Journal of Chromatography A, 1492 (2017) 98–105

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Background correction in separation techniques hyphenated to high-resolution mass spectrometry – Thorough correction with mass spectrometry scans recorded as profile spectra☆

Guillaume L. Erny[a,*], Tanize Acunha[b,c], Carolina Simó[b], Alejandro Cifuentes[b], Arminda Alves[a]

[a] LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[b] Laboratory of Foodomics, CIAL, CSIC, Nicolas Cabrera 9, 28049 Madrid, Spain
[c] CAPES Foundation, Ministry of Education of Brazil, 70040-020 Brasília, DF, Brazil

ARTICLE INFO

Article history:
Received 26 October 2016
Received in revised form 31 January 2017
Accepted 23 February 2017
Available online 1 March 2017

Keywords:
Chemometrics
Hyphenated techniques
Data mining
Profile spectrum
Baseline correction

ABSTRACT

Separation techniques hyphenated with high-resolution mass spectrometry have been a true revolution in analytical separation techniques. Such instruments not only provide unmatched resolution, but they also allow measuring the peaks accurate masses that permit identifying monoisotopic formulae. However, data files can be large, with a major contribution from background noise and background ions. Such unnecessary contribution to the overall signal can hide important features as well as decrease the accuracy of the centroid determination, especially with minor features. Thus, noise and baseline correction can be a valuable pre-processing step. The methodology that is described here, unlike any other approach, is used to correct the original dataset with the MS scans recorded as profiles spectrum. Using urine metabolic studies as examples, we demonstrate that this thorough correction reduces the data complexity by more than 90%. Such correction not only permits an improved visualisation of secondary peaks in the chromatographic domain, but it also facilitates the complete assignment of each MS scan which is invaluable to detect possible comigration/coeluting species.

1. Introduction

In separation techniques, the baseline should ideally be a constant shift of the signal intensity. In practice, however, especially with complex samples, various artefacts (e.g. drift, large featureless objects, dips, ghost peaks) can be visible. While some of those artefacts are related to instrumental problems and can be corrected, others are inherent to the matrix, the sample preparation (derivatization, digestion, dilution, etc.) or due to additives needed in the mobile phase or background electrolyte. Modern software is usually able to detect and measure the relevant peaks' figures of merit despite complicated background signals. In the most extreme cases, the method can be optimised to avoid baseline artefacts that are interfering with the part of the analytical signal that is of interest. However, even the simplest background signal can be challenging when using Parallel Factor Analysis (PARAFAC) and other multi-ways related chemometric tools [1–5]. Background signals are also problematic in hyphenated and two-dimensional separation techniques. For example, Pierce and colleagues demonstrated that baseline drift does not only hide important features, it also impairs the accurate determination of peaks areas and centres [6,7]. Similarly, Zhang et al. demonstrated the interest of intensive background subtraction for high-resolution LC/MS data to obtain clean product ion spectra [8].

Due to the growing importance of the above-cited techniques, many innovative algorithms for background correction have been proposed in the past decade [1,9–24]. While it is not the aim of this manuscript to compare those methods [21], two of those algorithms are particularly interesting because of their fast computing speed and their ability to model any baselines. Those were derived from the work of Eilers [18,25] and are the adaptive iteratively reweighted penalised least squares (airPLS) and the asymmetrically

☆ Selected paper from the 40th International Symposium on Capillary Chromatography and 13th GC×GC Symposium (RIVA 2016), 29 May–3 June 2016, Riva del Garda, Italy.

* Corresponding author.
E-mail address: guillaume@fe.up.pt (G.L. Erny).

http://dx.doi.org/10.1016/j.chroma.2017.02.052
0021-9673/

reweighted penalised least squares (arPLS) algorithms [15,19]. The aim of this work is to demonstrate the facility and rapidity in which a dataset obtained using separation techniques hyphenated with high-resolution mass spectrometry (CE-TOF MS and UHPLC-Q/TOF MS) can be thoroughly corrected using a modified arPLS algorithm. The key innovative concept in this work is to use MS scans recorded as profile spectrum rather than the usual centroid spectrum.

While MS instruments record scans as continuous profiles (known as profile spectrum), it is usual to transform each scan into a MS centroid spectrum, the prevalent bars type representation. To this effect, every peak within every MS scan is detected using a peak picking algorithm. For each of the detected peaks, only two figures of merits are measured and recorded: the peak centre and the peak maximum intensity. Such conversion allows reducing the size of the file by at least one order of magnitude. The peak centre is also an accurate measurement of the peak position in the m/z axis (hence the name accurate mass [26]), allowing higher selectivity [27] and better identification capability, this by equaling the accurate mass with putative isotopic formula [28,29]. Centroided data are typically calculated by the acquisition software and are often (if not always) the starting points in chemometric tools. However, Kaufmann and others [28,30,31] unambiguously proved that with complex samples, centroid values are not always exact. Most of the errors are due to poorly separated peaks in the MS profile that become merged as a single centroided object. Moreover, when using centroided data, ion profiles (extracted ion profile or pure ion profile) need to be reconstituted before performing the background correction. Various algorithms exist for extracting ion profiles [32–34] from complex datasets with MS scans recorded as centroid spectra. However, with those approaches, some of the data will be lost or merged in the process. On the other hand, working with MS scans recorded as profile spectra allows bypassing those issues. In this case, the m/z axis is the same for every scan and ion profiles are simply obtained by recording, in every scan, the intensity at a given m/z ordinate. In this manuscript, ion profiles refer to chromatographic-like profiles retrieved by selecting in each MS scan the intensity at a given m/z coordinate. They should not be mistaken with extracted ion profiles, single ion profiles or pure ion profiles constructed from MS scans that were converted to centroid spectrum format.

In this work, two metabolomics urine datasets will be used to highlight the interest of this approach. One dataset was obtained using CE-TOF MS instrument, the other with a UHPLC-Q/TOF MS instrument. All algorithms used in this work have been programmed using Matlab and a personal computer. The files are available at https://github.com/glerny/finneeFinnee2016/ as part of the Finnee toolbox. Additional information and tutorials will also be posted at https://finneeblog.wordpress.com/ as well as in the wiki associated with finnee2016 (https://github.com/glerny/Finnee2016/wiki).

2. Materials and methods

2.1. Programming

The Finnee toolbox, used for this work, was run under Matlab 2016b using a PC (Operating system: Windows 10 Pro 64-bit; CPU Intel Core i5 2400 @ 3.10 GHz; RAM: 8.00 GB Dual-Channel DDR3). The new functions programmed for this work are freely available in the Finnee repository https://github.com/glerny/finnee2016/. The Finnee toolbox uses the mzML file format. In this work those were obtained from the original data using either CompassXport (Bruker) or msConvert (freely available at http://proteowizard.sourceforge.net/).

2.2. CE-TOF MS analysis

CE-TOF MS analysis were carried out using a P/ACE 5010 CE system (Beckman, Fullerton, CA, USA) coupled to a time-of-flight (TOF) instrument "microTOF" from Bruker Daltonics (Bremen, Germany). The coupling was carried out via an ESI interface model G1607A from Agilent Technologies. Electrical contact at the electrospray needle tip was established via a sheath liquid composed of 2-propanol-water (50:50, v/v) delivered at a flow rate of 4 μL/min by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL, USA). Bare fused-silica capillary with 50 μm i.d. and 85 cm of total length was from Composite Metal Services (Worcester, England). The inner capillary wall was coated with a cationic TEDETAMA-co-HPMA copolymer. CE separation was performed at −20 kV in an acidic BGE (1 M formic acid adjusted to pH 2.4 with ammonium hydroxide). The nebuliser and drying gas conditions were 0.4 bar N_2 and 4 L/min N_2, respectively, maintaining the ESI chamber temperature at 200 °C. TOF MS was operated in the negative ion mode (capillary voltage was 4 kV) and spectra were acquired in the range of 50–700 m/z. External and internal calibration of the TOF MS instrument was performed by introducing a 5 mM sodium formate solution through the separation capillary. Masses for the calibration of the TOF MS were next: 180.9731, 248.9594, 316.9479, 384.9353, 452.9227 and 520.9102 m/z. The mass resolving power, RP(FWHM), was calculated at m/z = 243 equal to 8500. The numerical threshold was set to zero. The original data files were converted to mzML format using CompassXport. The mzML file is freely available [35].

2.3. UHPLC-Q/TOF MS analysis

UHPLC-Q/TOF analysis was carried out using a 1290 system (Agilent) coupled to a quadrupole-time-of-flight (Q/TOF) 6540 (Agilent) equipped with an orthogonal electrospray ionisation (ESI) source (Agilent Jet Stream, AJS). The separation was performed on a Zorbax Eclipse Plus C8 (2.1 × 100 mm, 1.8 μm) column using phase A (water with 0.1% (v/v) formic acid) and phase B (acetonitrile with 0.1% (v/v) formic acid) and following gradient program: the run was started at 0% B and maintained for 2 min. From 2–6 min phase, B increased linearly till 30%. From 6–8 min phase B was increased to 100% and maintained for 2 min (from 8 to 10). Before each run, the column was re-equilibrated for 3 min using the initial solvent composition. TOF-MS operation parameters were the following: capillary voltage, −4000 V; nebuliser pressure, 25 psi; drying gas flow rate, 7 L/min; gas temperature, 300 °C; skimmer voltage, 45 V; fragmentor voltage was 125 V; 50–1000 m/z mass scan in positive ionisation mode. External calibration of the TOF MS was carried out using a commercial mixture from Agilent with next m/z values: 118.0863, 322.0481, 622.0290, 1221.9906, 1521.9715, 1821.9523, 2121.9331, 2421.9139 and 2721.8948. The mass resolving power, RP(FWHM), was calculated at m/z = 265 equal to 16000. The numerical threshold was set to 200. The original data files were converted to mzML format using msConvert. The mzML file is freely available [36].

2.4. Chemicals and samples

All chemical used were of analytical reagent grade. Reagent and solvents employed in the preparation of the CE electrolytes, sheath liquid, and LC mobile phases were of MS grade: isopropanol, formic acid and ammonium hydroxide from Sigma-Aldrich (St. Louis, MO, USA) and acetonitrile from Labscan (Gliwice, Poland). The water was purified in a Milli-Q system from Millipore (Bedford, MA, USA). Urine sampleswere directly injected following simple filtration by using a 0.22 μm pore size regenerated cellulose filter.

3. Results and discussion

3.1. Selection of profiles

The first step in this work was to select the ion profiles that needed baseline subtraction. The initial motivation for such a task was to decrease the number of ion profiles to be corrected. For example, with the CE-TOF MS dataset, a MS scan is a 2 by 60512 array (m/z scan from 50 to 700, with an interval of 0.0044 at 50 m/z and 0.0167 at 700 m/z) with m/z values and intensities in the first and second column, respectively. The full dataset consists of 1265 scans spaced at relatively regular intervals, ion profiles are obtained by registering the intensity at given m/z values. Thus 60512 profiles will be collected, each containing 1265 points. Similarly, with the UHPLC-Q/TOF MS 141310 profiles (m/z scan from 50 to 1000, with an interval of 0.0025 at 50 and 0.0110 at 1000), each containing 3581 points, will have to be corrected. While the arPLS and airPLS algorithms are particularly fast to converge, it remains wise trying decreasing the number of ion profiles to correct.

While every ion profile is of the same length, with MS instruments, a zero value indicates that no ion was detected at this time and this m/z interval. The relative amount of zero values within a given profile is a strong indicator of the presence of background ions. Indeed, in the absence of background ions, the intensities in the profile will be different of zero only when an analyte entered the ion chamber and generated an ion whose m/z peak will be within the mass interval of interest. With high-resolution mass analyser, in the absence of noise and background ions, the signal will contain many zero values. However, if background ions are present, ionic profiles will contain few zero values as, by definition, the background ions are present during most of the separation time. Fig. 1 illustrates the repartition of profiles as a function of their relative amount of non-zeros values. In this figure, the profile relative lengths are represented in abscissa (non-zeros divided by the total number of values; discrete value with a 1% increment), and the number of profiles with such relative length are represented in ordinate. A similar figure was obtained with the CE-TOF MS data (additional figures, especially related to the CE-TOF MS are presented as supplementary materials, SM1). This graph is characteristic of the distribution of profiles, with no profiles with less than 10%, then a spike of profiles between 10 and 15% and a majority of profiles between 30 and 60%. This representation can be used to filter the dataset, by as in Fig. 2, grouping profiles based on their relative length.

Fig. 2(A–E) shows the base peak profiles (BPP) calculated using profiles with relative length between (A) 0 and 15%, (B) 0 and 25%, (C) 25 and 50%, (D) 50 and 75% and (E) 75% and 100%. BPPs in Fig. 2A and B contain a minimal amount of information, showing that profiles with a relative length below 25% contain neither background ions nor analytes and cen be used to estimate the instrumental noise in the absence of peaks.To do so, all the intensities from profiles with a relatve length below 25% were concatenate into a single vector.The noise was measured using six times the standard deviation of all values. The instrumental noise was equal to 50 a.u.for the UHPLC-Q/TOF MS and 20 a.u. for the CE-TOF MS. As was previously discussed, the BPP depicted in Fig. 2E presents a high baseline level, but also a strong analytical signal. Profiles used to construct this representation seem the most relevant for baseline correction. However, for a thorough correction of the dataset, profiles that contain more than 50% of non-zeros values were also included. Thus, with the UHPLC-Q/TOF MS dataset, 37,000 profiles are selected (for a total of 141310) for baseline correction. With the CE-TOF MS dataset, 3909 profiles were selected (for a total of 60512). It should be emphasised that while it may seem attractive to correct all profiles, most baseline correction algorithms will not perform well with profiles that only contain peaks without any baseline points.

3.2. Baseline correction algorithms

Basic requirements for the background correction algorithm are good robustness, high response speed, flexible; and parameters should be the same for every ion profile within a single dataset. While most existing algorithms were tested in this work, the airPLS and arPLS were of particular interest due to their easiness, good performance with various baseline drift profiles, and rapidity. Fig. 3A and C show the baseline obtained with both algorithms with profiles representative of the most complicated baselines found in the two datasets. This figure clearly indicates that the airPLS algorithm is not adequate when negative peaks, such artefacts (very common in CE analysis), are present (Fig. 3C). Moreover, the position of the baseline about the background noise depends on the

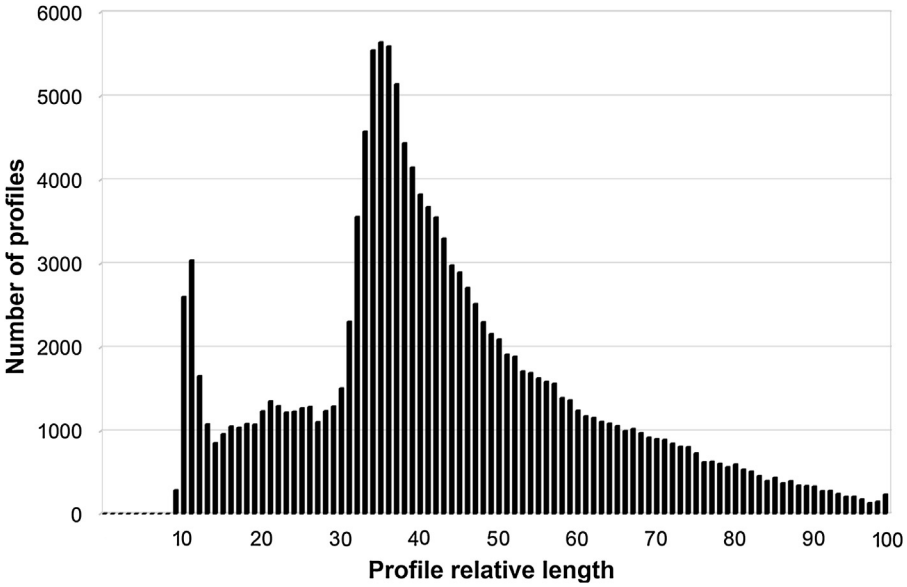

Fig. 1. Amount of profiles within the UHPLC-Q/TOF MS dataset with a given relative length, where the relative length is, for a given profile, the number of nonnull values divide by the total number of values.

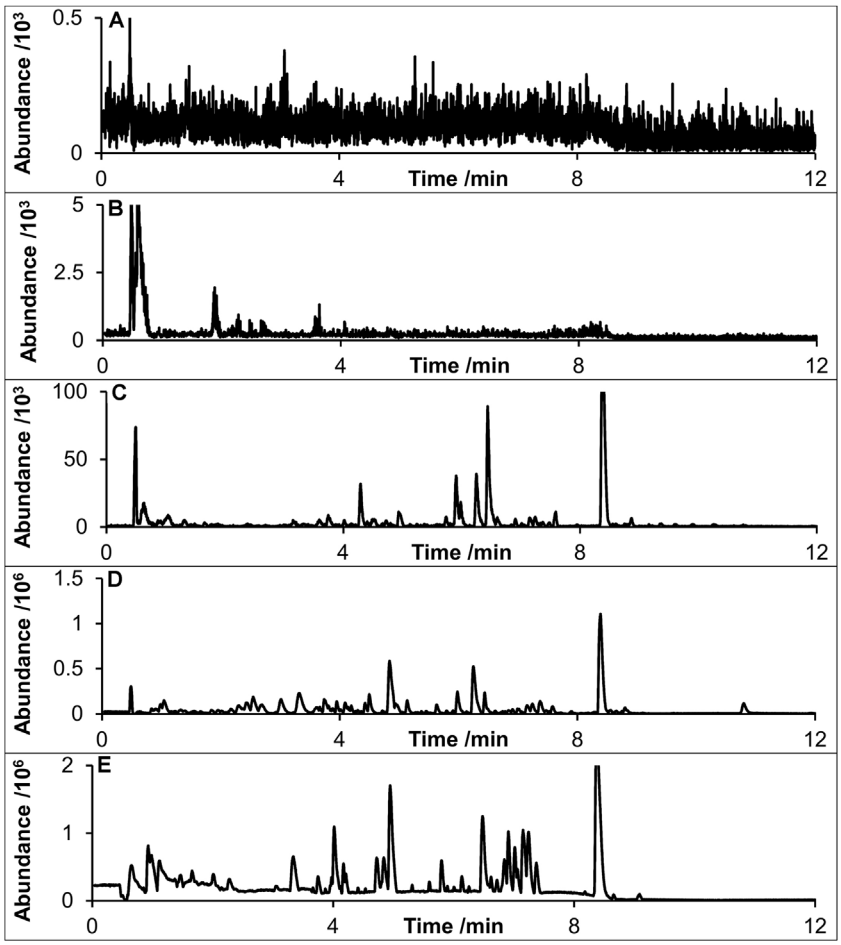

Fig. 2. Base peak profiles using selected ion profiles filtered based on the relative length. BPP calculated with profiles with relative lengths between (A) 0 and 15%, (B) 0 and 25%, (C) 25 and 50%, (D) 50 and 75% and (E) 75 and 100%.

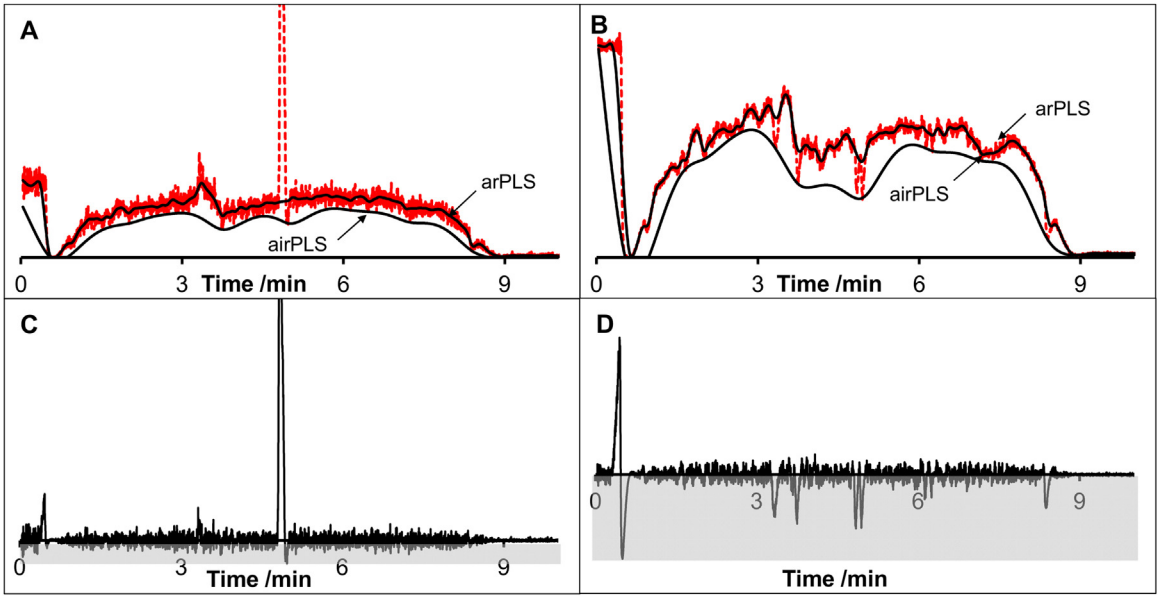

Fig. 3. (A) and (B) selected ion profiles and baselines calculated using the arPLS and airPLS algorithms, (C) and (D) corrected profiles using the arPLS algorithms.

asymmetry parameters. Such a parameter will have to be optimised for every profile depending on the signal to noise ratio. The arPLS appears better with profiles found in the two datasets. Not only this method works correctly with negative peaks, but the smoothness, the only necessary parameter, can be optimised once for all profiles within the dataset. The main limitation of the airPLS algorithm is that negative values can be under evaluated and will depend on the asymmetry factor. While the airPLS algorithm is suitable for convo-

luted profiles with many sharp peaks, the arPLS is clearly superior for profiles that consist of few peaks. Thus the arPLS algorithm was used to correct the selected ion profiles. It should be noted that the arPLS algorithm utilised in this work is slightly different from the original version. The code is introduced in supplementary material (SM2).

Figs. 3C and D show the baseline corrected profiles with the arPLS algorithm; all negative values (zone in grey) are arbitrary set to zero. Those figures indicate that the arPLS can correct baselines with very erratic fluctuations. Nevertheless, in cases of sharp variations of the baseline intensity a ghost peak may appear in the corrected profile (e.g. at 0.5 min in Fig. 3D). The intensity of this peak can be decreased using a lower smoothness value. However, an improperly low value will also modify the intensity of the analytical signal.

3.3. Thorough correction

Fig. 4 is a schematic representation of the algorithm utilised for the thorough baseline correction of the original MS dataset. The steps are as follow:

i Selection of profiles using the percentage of non-zeros values.
ii Estimation of the baseline for each profile using the arPLS algorithm. Each profile is corrected, and negative values are set to zeros.
iii MS scans are corrected by replacing the intensities in the *m*/*z* coordinates of the selected profiles.

Fig. 5 compares the total ion profile (TIP) calculated with (A) the original dataset and (B) the dataset where MS scans were baseline corrected. While the baseline level decreases after baseline correction, the level is still higher than zero. Rather than due to uncorrected (or poorly corrected) baselines this shift is believed to be due to the contribution from noise from every channel. Because individual scans are now corrected from background ions, a very simple noise removal strategy can be designed. Noise removal was done as follow: first, the intensity of the noise in the absence of peaks is estimated. Then, a moving window of size m*n is used to

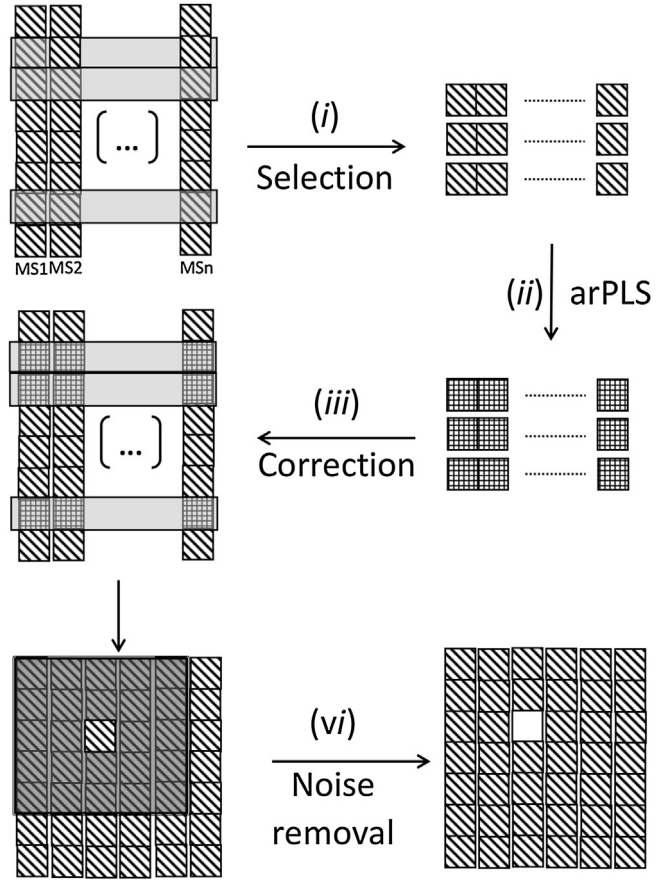

Fig. 4. Schematic representation of the baseline correction methodology. (*i*) Selection of the profiles that need to be corrected from baseline drift; (*ii*) correction using the arPLS algorithm; (*iii*) the selected profiles are replaced by the corrected profiles in the dataset: (*iv*) noise removal: for each point if the intensities within a set window are below the noise threshold, the intensity of the point of interest is set to zero.

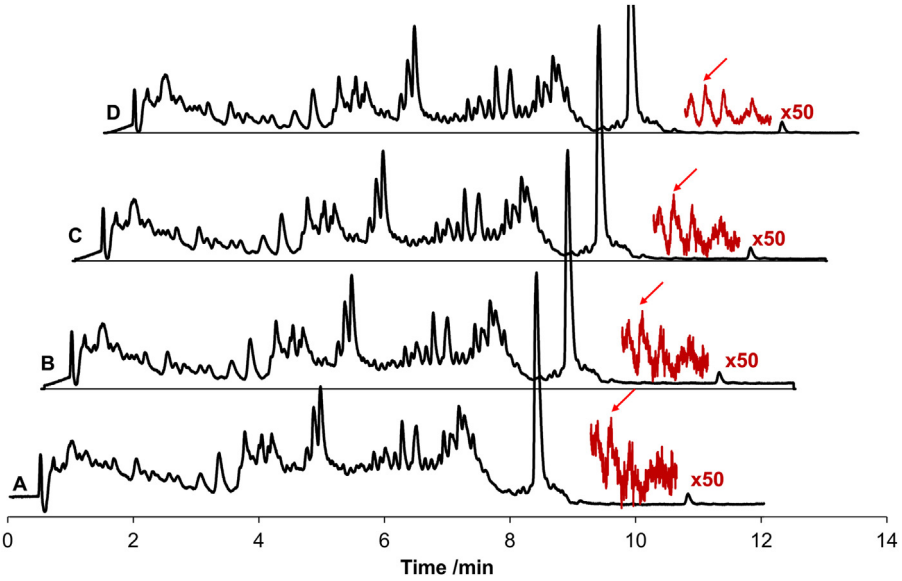

Fig. 5. Total ion profiles from the UHPLC-Q/TOF MS dataset with (A) original data, (B) corrected profiles, (C) corrected profiles and noise reduction using a noise value of 50 and a 7*5 moving window and (D) corrected profiles and noise reduction using a noise value of 200 and a 11*7 moving window. Traces in light grey are a 50-time enhancement of the signal between 9 and 10.5 min; the arrows indicate the peak of interest.

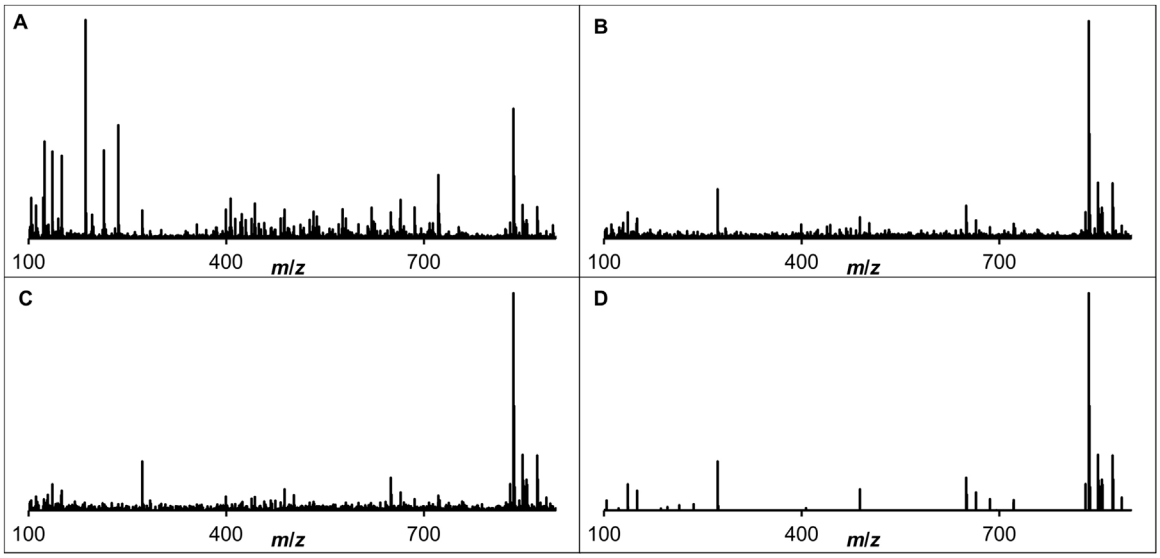

Fig. 6. Full MS scans (profile spectrum) recorded at 9.69 min with (A) original data, (B) corrected profiles, (C) corrected profiles and noise reduction using a noise value of 50 and (D) corrected profiles and noise reduction using a noise value of 200.

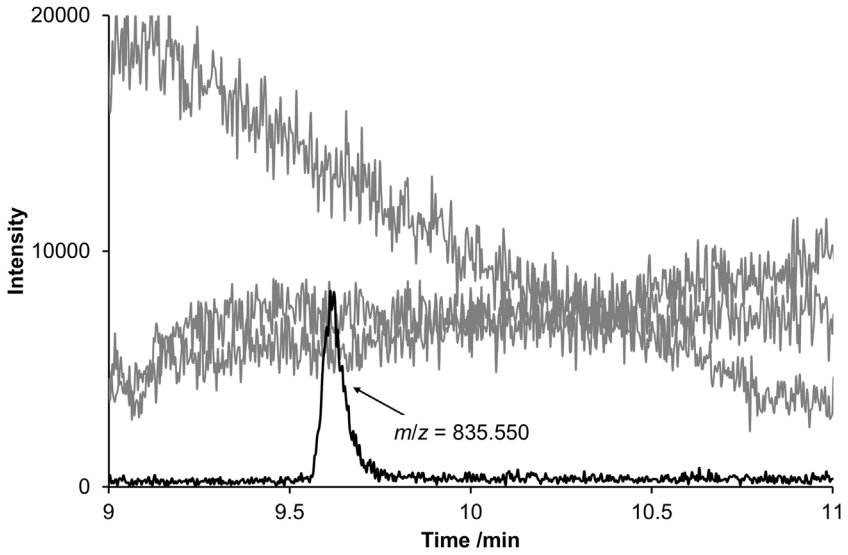

Fig. 7. Ion profiles from the original dataset selected based on the four most intense m/z peak in Fig. 6A. The ion profile obtained at $m/z = 835.550$ corresponds to the base peak after baseline correction (Fig. 6B–D).

measure the maximum intensity within this window. This window is applied to every data point and if the maximum window's intensity is lower than three times the estimated noise, the intensity of the investigated point is set to zero. The TIP, after noise correction, obtained with threshold noise values of 50 and 200 can be observed in Fig. 5C and D, respectively. The noise threshold of 50 is the instrumental noise as previously estimated (see 3.1). As it can be seen in Fig. 5C, while there is a slight improvement in comparison to Fig. 5C, the noise remains high. This is not the case when using a noise threshold of 200, as it can be observed in Fig. 5D. The combination of baseline and noise removal allowed correction from baseline fluctuation. As visible in the close-up (red traces in Fig. 5), this allows detecting easily minor peaks. The noise removal allowed to reduce the size of the data (i.e. value set to zero) by more than 98% with the UHPLC-Q/TOF MS (noise value 200; window size of 11*7). While removing unnecessary information is clearly attractive, it is evident that the data should be minimally modified for such approach to be valuable. The effect of baseline correction and noise removal algorithm on the MS scan is investigated in Fig. 6. The MS scans at 9.69 min (A, indicated by the arrows in Fig. 5) is compared to (B) the scan after baseline correction but without noise removal, (C) baseline and noise corrected with a noise parameter of 50 (windows of 7*5) and (D) baseline and noise corrected with a noise parameter of 200 (window of 11*7).

The baseline correction allows a drastic simplification of the MS scans. Usefulness of such data reduction methodology is exemplified in Fig. 7 where the ion profiles calculated based on the four most intense MS peaks in Fig. 6A are superposed. As it can be seen three of them are related to background ions, while one exhibits a peak at 9.69 min. The m/z index of this profile corresponds to the base peak after baseline correction in Fig. 6B–D. In this example, and as demonstrated in Fig. 7, the corrected scans allow to easily and reliably obtaining the base peak ions that were previously hidden by background ions. While not as striking, this is also true for

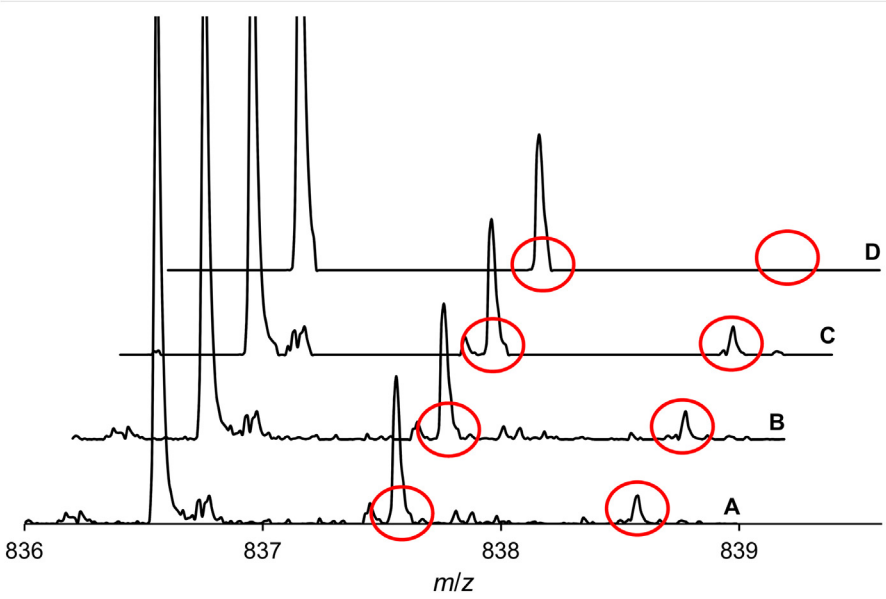

Fig. 8. Zoom between *m*/*z* 836 and 840 of the MS scans presented in Fig. 5. (A) original data, (B) corrected profiles, (C) corrected profiles and noise reduction using a noise value of 50 and (D) corrected profiles and noise reduction using a noise value of 200.

Table 1
Computing speed depending on the number of profiles to correct and dataset size.

CE-TOF MS		UHPLC-Q/TOFMS	
File size (mzML file)	650,000 kB	File size (mzML file)	2,900,000 kB
Selection of profiles	<1 min	Selection of profiles	5 min
Correction (1830 prof.)	<1 min	Correction (9000 prof.)	6 min
Correction (7145 prof.)	1.5 min	Correction (37000 prof.)	15 min
Data reduction by	99.4%	Data reduction by	98.0%

any scan, even with high-intensity peaks. In this case, baseline and noise removal facilitate a thorough analysis of the scans allowing to detect possible co-migration/co-elution.

In Fig. 8, the effects of the baseline and noise removal algorithms on the peak shapes were also investigated. In this figure, the shapes of the three first isotopic peaks were studied under different data manipulation. The filter is very conservative; the peak shapes barely change despite the removal of many data of low intensity. However, it is evident, that if using high values for the noise parameter, the peaks of lowest intensities may be filtered out as it is the case in Fig. 7D. Nevertheless, this may still be beneficial as it allows to reduce the data size by more the 98%, which not only allows much clearer TIP but should also be useful for subsequent chemometrics routine, including for centroid algorithms.

3.4. Computing speed

Computing speed is a crucial factor, however, because speed can vary a lot depending on the computer and complexity of the dataset, the speeds given in Table 1 are only indicative. Time has been measured using a personal computer and using the Matlab tic/toc functions that measure the CPU time. With a maximum time of 15 min, it demonstrates that manipulating large dataset can be done in reasonable times. The total computing speed varied from less than 2 min with the small CE-TOF MS dataset (0.7 GB) to 20 min with the larger dataset (2.9 GB), this excluding the conversion of the mzML file to the Matlab structure as previously described. Computing speed will also depend on the number or profiles to correct. While it is tempting to minimise this speed by reducing the number of profiles to correct, a thorough correction is clearly advantageous but will require overestimating the number of profiles to correct.

4. Conclusions

In this manuscript, the arPLS, a baseline correction algorithms and a new noise removal algorithm have been applied to thoroughly correct urine metabolome datasets. The baseline and noise correction allows a drastic reduction of the data complexity by more than 98%. Such data transformation allows easily detecting minor peaks that were previously hidden. The arPLS has been found to be reliable, robust and fast allowing to correct those large datasets in a reasonable time, e.g. less than 20 min for the larger dataset. The use of profile spectra instead of centroid spectra allows to easily obtain ion profiles. However, it is evident that centroid algorithms will then have to be performed in the corrected data set to calculate the accurate masses.

Acknowledgments

This work was financially supported by the projects: (i) POCI-01-0145-FEDER-006939 (Laboratory for Process Engineering, Environment, Biotechnology and Energy – UID/EQU/00511/2013) funded by the European Regional Development Fund (ERDF), through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds, through FCT – Fundaça˜o para a Ciência e a Tecnologia; (ii) NORTE-01-0145-FEDER-000005LEPABE-2-ECO-INNOVATION, LEPABE-2-*E*CO-INNOVATION, supported by North Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF); (iii) AGL2014-53609-P project (Ministerio de Educación y Ciencia, Spain); (iv) IF/00528/2013, supported by FEDER funds through the Operational Programme for Human Potential and by National Funds through FCT – Foundation for Science and Technology. T. A. thanks the CAPES Foundation, Ministry of Education of Brazil for her pre-doctoral Scholarship – Proc. N° 1532/13-8.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2017.02.052.

References

[1] H.-Y. Fu, H.-D. Li, Y.-J. Yu, B. Wang, P. Lu, H.-P. Cui, et al., Simple automatic strategy for background drift correction in chromatographic data analysis, J. Chromatogr. A 2016 (1449) 89–99, http://dx.doi.org/10.1016/j.chroma.2016.04.054.

[2] J.W. McIlroy, R.W. Smith, V.L. McGuffin, Assessing the effect of data pretreatment procedures for principal components analysis of chromatographic data, Forensic Sci. Int. 257 (2015) 1–12, http://dx.doi.org/10.1016/j.forsciint.2015.07.038.

[3] J. Kuligowski, G. Quintás, R. Tauler, B. Lendl, M. De La Guardia, Background correction and multivariate curve resolution of online liquid chromatography with infrared spectrometric detection, Anal. Chem. 83 (2011) 4855–4862, http://dx.doi.org/10.1021/ac2004407.

[4] M. Jalali-Heravi, H. Parastar, Recent trends in application of multivariate curve resolution approaches for improving gas chromatography-mass spectrometry analysis of essential oils, Talanta 85 (2011) 835–849, http://dx.doi.org/10.1016/j.talanta.2011.05.045.

[5] R.B. Martí, J.F. Baldrich, Fundamentals of PARAFAC, in: A. Muñoz de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri (Eds.), Data Handl. Sci. Technol., first edition, Elsevier B.V., Amsterdam, Netherland, 2015, pp. 7–35, http://dx.doi.org/10.1016/B978-0-444-63527-3.00001-1.

[6] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, J. Chromatogr. A 1255 (2012) 3–11, http://dx.doi.org/10.1016/j.chroma.2012.05.050.

[7] R.C. Allen, M.G. John, S.C. Rutan, M.R. Filgueira, P.W. Carr, Effect of background correction on peak detection and quantification in online comprehensive two-dimensional liquid chromatography using diode array detection, J. Chromatogr. A 1254 (2012) 51–61, http://dx.doi.org/10.1016/j.chroma.2012.07.034.

[8] H. Zhang, M. Grubb, W. Wu, J. Josephs, W.G. Humphreys, Algorithm for thorough background subtraction of high-Resolution LC/MS data: application to obtain clean product ion spectra from nonselective collision-Induced dissociation experiments, Anal. Chem. 81 (2009) 2695–2700, http://dx.doi.org/10.1021/ac8027189.

[9] R. Jabeen, Automated Baseline Estimation for Analytical Signals, Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science, Dalhousie University, 2013, http://dx.doi.org/10.1016/S0022-3913(12)00047-9.

[10] X. Liu, Z. Zhang, P.F.M. Sousa, C. Chen, M. Ouyang, Y. Wei, et al., Selective iteratively reweighted quantile regression for baseline correction, Anal. Bioanal. Chem. 406 (2014) 1985–1998, http://dx.doi.org/10.1007/s00216-013-7610-x.

[11] Y. Liu, W. Cai, X. Shao, Intelligent background correction using an adaptive lifting wavelet, Chemom. Intell. Lab. Syst. 125 (2013) 11–17, http://dx.doi.org/10.1016/j.chemolab.2013.03.010.

[12] M. Lopatka, A. Barcaru, M.J. Sjerps, G. Vivó-Truyols, Leveraging probabilistic peak detection to estimate baseline drift in complex chromatographic samples, J. Chromatogr. A 1431 (2016) 122–130, http://dx.doi.org/10.1016/j.chroma.2015.12.063.

[13] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS), Chemom. Intell. Lab. Syst. 139 (2014) 156–167, http://dx.doi.org/10.1016/j.chemolab.2014.09.014.

[14] Z. Wang, M. Zhang, P.D.B. Harrington, Comparison of three algorithms for the baseline correction of hyphenated data objects, Anal. Chem. 86 (2014) 9050–9057, http://dx.doi.org/10.1021/ac501658k.

[15] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, Analyst 135 (2010) 1138–1146, http://dx.doi.org/10.1039/b922045c.

[16] Y.-J. Yu, Q.-L. Xia, S. Wang, B. Wang, F.-W. Xie, X.-B. Zhang, et al., Chemometric strategy for automatic chromatographic peak detection and background drift correction in chromatographic data, J. Chromatogr. A 1359 (2014) 262–270, http://dx.doi.org/10.1016/j.chroma.2014.07.053.

[17] X. Liu, Z. Zhang, Y. Liang, P.F.M. Sousa, Y. Yun, L. Yu, Baseline correction of high resolution spectral profile data based on exponential smoothing, Chemom. Intell. Lab. Syst. 139 (2014) 97–108, http://dx.doi.org/10.1016/j.chemolab.2014.09.018.

[18] P.H.C. Eilers, H.F.M. Boelens, Baseline correction with asymmetric least squares smoothing, in: Unpubl. Manuscr, 2005, pp. 1–26 (accessed September 26, 2016) https://www.researchgate.net/publication/228961729_Baseline_Correction_with_Asymmetric_Least_Squares_Smoothing.

[19] S.-J. Baek, A. Park, Y.-J. Ahn, J. Choo, Baseline correction using asymmetrically reweighted penalized least squares smoothing, Analyst 140 (2015) 250–257, http://dx.doi.org/10.1039/C4AN01061B.

[20] J. Kuligowski, G. Quintás, S. Garrigues, M. de la Guardia, New background correction approach based on polynomial regressions for on-line liquid chromatography-Fourier transform infrared spectrometry, J. Chromatogr. A 1216 (2009) 3122–3130, http://dx.doi.org/10.1016/j.chroma.2009.01.110.

[21] L. Komsta, Comparison of several methods of chromatographic baseline removal with a new approach based on quantile regression, Chromatographia 73 (2011) 721–731, http://dx.doi.org/10.1007/s10337-011-1962-1.

[22] J. Kuligowski, G. Quintás, S. Garrigues, M. De La Guardia, Application of point-to-point matching algorithms for background correction in on-line liquid chromatography-Fourier transform infrared spectrometry (LC-FTIR), Talanta 80 (2010) 1771–1776, http://dx.doi.org/10.1016/j.talanta.2009.10.021.

[23] G.L. Erny, V. Calisto, V.I. Esteves, Noise normalisation in capillary electrophoresis using a diode array detector, J. Sep. Sci. 34 (2011) 1703–1707, http://dx.doi.org/10.1002/jssc.201100243.

[24] Z. Xu, X. Sun, P.B. Harrington, Baseline correction method using an orthogonal basis for gas chromatography/mass spectrometry data, Anal. Chem. 83 (2011) 7464–7471, http://dx.doi.org/10.1021/ac2016745.

[25] P.H.C. Eilers, A perfect smoother, Anal. Chem. 75 (2003) 3631–3636, http://dx.doi.org/10.1021/ac034173t.

[26] K.K. Murray, R.K. Boyd, M.N. Eberlin, G.J. Langley, L. Li, Y. Naito, Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013)*, Pure Appl. Chem. 85 (2013) 1515–1609, http://dx.doi.org/10.1351/PAC-REC-06-04-06.

[27] L. Vergeynst, H. Van Langenhove, P. Joos, K. Demeestere, Accurate mass determination, quantification and determination of detection limits in liquid chromatography–high-resolution time-of-flight mass spectrometry: challenges and practical solutions, Anal. Chim. Acta 789 (2013) 74–82, http://dx.doi.org/10.1016/j.aca.2013.06.024.

[28] A. Kaufmann, The current role of high-resolution mass spectrometry in food analysis, Anal. Bioanal. Chem. 403 (2012) 1233–1249, http://dx.doi.org/10.1007/s00216-011-5629-4.

[29] I. Ojanperä, M. Kolmonen, A. Pelander, Current use of high-resolution mass spectrometry in drug screening relevant to clinical and forensic toxicology and doping control, Anal. Bioanal. Chem. (2012), http://dx.doi.org/10.1007/s00216-012-5726-z.

[30] A.W.T. Bristow, K.S. Webb, Intercomparison study on accurate mass measurement of small molecules in mass spectrometry, J. Am. Soc. Mass Spectrom. 14 (2003) 1086–1098, http://dx.doi.org/10.1016/S1044-0305(03)00403-3.

[31] A. Kaufmann, P. Butcher, Strategies to avoid false negative findings in residue analysis using liquid chromatography coupled to time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 20 (2006) 3566–3572, http://dx.doi.org/10.1002/rcm.2762.

[32] S.-Y.Y. Wang, C.-H.H. Kuo, Y.J. Tseng, Ion trace detection algorithm to extract pure ion chromatograms to improve untargeted peak detection quality for liquid chromatography/time-of-flight mass spectrometry-based metabolomics data, Anal. Chem. 87 (2015) 3048–3055, http://dx.doi.org/10.1021/ac504711d.

[33] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinf. 11 (2010) 395, http://dx.doi.org/10.1186/1471-2105-11-395.

[34] R. Tautenhahn, C. Bottcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinf. 9 (2008) 16, http://dx.doi.org/10.1186/1471-2105-9-504.

[35] G. Erny, CE-TOFMS data from a urine sample – mzML files with scans in profile or centroid spectrum format, Mendeley Data (2017) v2, http://dx.doi.org/10.17632/cb4hv9cp2c.2.

[36] G. Erny, UHPLC-QTOFMS data from a urine sample – mzML files with scans in profile or centroid spectrum format, Mendeley Data (2017) v1, http://dx.doi.org/10.17632/6rn82jdv8d.1.