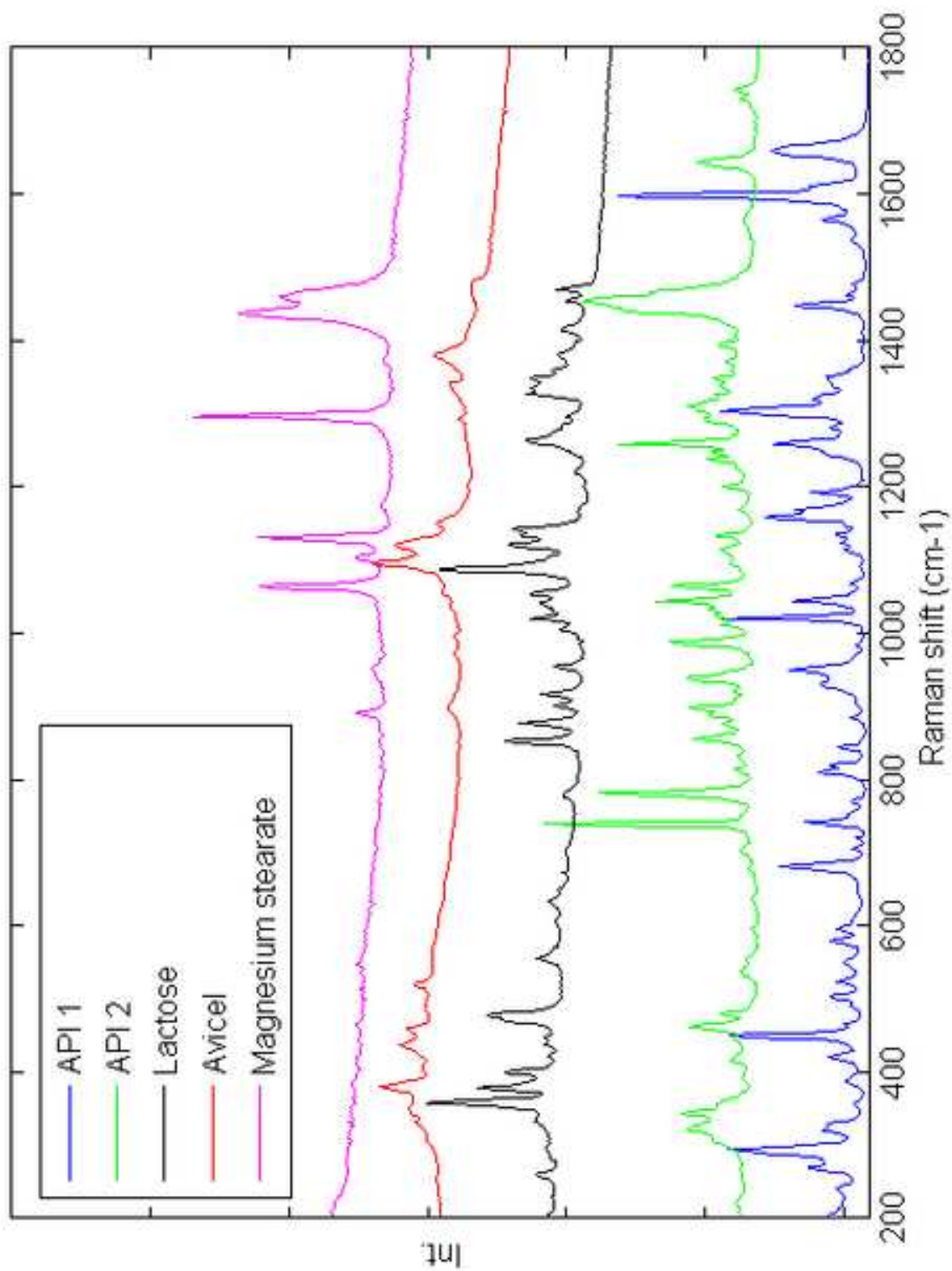


Figure(s)
[Click here to download high resolution image](#)

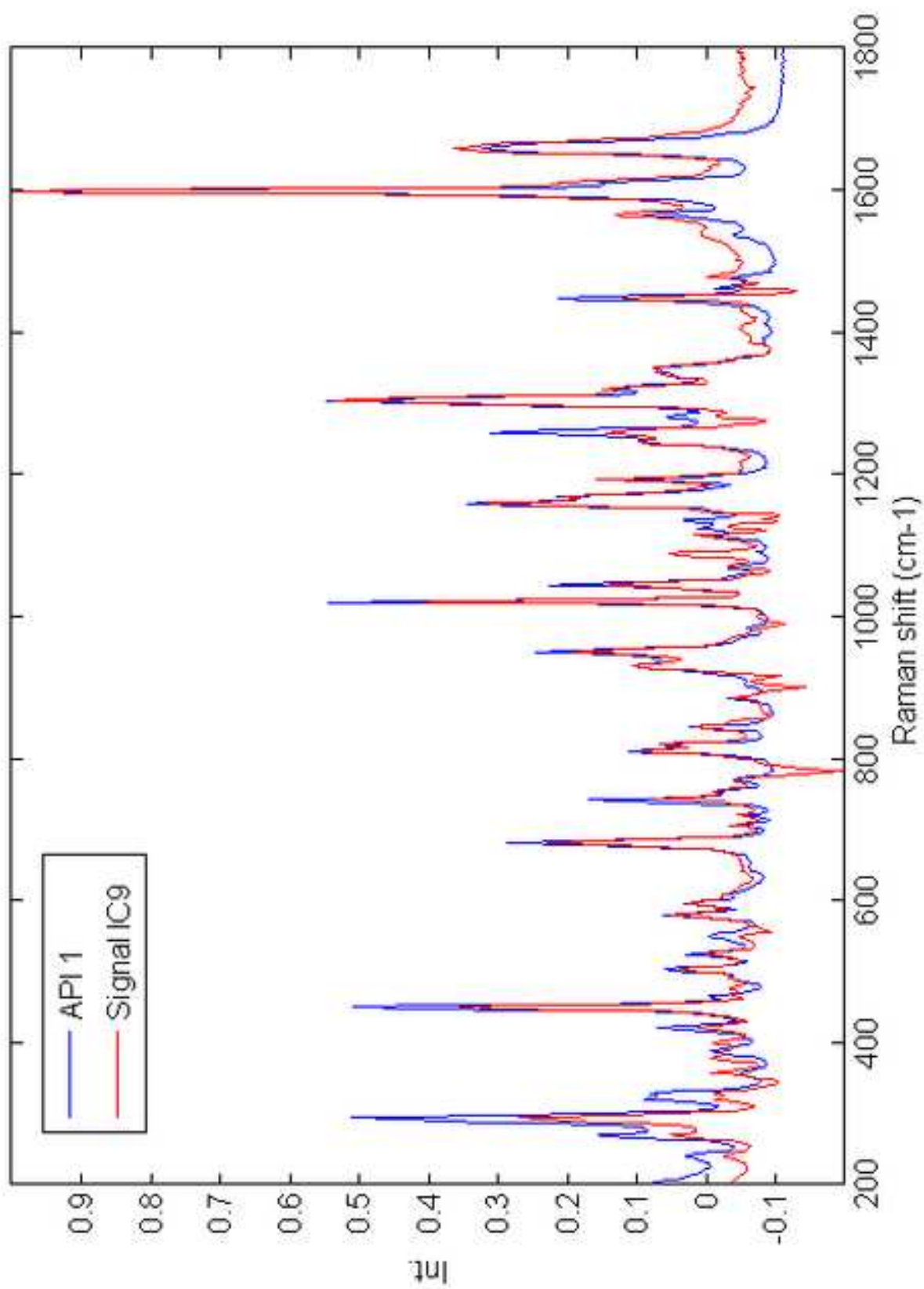


Figure(s)
[Click here to download high resolution image](#)

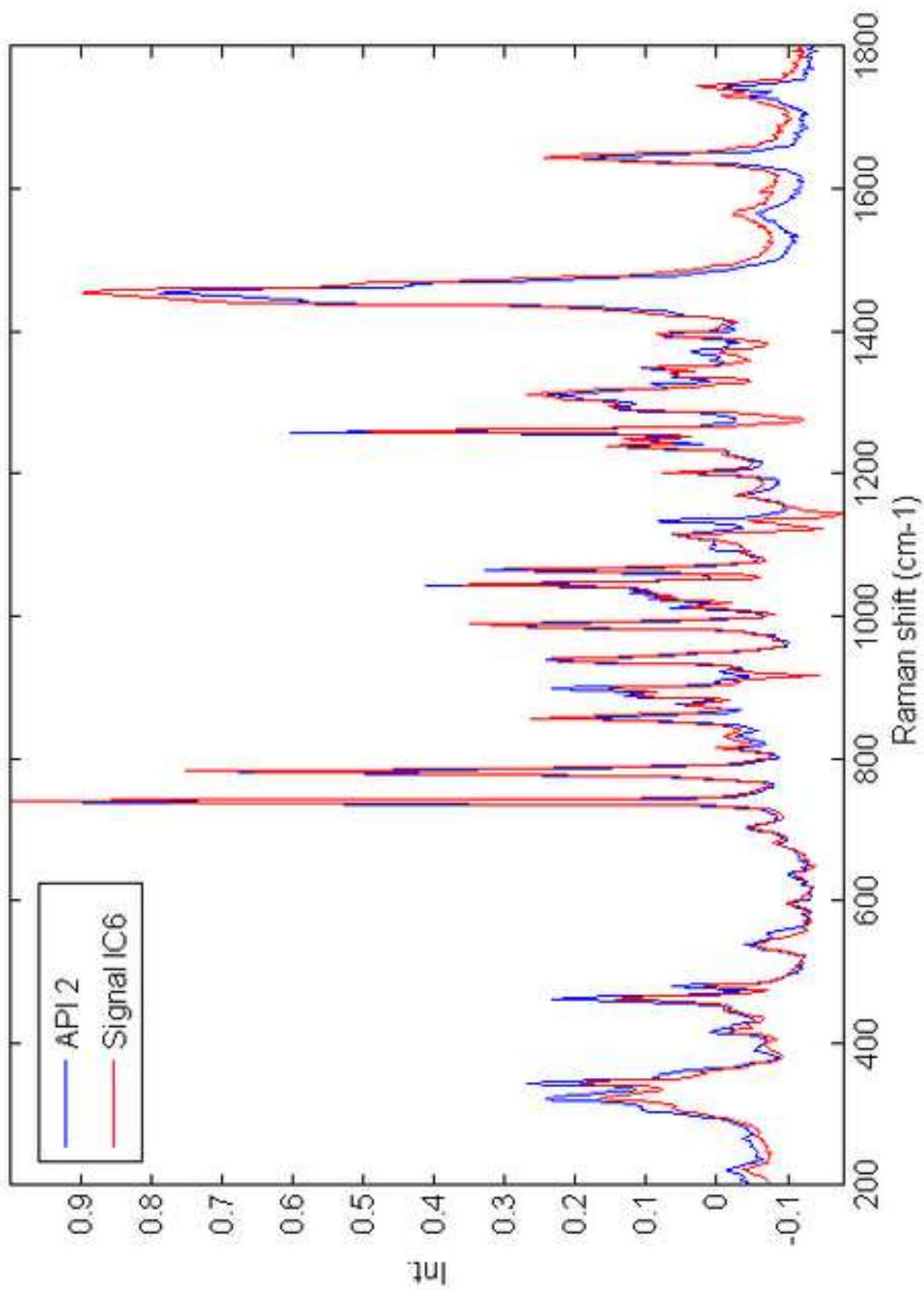




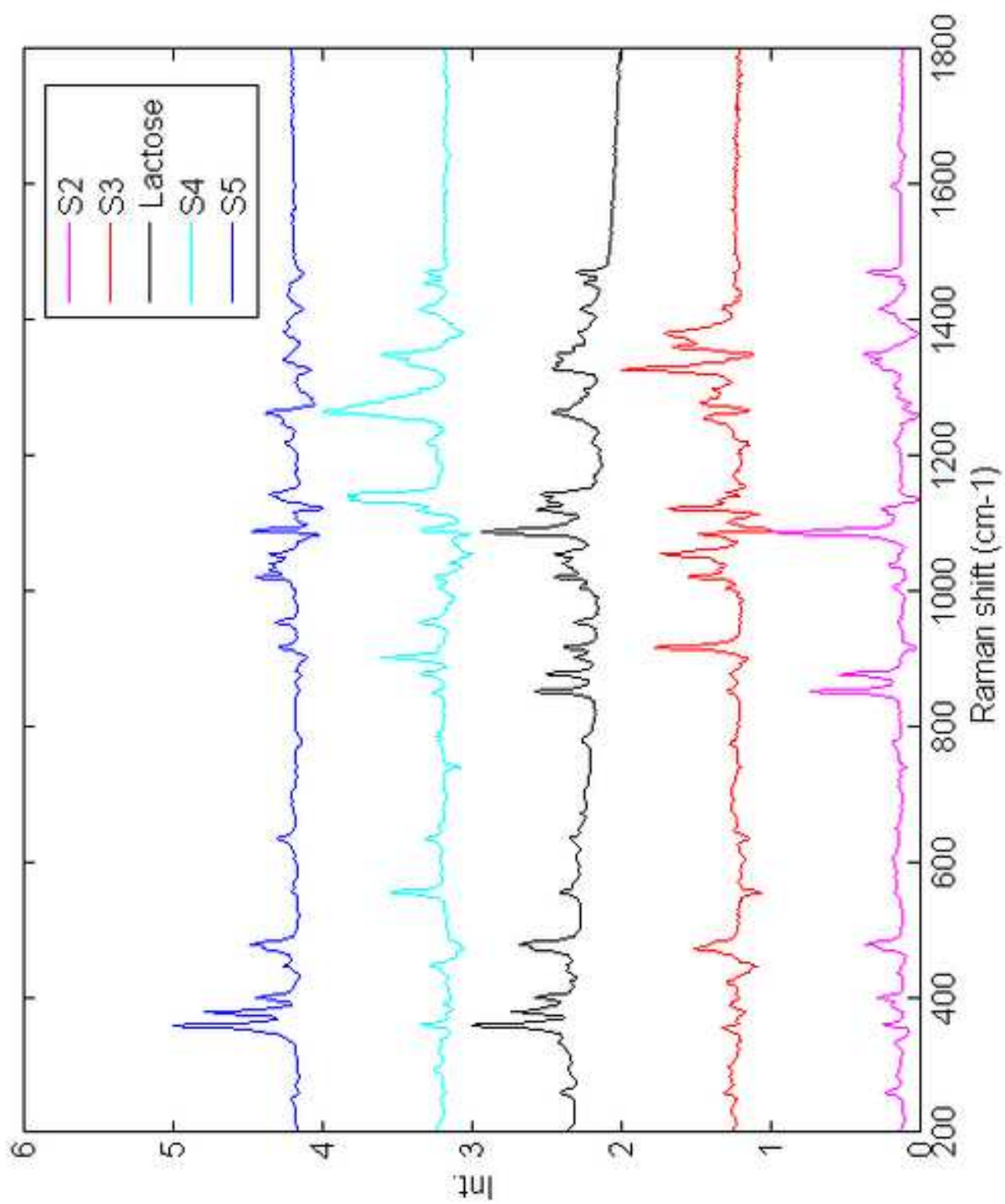
Figure(s)
[Click here to download high resolution image](#)



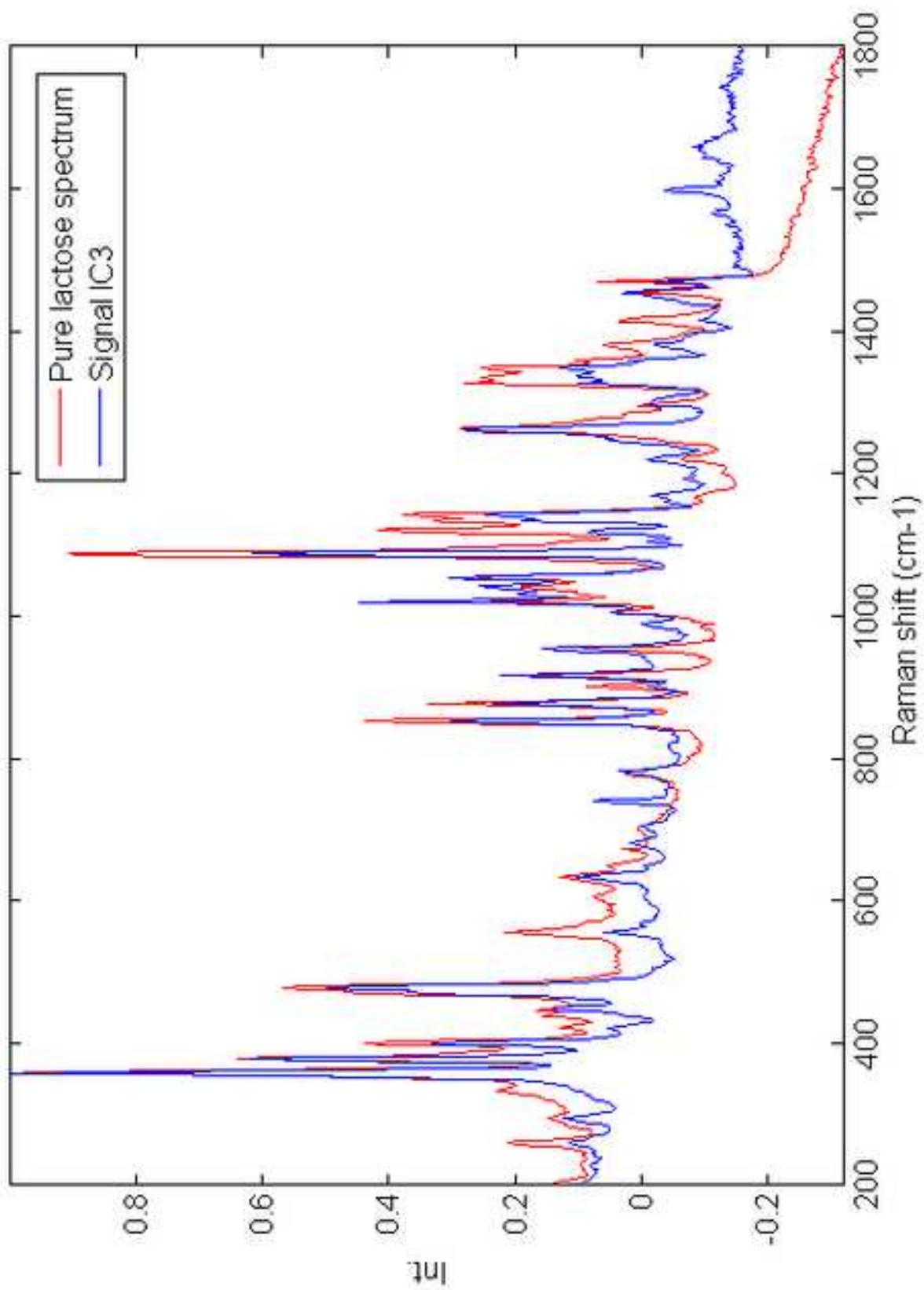
Figure(s)
[Click here to download high resolution image](#)



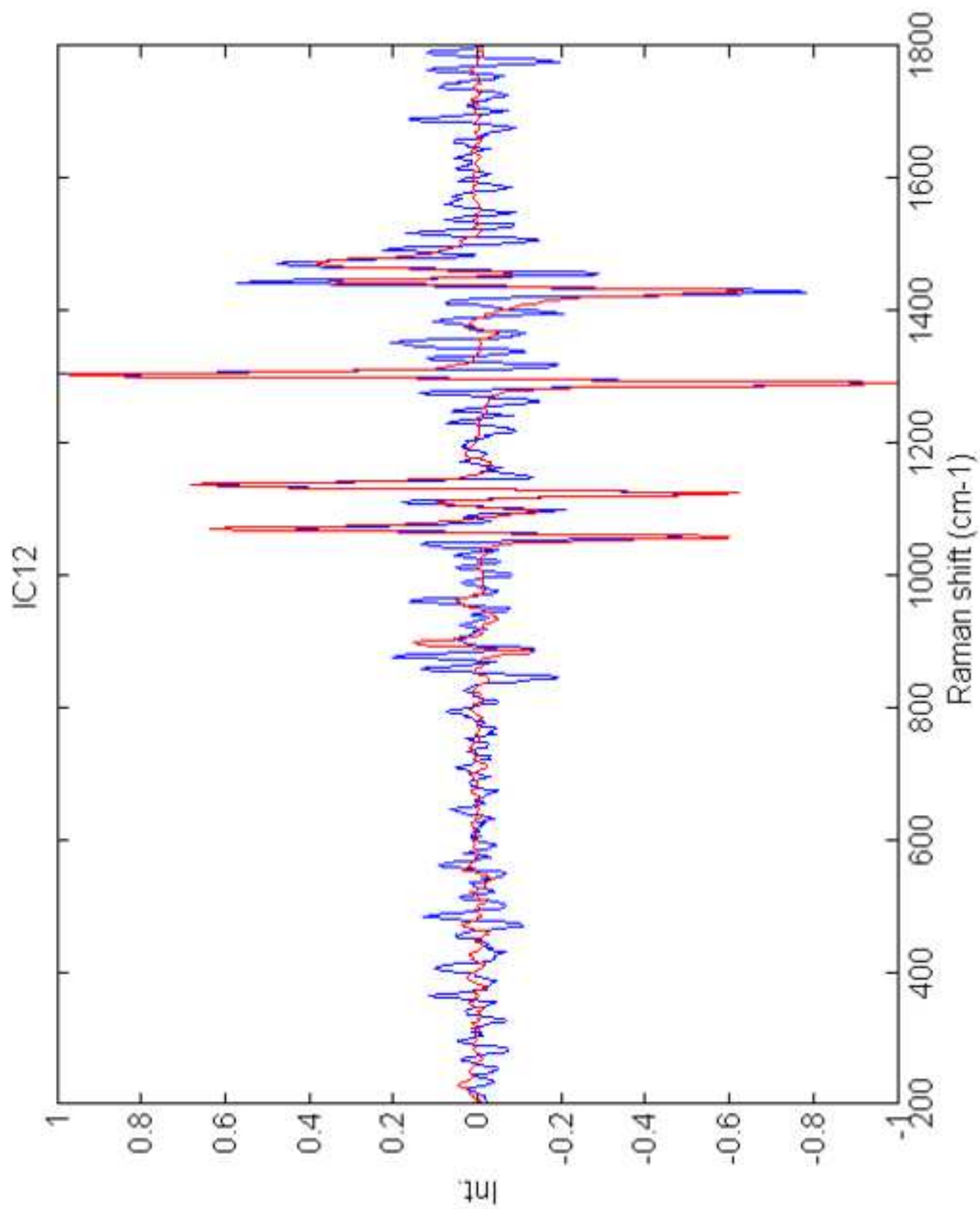
Figure(s)
[Click here to download high resolution image](#)



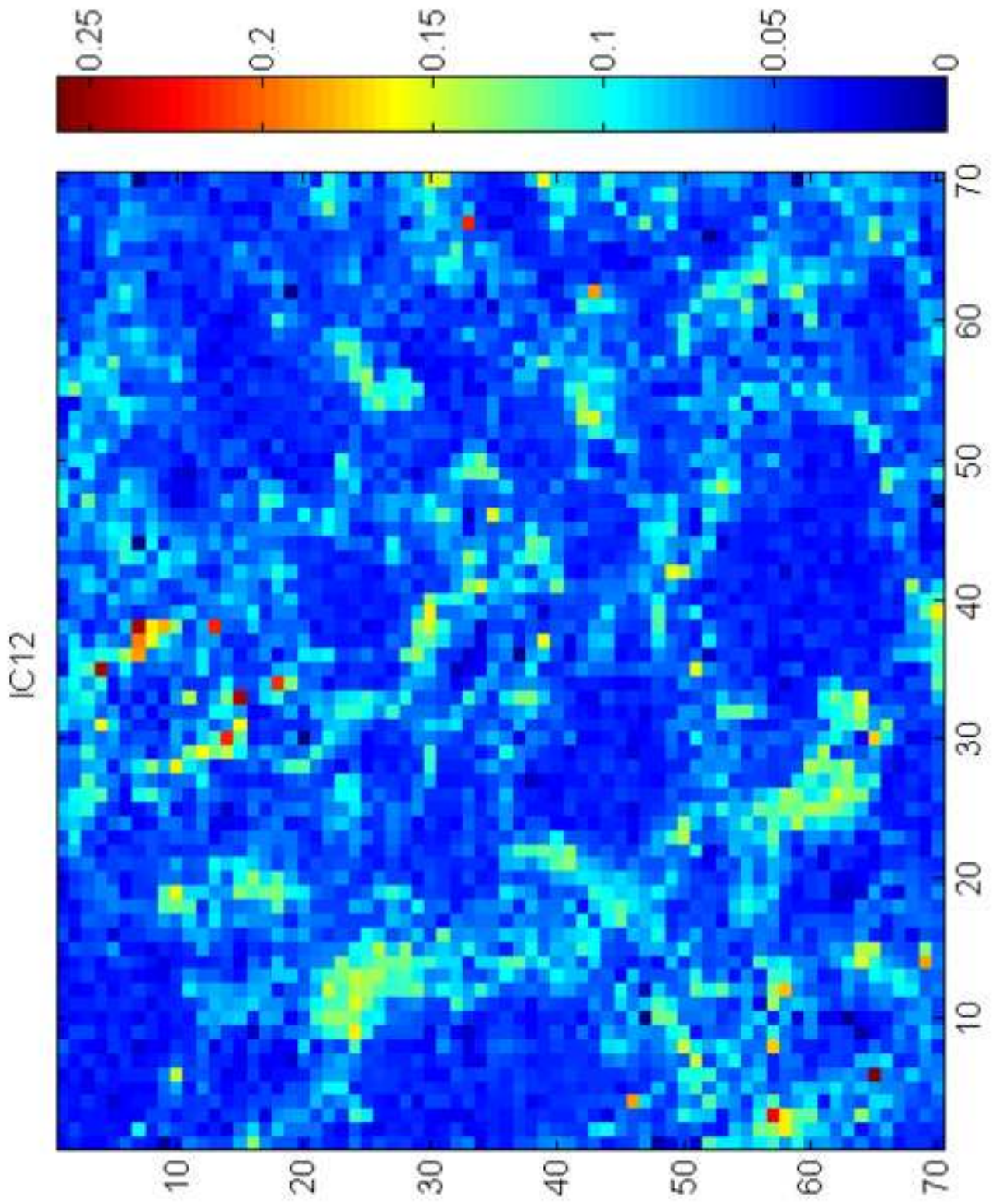
Figure(s)
[Click here to download high resolution image](#)



Figure(s)
[Click here to download high resolution image](#)



Figure(s)
[Click here to download high resolution image](#)



Figure(s)
[Click here to download high resolution image](#)