Automated resolution of chromatographic signals by independent component analysis - orthogonal signal deconvolution in comprehensive gas chromatography/mass spectrometry-based metabolomics

Xavier Domingo-Almenara[a,c,*], Alexandre Perera[b], Noelia Ramírez[a,c], Jesus Brezmes[a,c]

[a] *Metabolomics Platform - IISPV, Department of Electrical and Automation Engineering (DEEEA). Universitat Rovira i Virgili, Tarragona, Catalonia (Spain).*

[b] *B2SLAB. Department d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, CIBER–BBN, Universitat Politècnica de Catalunya, Barcelona, Catalonia, (Spain)*

[c] *CIBERDEM, Biomedical Research Networking Center in Diabetes and Associated Metabolic Disorders, Madrid (Spain).*

Abstract

Comprehensive gas chromatography - mass spectrometry (GCxGC-MS) provides a different perspective in metabolomics profiling of samples. However, algorithms for GCxGC-MS data processing are needed in order to automatically process the data and extract the purest information about the compounds appearing in complex biological samples. This study shows the capability of independent component analysis - orthogonal signal deconvolution (ICA-OSD), an algorithm based on blind source separation and distributed in an R package called *osd*, to extract the spectra of the compounds appearing in GCxGC-MS chromatograms in an automated manner. We studied the performance of ICA-OSD by the quantification of 38 metabolites through a set of 20 Jurkat cell samples analyzed by GCxGC-MS. The quantification by ICA-OSD was compared with a supervised quantification by selective ions, and most of the R^2 coefficients of determination were in good agreement (R^2>0.90) while up to 24 cases exhibited an excellent linear relation (R^2>0.95). We concluded that ICA-OSD can be used to resolve co-eluted compounds in GCxGC-MS.

Keywords: comprehensive gas chromatography, orthogonal signal deconvolution, multivariate curve resolution, compound deconvolution, independent component analysis.

1. Introduction

Metabolomics is the study of low molecular weight compounds in biological systems [1]. Particularly, metabolomics focuses on comparing healthy versus metabolomic disease organisms and, therefore, it attempts to discover predictive biomarkers by detecting early biochemical changes before the appearance of the disease [2]. For that purpose, metabolomics experimental designs include non-targeted analysis of the samples as there is no prior knowledge of the metabolites that may be involved not only in fully developed metabolomic diseases, but also in pre-symptomatic stages.

[*] Corresponding author: xavier.domingo@urv.cat. Phone +34 977559619. Fax: +34 977559605

Analytical techniques to identify and quantify metabolites include the best-established gas chromatography-mass spectrometry (GC-MS). Gas chromatography separates the compounds contained in a sample while passing through a chromatographic column. However, when two or more compounds do not completely separate chromatographically, those compounds are known to be co-eluted, and this clearly affects the correct quantification and identification of the metabolites. In that sense, comprehensive gas chromatography - mass spectrometry (GCxGC-MS) [3, 4] was devised to minimize co-elution. In GCxGC-MS, the sample pass through two chromatographic columns with orthogonal polarity properties, which improves the compound separation and it leads to an increased compound detection capacity as co-elution is diminished.

However, compounds in the samples usually appear at trace levels and different sources of noise derived from the instrument and the sample biological matrix may interfere with the correct identification of the compounds. In the same way, GCxGC-MS generates large quantity of data and its interpretation can not be conducted manually. In that sense, GCxGC-MS data processing algorithms are needed to turn the chromatographic signals into interpretable biological information. Besides, GCxGC-MS samples are composed by a large amount of data in comparison with GC-MS samples, and algorithms for GCxGC-MS data processing should be optimized for a fast data processing.

As reviewed in [5], some of the existing data processing algorithms that can be applied to resolve mixtures in comprehensive gas chromatography include PARAFAC [6] and multivariate curve resolution - alternating least squares (MCR-ALS) [7]. Contrarily to MCR, PARAFAC can be only applicable to a three-way data set, i.e., PARAFAC can not resolve a single GCxGC-MS sample.

In the past years, independent component analysis (ICA) [8] has been introduced as an alternative to the traditional MCR for GC-MS data analysis [9, 10, 11]. ICA is a blind source separation (BSS) technique used to separate linearly mixed sources, i.e., it is capable of separating and retrieve the original compound sources - elution profile or spectra - from a mass spectra chromatogram. Whereas MCR–ALS resolves a chromatographic mixture by minimizing the residual error between the data and the predicted model, ICA uses another type of measure which is the statistical independence, and it estimates the original compound sources by maximizing the independence between components. ICA is widely applied in biomedical sciences, including data processing in electroencephalography recordings [12, 13, 14], and it is also one of the most reported algorithms for resolution of spectroscopy mixtures. More recently, we have developed a new method known as independent component analysis - orthogonal signal deconvolution (ICA-OSD) [15], embedded in an R package, that uses a combination of ICA and principal component analysis (PCA) to identify co-eluted compounds in GC-MS. In ICA-OSD, PCA is proposed as an alternative to the typical use of least squares (LS) in MCR-ALS. The application of LS for spectra extraction has different drawbacks, detailed in [15], which can be summarized in the fact that no correlation or covariance information is taken into account when applying LS, and therefore LS may find difficulties in distinguishing noise and the different compound fragments. This may lead to introducing a bias into the LS regressors specially in situations of co-elution or under undue biological matrix interference. Besides, whereas the current ICA-based methods consider the spectra as the independent source in the chromatograms, in ICA-OSD we implemented a different approach where we assumed that the elution profile was the independent source, as opposite to the spectra. In that sense, we used ICA to extract the elution profiles and then determine the spectra by means of OSD. Finally, ICA-OSD shown itself as a computationally faster alternative to MCR-ALS. Up to the date, the capability of independent component analysis - orthogonal signal deconvolution for compound quantification in chromatographic signals has not been studied.

In this paper we propose an automated method to deconvolve compounds appearing in GCxGC-MS samples by independent component analysis - orthogonal signal deconvolution.

2. Materials and methods

2.1. Materials

The performance of ICA-OSD was evaluated through a set of 38 metabolites appearing in 20 Jurkat cell samples extracted from human acute T cell lymphoblastic leukemia cell line Jurkat. The samples of this experiment were previously used to report the intersection of phosphoethanolamine with menaquinone-triggered apoptosis by Styczynski *et al.* [16]. More details on the dataset, sample preparation and methods can be found in the original study.

2.2. Data analysis and pre-processing

ICA-OSD was used to automatically extract and deconvolve the compounds concentration profiles and spectra. The GCxGC-MS chromatograms were processed by analyzing each modulation cycle separately. Each modulation cycle was first divided in chromatographic peak features (CPFs) using the same criteria as in [17]. The different CPFs contained several compounds, so the algorithm had to deconvolve them in case of co-elution. The number of factors or components for ICA was determined by evaluation of residual sum of squares (described in Section 3.2).

The chromatograms were automatically processed by ICA-OSD. From the ICA-OSD output we only took into account those metabolites appearing in at least 15 of the 20 samples, so a total of 38 compounds with KEGG number (Kyoto Encyclopedia of Genes and Genomes) were identified. Metabolite identities were curated by spectral similarity with the reference spectra and retention index error by retention time standardization using fatty acid methyl esters (FAME) standards. However, the identity was not confirmed with the analysis of reference standards and therefore, the list of identified metabolites is putative, and a name is assigned to facilitate the interpretation of the results. For this sub-set of 38 compounds, reference relative compound concentration - relative across samples - was determined by the area of a selective ion. The most selective ion was manually determined for each compound.

The spectra determined by ICA–OSD were compared using the dot product [18] against the Golm Metabolome Database (GMD) [19] MS spectra library. The masses 73, 74, 75, 147, 148, and 149 m/z were excluded before processing the sample, since they are ubiquitous mass fragments typically generated from compounds carrying a trimethylsilyl moiety [19]. They were also excluded in the identification. Only the fragments from m/z 70 to 600 were taken into account when comparing reference and empirical spectra, since this is the m/z range included in the downloadable GOLM database. Also, chromatographic signals were filtered using a Savitzky–Golay filter [20]. The ICA algorithm used was the joint approximate diagonalization of eigenvalues (JADE) [21].

3. Computational methods and theory

This section describes the ICA-OSD algorithm together with the methodology to determine the number of compounds.

3.1. Resolution of GCxGC-MS mixtures by independent component analysis – orthogonal signal deconvolution

Orthogonal signal deconvolution (OSD) is a multivariate method which purpose is to extract and deconvolve the spectrum of a given compound only with the information relative to the compound elution profile. OSD is based on principal component analysis, avoiding thus, the use of least squares used in multivariate curve resolution - alternating least squares (MCR-ALS). Here, the elution profiles are determined by ICA to later determine the spectra using OSD, and in this manner we will refer the complete approach as ICA-OSD.

ICA is mathematically expressed as:

$$X = AZ^T \tag{1}$$

where X (N×M) is the matrix containing the mixture of compounds, A (N×k) is the mixing matrix and Z^T(k×M) is the source matrix. N and M are the number of rows and columns of the data matrix X, and k denotes the number of components or compounds in the model. Each row in X holds a m/z channel whereas each column holds the retention time scans. ICA decomposes the data matrix by finding the independent sources contained in X.

As mentioned above, generally ICA-based approaches are based on extracting first the spectra using ICA - the spectra are considered the independent sources - to later estimate the elution profile using different approaches. In our ICA-OSD implementation, the elution profiles of the compounds are considered the independent sources and thus Z^T holds the elution profile for each compound. Since the elution profiles determined by ICA may be affected by the ICA ambiguity of negativity, the sources in Z^T that express more negative variance than positive are negatively rotated. Moreover, all the components in Z^T are submitted to unimodality constraint to force one local maxima per source. ICA has a second ambiguity related to variance (energy) indetermination, which means that the energy of the recovered compound profiles do not correspond to the real energy of that component. To overcome that, a least squares regression is performed with the estimated sources hold in Z^T against the base ion chromatogram of the matrix X. The base ion chromatogram or BIC is determined by representing the maximum *m/z* value for each point in the chromatogram.

Once the elution profiles are determined, OSD is applied to extract each corresponding spectra. In OSD, an X'_j sub-data matrix is determined for each compound j in Z^T. This sub-data matrix comprises only the data from X in which the compound profile in Z^T_j is non-zero - the elution profile in Z^T is used as a mask to suppress the surrounding data non-related to the compound -. A PCA is performed over the sub-data matrix to determine the spectra associated to each compound. PCA can be mathematically expressed as:

$$X'_j = YW^T \tag{2}$$

where X′(N×M) is the sub-data matrix to decompose, Y(N×M) is the score matrix and W(M×M) is the loading or eigenvectors matrix. For each compound profile, the PCA decorrelates the information of the sub-data matrix and decomposes it into a matrix W^T (Eq. 3) which is a set of orthogonal spectra and a matrix Y which is associated to the retention time covariance response for each spectrum in W^T. The matrix W^T holds the spectra of the compound of interest together with the spectra of the different sources of noise - such as co-eluted substances or biological matrix interference -. To determine which spectrum is related to the compound of interest we compute the correlation between the profile of the compound in $Z^T{}_j$ and the information of the covariance responses determined by the PCA in Y. The component with the highest absolute correlation is the candidate spectra for the compound of interest.

OSD can be summarized in the following steps:

1. Given a Z_j^T compound elution profile, determine a X_j sub-data matrix comprised only of the data of the retention time in which the compound is eluting.

2. Apply a PCA over X_j. The result is a score matrix Y and loading matrix W.

3. Determine the correlation coefficient between Z_j^T and each component in Y and select the component h with the highest absolute correlation value.

4. Select the component h in W, rotate W_h according to the sign of the previous determined correlation coefficient, and clip to zero all the negative values. W_h is now considered to be the spectrum of Z_j^T.

After the spectra are determined, the elution profiles are refined by the application of a NNLS regression of all the spectra against the data matrix X.

3.2. Determination of number of components

To define the ICA model, it requires a fixed number of components. The number of components is closely related to the number of compounds present in the mixture, as usually the model to define the data is not only constructed by pure compounds but also by baseline, noise, or other interferences. An iterative residual sum of squares (RSS) approach was used to automatically determine the number of components for the ICA model. The RSS can be expressed as:

$$RSS(k) = \sum_{i=1}^{N}(X - X^*(k))^2 \tag{3}$$

where, X is the original mixture matrix, $X^*(k)$ is the resolved matrix by ICA-OSD using k components, and N is the total length of the unfolded X matrix. For each k in $k = 1, 2, ..., N$, ICA-OSD resolves the X data with k components and it determines the RSS. This method yields a decreasing RSS curve that tends to a minimum. The proper number of factors is determined when the addition of more components does not significantly decrease the explained variance, i.e., when the RSS error reaches a certain threshold.

4. Results and discussion

The chromatographic data was automatically processed with our proposed method ICA-OSD. Metabolites eluting in more than one modulation cycle were associated based on their identity and quantified together (sum of concentrations). The metabolites across samples were aligned also based on their identity. Table 1 shows the list of the identified compounds along with their 1st and 2nd retention times and the identification match factor (MF). The identification match factor is determined by dot product between the averaged compound spectra across samples and the reference spectra (Golm Metabolome Database GMD). The closer the score to one hundred, the more exact and pure the spectra extracted. The table also shows the linear regression coefficient of determination (R^2) between our empirical method ICA-OSD and the selective ion area (reference model). In order to demonstrate the ICA-OSD quantification capability along a wide dynamic range of metabolite concentration, we determined the relative compound concentration (Rel. C.) which is the quotient between the mean concentration of each compound and the mean concentration of all the compounds listed in the table.

In this study, we use the coefficient of determination R^2 as a metric to describe the relative deviation between our proposed method for quantification (ICA-OSD) and our reference model (selective ion). From the given results, most of the R2 coefficients are in good agreement (R^2>0.90) while up to 24 cases exhibit an excellent linear relation (R^2>0.95). Overall, ICA-OSD conducted a reliable quantification of compounds even when those occurred at low concentration or appeared co-eluted.

The efficiency of ICA-OSD is directly conditioned by the degree of noise and co-elution with other compounds. To illustrate this, and the operation of ICA-OSD for compound deconvolution we shown two different examples of co-elution situations in GCxGC-MS. Figure 1 shows the total ion chromatogram (BIC) in dotted grey line, and the resolved compound elution profiles by ICA-OSD in color lines, of two selected retention time windows from different modulation cycles.

In Figure 1 (a), three compounds appear under the same chromatographic peak, those three compounds were resolved by ICA-OSD and one of them was identified as erythritol (4TMS). Similarly, in Figure 1 (b) three compounds appear co-eluted but resolved by ICA-OSD; on of them was identified as myo-inositol (6TMS). The resolved spectra for erythritol and myo-inositol are shown in Figure 2 where we can visually compare the empirical (black and positive) and the reference (color and negative) spectra. In both cases ICA-OSD successfully extracted the spectra needed to properly identify both compounds. In the Figure 1 (a) case, erythritol appears low concentrated and in co-elution with a more intense compound. Despite that, ICA-OSD is capable of extracting a sufficient pure spectrum to allow a correct identification, with a match score of 98 % - for the given sample case -. In Figure 1 (b), myo-inositol appears strongly interfered by another more concentrated compound. As a result, ICA-OSD fails in correctly associate the fragments between m/z 100 and 150 (Figure 2 (b)), which appears in the reference spectrum but they do not appear in the empirical spectrum. Also, the ions m/z 305 and 318 appears to be interfered, and their relative intensities differ from the reference pattern. Consequently, the match score of myo-inositol in this given case is 87 %. This is a clear example of the problems for the correct identification of metabolites that co-elution brings. The identification performance can be assessed also in an example of a set of spectra extracted by ICA-OSD shown in Figure 3, where we can visually compare the empirical (black and positive) and the reference (color and negative) spectra for each compound. The figure shows the spectra extracted for lactic acid (2TMS), phosphoric acid (3TMS), fumaric acid (2TMS) and glycerol (3TMS), and this exemplifies the capability of ICA-OSD to successfully extract spectra from chromatographic mixtures.

As mentioned before, one of the most important factors that difficulties the identification is co-elution. In those cases, the spectrum of each compound has to be correctly separated - resolved or deconvolved - from co-eluted compounds or other noise interferences. Despite that one of the differential characteristics of GCxGC-MS with respect to GC-MS is the reduction of the co-elution problem, we still find co-eluted peaks across the second retention time dimension. Here we show how ICA-OSD is also an effective method for the resolution of chromatographic signals including those generated by GCxGC-MS. Due to noise and other interferences, OSD may fail in correctly classify the m/z when deconvolving spectra. This means that OSD would fail in associating a certain m/z to a compound where other methods based on least squares, such as MCR-ALS would probably not, as OSD is a more conservative approach . On the contrary, OSD brings more accuracy generally in co-eluted situations as attempts to differentiate which ions correspond to the compound of interest [15].

Here we applied ICA-OSD in each modulation cycle separately. We later grouped the compounds appearing in different modulation cycles according to their identity. This may also affect the quantification of compounds as the same compound can be identified with a different name between or within samples. Automatic alignment or grouping of compounds between and within samples after deconvolution is still an important problem that has to be tackled.

5. Conclusions

We previously shown that ICA-OSD was able to successfully extract the spectra from co-eluted compounds in GC-MS [15], but the capability of ICA-OSD to quantify metabolites was not evaluated. In this study we evaluate a method to automatically resolve chromatographic data in GCxGC-MS samples with ICA-OSD. Besides, ICA-OSD is an efficient method in terms of speed of execution as previously shown in [15], which is an important advantage for GCxGC-MS data processing due to the large amount of data that metabolomics experiments generate with this analytical platform. This study concludes that ICA-OSD can be used to resolve co-eluted compounds in GCxGC/MS-based metabolomics samples.

Acknowledgements: We would like to show our gratitude to the Dr. M. Styczynski and Mrs. A. Su from the School of Chemical & Biomolecular Engineering (Georgia Institute of Technology), for their support and for providing the GCxGC raw data, and to Dr. R. Ras and Mrs. S. Mariné from the Center for Omics Sciences in Reus for their constant support and scientific advising. We also thank the anonymous reviewers of the earlier versions of this manuscript for their constructive criticisms and suggestions that allowed us to significantly improve the paper. This research was partially funded by MINECO grant TEC2012–31074 and TEC2015-69076-P (to JB) and TEC2013–44666–R and TEC2014-60337-R (to AP). CIBER–BBN is an initiative of the Spanish ISCIII. NR acknowledges the financial support of the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement (No. 660034). XD also acknowledge the University Rovira and Virgili Martí and Franquès grants for the financial support.

Availability: The *osd* R package can be freely downloaded from http://www.metabolomicsplatform.com/applications.

[1] G.J. Patti, O. Yanes, G. Siuzdak Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13 (4):63–269, 2012.

[2] Aihua Zhang, Hui Sun, and Xijun Wang. Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Analytical and Bioanalytical Chemistry*, 404(4):1239–1245, September 2012.

[3] Luigi Mondello, Peter Quinto Tranchida, Paola Dugo, and Giovanni Dugo. Comprehensive two-dimensional gas chromatography-mass spectrometry: a review. *Mass Spectrometry Reviews*, 27(2):101–124, April 2008.

[4] John V. Seeley and Stacy K. Seeley. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Analytical Chemistry*, 85(2):557–578, 2012.

[5] J. T. V. Matos, Regina M. B. O. Duarte, and Armando C. Duarte. Trends in data processing of comprehensive two-dimensional chromatography: State of the art. *Journal of Chromatography B*, 910:31–45, December 2012.

[6] Nicolaas (Klaas) M. Faber, Rasmus Bro, and Philip K. Hopke. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemometrics and Intelligent Laboratory Systems*, 65(1):119–137, January 2003.

[7] A. de Juan, J. Jaumot, R. Tauler. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Anal. Methods*, 6:4964, 2014.

[8] Stephen Roberts and Richard Everson. *Independent Component Analysis: Principles and Practice*. Cambridge University Press, March 2001.

[9] Guoqing Wang, Wensheng Cai, and Xueguang Shao. A primary study on resolution of overlapping GC-MS signal using mean-field approach independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, 82(1-2):137–144, May 2006.

[10] Zhichao Liu, Wensheng Cai, and Xueguang Shao. Sequential extraction of mass spectra and chromatographic profiles from overlapping gas chromatography-mass spectroscopy signals. *Journal of Chromatography A*, 1190(1-2):358–364, May 2008.

[11] Xueguang Shao, Zhichao Liu, and Wensheng Cai. Resolving multi-component overlapping GC-MS signals by immune algorithms. *TrAC Trends in Analytical Chemistry*, 28(11):1312–1321, December 2009.

[12] Silvia Comani, Dante Mantini, Paris Pennesi, Antonio Lagatta, and Giovanni Cancellieri. Independent component analysis: fetal signal reconstruction from magnetocardiographic recordings. *Computer Methods and Programs in Biomedicine*, 75(2):163–177, August 2004.

[13] F. J. Martinez-Murcia, J. M. Gorriz, J. Ramirez, C. G. Puntonet, and I. A. Illan. Functional activity maps based on significance measures and Independent Component Analysis. *Computer Methods and Programs in Biomedicine*, 111(1):255–268, July 2013.

[14] S. Spasic, Lj. Nikolic, D. Mutavdzic, and J. Saponjic. Independent complexity patterns in single neuron activity induced by static magnetic field. *Computer Methods and Programs in Biomedicine*, 104(2):212–218, November 2011.

[15] X. Domingo-Almenara, A. Perera, N. Ramirez, N. Canellas, X. Correig, J. Brezmes. Compound identification in gas chromatography/mass spectrometry-based metabolomics by blind source separation. *Journal of Chromatography A*, 28; 1409:226–33, 2015

[16] S. Dhakshinamoorthy, N. Dinh, J. Skolnick, M.P. Styczynski. Metabolomics identifies the intersection of phosphoethanolamine with menaquinone-triggered apoptosis in an in vitro model of leukemia. *Molecular Biosystems*, 11 (2015): 2406–2416

[17] Y. Ni, Y. Qiu, W. Jiang, K. Suttlemyre, M. Su, W. Zhang, W. Jia, X. Du, ADAP- GC 2.0: deconvolution of coeluting metabolites from GC/TOF-MS data for metabolomics studies, *Anal. Chem*, 84 (15) (2012) 6619–6629

[18] Katty X. Wan, Ilan Vidavsky, and Michael L. Gross. Comparing similar spectra: from similarity index to spectral contrast angle. *Journal of the American Society for Mass Spectrometry*, 13(1):85–88, January 2002.

[19] Jan Hummel, Nadine Strehmel, Joachim Selbig, Dirk Walther, and Joachim Kopka. Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics*, 6(2):322–333, June 2010.

[20] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures *Anal. Chem*, 36 (8) (1964) 1627–1639,

[21] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *Radar and Signal Processing, IEE Proceedings F*, 140(6):362–370, December 1993.

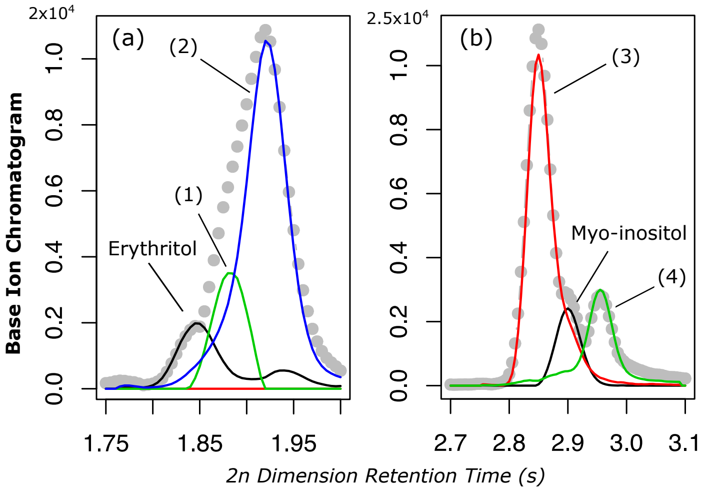

Figure 1: Two cases of co-elution resolved by ICA–OSD. The dotted grey line represents the BIC whereas the resolved profiles are shown in the solid-colored line. In (a), erythritol appear in co-elution with other unkown compounds (1, 2). In (b), myo-inositol appear also in co-eluted with an unkown compound (3, 4).

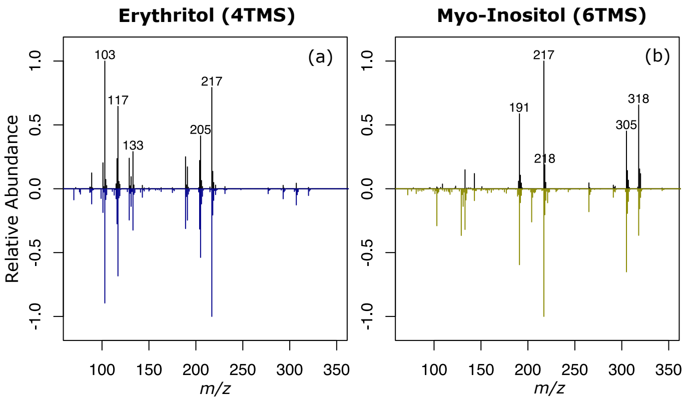

Figure 2: Representation of the extracted spectra (black) by ICA-OSD and the reference GMD spectra (color), for the cases shown in Figure 1, erythritol and myo-inositol. Reference spectra are shown negatively rotated in the same axis for a better visual appreciation.

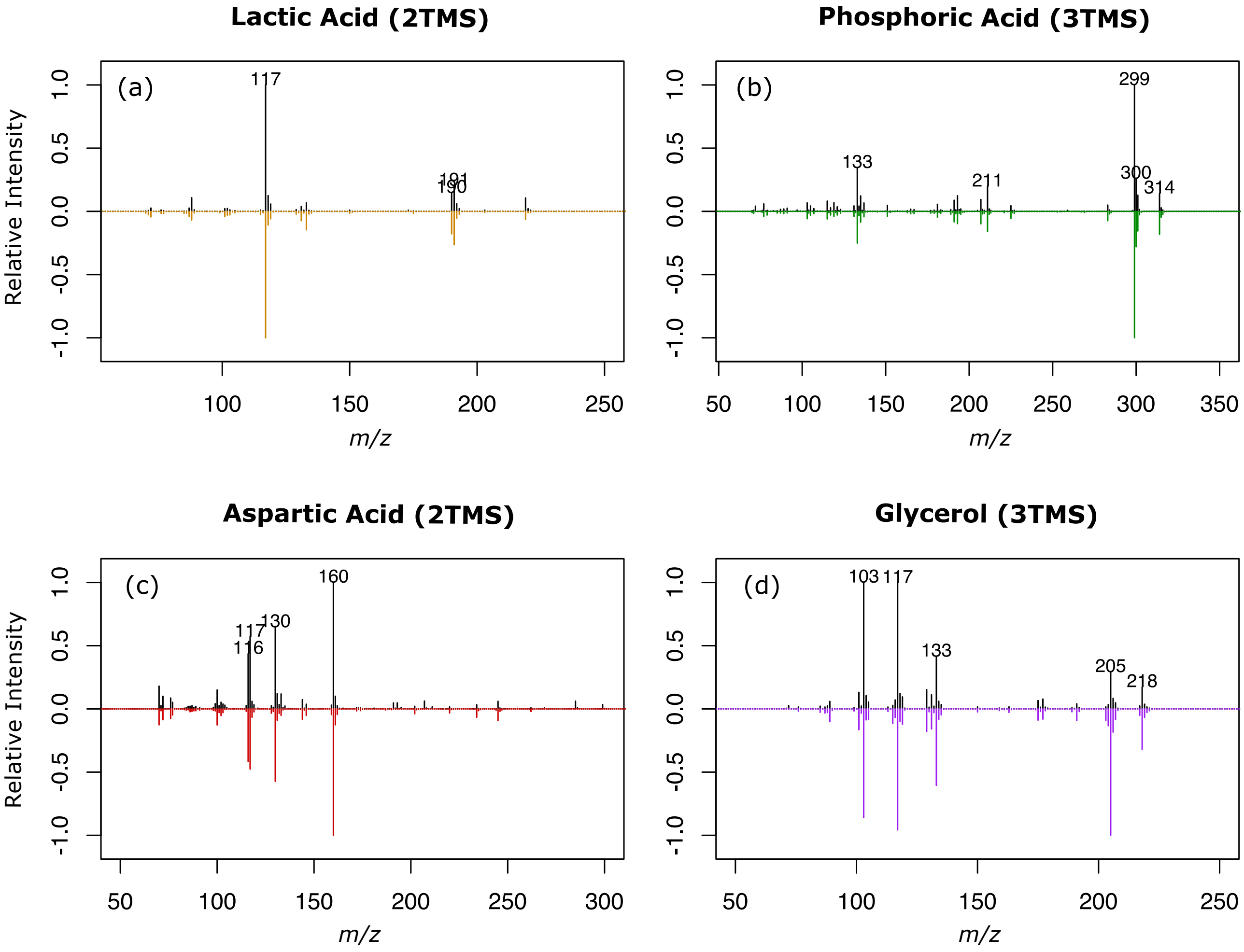

Figure 3: Representation of a set of extracted average - across samples - spectra (black) by ICA-OSD and the reference GMD spectra (color). Reference spectra are shown negatively rotated in the same axis for a better visual appreciation.

Table 1: List of identified compounds in Jurkat cell samples. MF is the match factor, R^2 is the linear regression coefficient, and Rel. C is the relative concentration.

No.	Rt1	Rt2	Name	MF	R^2	Rel. C (%)
1	4.5	1.8	Boric acid (3TMS)	92	0.96	137.28
2	4.58	2.9	Alanine (2TMS)	95	0.82	24.87
3	5.33	1.99	Valine (1TMS)	98	0.92	27.06
4	5.58	2.06	Lactic acid (2TMS)	99	0.99	939.18
5	5.75	2.12	Glycolic acid (2TMS)	98	0.90	61.17
6	5.92	1.94	Ethanolamine (3TMS)	87	0.84	16.26
7	6.5	1.9	Isovaleric acid, 2-oxo- (1MEOX) (1TMS) MP	89	0.98	101.7
8	6.67	2.38	Furan-2-carboxylic acid (1TMS)	98	1.00	25.02
9	7.5	2.78	Phosphoric acid (3TMS)	98	0.97	12.84
10	7.6	1.86	Glycerol (3TMS)	90	0.98	1294.11
11	8.1	2.38	Succinic acid (2TMS)	98	0.85	49.99
12	8.6	2.12	Nonanoic acid (1TMS)	91	0.98	105.88
13	9.1	2.04	Threonine, allo- (3TMS)	98	0.90	12.23
14	9.5	2.48	Aspartic acid (2TMS)	95	0.85	20.99
15	9.6	2.06	Malic acid (3TMS)	72	0.99	13.83
16	9.8	2.11	Decanoic acid (1TMS)	96	1.00	11.63
17	10.6	1.86	Erythritol (4TMS)	97	0.99	56.44
18	11.4	2.48	Proline [+CO2] (2TMS)	99	0.98	7.99
19	11.6	2.54	Hypotaurine (3TMS)	97	0.98	74.63
20	11.8	2.26	Glutamic acid (3TMS)	98	0.99	93.16
21	12.23	3.79	Pyroglutamic acid (2TMS)	99	0.89	112.16
22	12.23	3.05	Proline, 4-hydroxy-, cis- (3TMS)	98	0.82	14.45
23	12.82	4.28	Glutamic acid (2TMS)	97	0.97	17.37
24	13.23	3.26	Glutamic acid (3TMS)	98	0.99	80.1
25	13.48	3.01	Dodecanoic acid (1TMS)	98	0.94	25.84
26	13.9	3.65	Pyrophosphate (4TMS)	96	0.99	5.25
27	14.23	3.94	Glucose, 2-amino-2-deoxy- (4TMS) MP	91	0.99	8.38
28	14.57	2.89	Xylitol (5TMS)	98	0.92	24.63
29	14.98	3.41	Glycerol-3-phosphate (4TMS)	98	0.92	93.59
30	15.4	3	Ornithine (4TMS)	97	1.00	3.95
31	15.57	3.02	Tetradecanoic acid (1TMS)	98	0.97	154.25
32	16.07	3.25	Tyrosine (2TMS)	99	0.84	3.84
33	16.15	2.85	Psicose (1MEOX) (5TMS) BP	99	0.96	270.47
34	16.4	2.85	Glucose (1MEOX) (5TMS) MP	97	1.00	149.68
35	16.48	2.83	Mannose (1MEOX) (5TMS) MP	98	1.00	66.15
36	17.65	2.9	Inositol, allo- (6TMS)	94	0.95	19.81
37	18.98	2.98	Octadecenoic acid, 9-(Z)- (1TMS)	91	0.89	30.99
38	22.9	2.8	Sucrose (8TMS)	94	1.00	3.43