**RESEARCH PAPER**

# Linking MS1 and MS2 signals in positive and negative modes of LC-HRMS in untargeted metabolomics using the ROIMCR approach

Flávia Yoshie Yamamoto[1,2] · Carlos Pérez-López[1] · Ana Lopez-Antia[1] · Silvia Lacorte[1] · Denis Moledo de Souza Abessa[2] · Romà Tauler[1]

**Abstract**

Data-independent acquisition (DIA) mode in liquid chromatography (LC) high-resolution mass spectrometry (HRMS) has emerged as a powerful strategy in untargeted metabolomics for detecting a broad range of metabolites. However, the use of this approach also represents a challenge in the analysis of the large datasets generated. The regions of interest (ROI) multivariate curve resolution (MCR) approach can help in the identification and characterization of unknown metabolites in their mixtures by linking their MS1 and MS2 DIA spectral signals. In this study, it is proposed for the first time the analysis of MS1 and MS2 DIA signals in positive and negative electrospray ionization modes simultaneously to increase the coverage of possible metabolites present in biological systems. In this work, this approach has been tested for the detection and identification of the amino acids present in a standard mixture solution and in fish embryo samples. The ROIMCR analysis allowed for the identification of all amino acids present in the analyzed mixtures in both positive and negative modes. The methodology allowed for the direct linking and correspondence between the MS signals in their different acquisition modes. Overall, this approach confirmed the advantages and possibilities of performing the proposed ROIMCR simultaneous analysis of mass spectrometry signals in their differing acquisition modes in untargeted metabolomics studies.

**Keywords** Untargeted metabolomics · All-ion fragmentation (AIF) · Positive and negative acquisition modes · Metabolite coverage improvement · ROIMCR, Regions of interest multivariate curve resolution

## Introduction

New advances in OMICs techniques have increased the comprehension of biological mechanisms as responses to multiple disturbing factors affecting the health of the organisms [1–3]. In this context, the extensive analysis of metabolites represents a powerful approach to unveil new responses of organisms in diverse research fields, such as environmental sciences and clinical studies [4, 5]

Liquid chromatography with high-resolution mass spectrometry (LC-HRMS), either with Orbitrap or time-of-flight (TOF) mass analyzers, represents one of the main choices for the analysis of metabolites in biological systems due to

its high sensitivity and specificity [6–8]. While a targeted approach allows the investigation of specific metabolites and potentially affected metabolic pathways, an untargeted analysis assesses a broader range of small molecules, identifying a greater number of affected metabolites and consequently exploring new affected pathways [9].

Regarding the ionization approach, electrospray ionization (ESI) allows the formation of positively or negatively charged ions depending on the polarity of the applied voltage. The particular physiochemical properties of each molecule may determine the most adequate ionization acquisition mode, with some presenting better ionization efficiency (IE) in the positive mode and others in the negative [10]. Most commonly, the acquisition is performed either in positive or negative ESI depending on the electronegativity, basicity, and size of the molecules to be analyzed [11], which may limit the identification of all constituents of biological systems if only one ionization mode is used.

The inherent large-sized complex dataset generated by the LC-HRMS, especially for data-independent acquisition (DIA) mode, is a great challenge for properly interpreting

✉ Romà Tauler
roma.tauler@idaea.csic.es

1 Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona, 18-26, 08034 Barcelona, Spain

2 Institute of Biosciences, São Paulo State University, São Vicente, Brazil

the most relevant changes of treated organisms under an experimental condition. Different computer workflows have been proposed for extracting the significant signal features and their chemical assignation in LC-HRMS datasets [12–16]. The regions of interest multivariate curve resolution (ROIMCR) chemometrics approach is based on the bilinear model intrinsic data structure of the LC–MS data [7, 17–20] and represents an effective strategy for compressing and resolving data without losing relevant information and maintaining instrumental mass accuracy. Firstly, to reduce the dimensionality of the data matrices, the "regions of interest" (ROI) approach considers some parameters, such as signal intensity threshold, mass error tolerance, and a minimum number of detected occurrences [17, 21]. This initial "filtering" of the data selects the most relevant signals (features) and eliminates possible interferences and background noise generated by the equipment and solvents used. Following data compression, the resolution of the elution and spectra profiles of the chemical constituents in the analyzed mixtures can be achieved with the multivariate curve resolution alternating least squares (MCR-ALS) method, as it has been extensively reported [22–25].

Within the untargeted metabolomics approach, our research group has recently shown for the first time the possible simultaneous analysis of MS1 and MS2 DIA raw data obtained by LC/Q-TOF by the ROIMCR approach for the identification of per- and poly-fluoroalkyl substances [20]. This new approach allows an improved linking of MS1 and MS2 DIA spectra of the resolved components by ROIMCR, i.e., linking the MS1 signals of the precursor ions with the MS2 signals of their fragment ions for every chemical compound of the analyzed sample. In this sense, a step forward in this approach is the simultaneous analysis of DIA acquisition in positive and negative modes, which has not yet been exploited. Therefore, in this work, we propose the simultaneous analysis of MS1 and MS2 DIA for both positive and negative modes, to increase the coverage of possible metabolites identified as well as to improve the efficiency of the data analysis procedures in untargeted metabolomics studies.

## Material and methods

First, one synthetic standard mixture of amino acids was prepared and analyzed to set up the ionization and acquisition parameters to obtain higher sensitivity and accuracy by the developed LC/Q-TOF method. Then, 7 fish embryo replicates (*Rhamdia quelen*) were analyzed to comprehensively identify the amino acids' signatures in the biological samples. For both situations, the simultaneous analysis of MS1 and MS2 DIA datasets for positive

and negative modes was attempted using the proposed ROIMCR approach.

## Amino acid analytical standards

The amino acid standard solution containing 17 amino acids and ammonium chloride was purchased from Sigma Aldrich® (AAS18). The concentrated original solution of 2.5 μM mL$^{-1}$ (except by L-cystine at 1.25 μM mL$^{-1}$) was diluted (5 × with 50:50 ACN/water) to reach a concentration of approximately 2 μg/mL (detailed information available in the Supplementary Material). Two internal standards, methionine sulfone (Sigma Aldrich®) and the 1,4-piperazinediethanesulfonic acid (PIPES, Sigma Aldrich®), were prepared at a concentration of 1 mg/mL and were spiked at a final concentration of 1 μg/mL.

## Metabolites' extraction in fish embryo samples

For the analysis of amino acids in biological samples, fish embryos (*Rhamdia quelen*) were obtained in the Panama Fish Farm (Paulo Lopes, Brazil) in March 2021 using the same procedures for fish breeding as described by Yamamoto et al. [26]. Amino acids and other metabolites were extracted from 7 pooled samples (approximately 50 mg of lyophilized embryos for each sample) according to the method proposed by Ortiz-Villanueva et al. [27] and are fully detailed in the Supporting Information.

## LC/Q-TOF analysis

Analysis was performed using a UPLC/Q-TOF (Bruker Impact II). An Acquity UPLC BEH HILIC column (2.1 × 50 mm, 1.7 μm) from Waters™ was used to analyze amino acids and other metabolites in the standard solution and fish embryo extracted samples according to the protocol proposed by WATERS [28] with few modifications (Table S1 in Supporting Information). Acquisition (FSR > 60,000) was performed in full scan mass over 60 to 1000 m/z (Supporting Information Table S2). The ionization source operated in positive (4 eV of ion energy, 20–30 eV switching normalized collision energy) and negative mode (6 eV of ion energy, 24–36 eV switching normalized collision energy) alternating MS scans of the precursor ions (MS1) and all-ion fragmentation (MS2). Acquisitions in positive and negative ESI were performed in different chromatographic runs, as Bruker's Q-TOF technology does not allow simultaneous acquisition in both modes. All solvents were analytical grade or higher purity.

## Data analysis

LC/Q-TOF data obtained in the analysis of the amino acid standards and fish embryo samples were independently processed by the ROIMCR chemometric method [17, 22, 24, 29], which consists of a first step (regions of interest, ROI) to select the most relevant MS instrumental signals reducing the dimensionality of the data [21] followed by the application of the multivariate curve resolution (MCR) method according to Jaumot et al. [18] and Tauler [30] to resolve the elution and mass spectra profiles of the chemical

constituents of the analyzed samples. In this work, MS1 and MS2 data acquired independently (DIA) were simultaneously analyzed by ROIMCR either separately in positive (+) and negative (-) ionization modes or simultaneously, in MS1(+), MS1(-), MS2(+), and MS2(-) modes. The procedure has the following steps (Fig. 1).

### Import of data

Firstly, the raw data obtained through the Bruker LC/Q-TOF were converted to a NetCDF data file (or mzXML files)
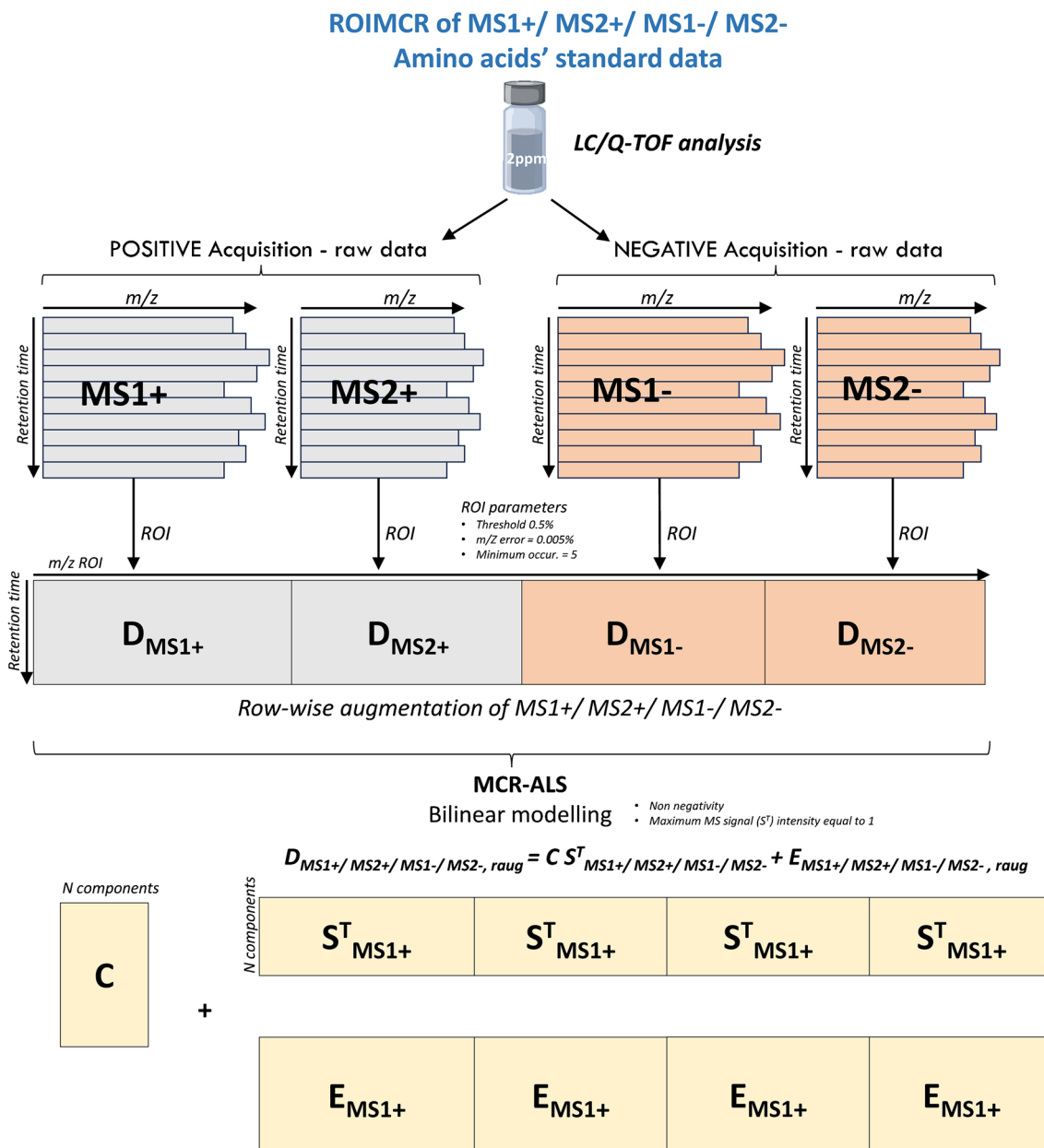


**Fig. 1** Row-wise matrix augmentation and MCR bilinear model factor decomposition of MS1(+), MS2(+), MS1(-), and MS2(-) data acquired in the analysis of the amino acid standard mixture

through the Compass DataAnalysis software 5.2 (Bruker, Germany) and imported into MATLAB R2020a computing and visualization environment. This could be performed either using the MSroi app as described by Pérez-Cova et al. [21] in MATLAB or via mass spectrometry directly with the functions of the Bioinformatics toolbox (The Mathworks, Inc., 2020b).

The data from the amino acids' standard were imported using the MSroi GUI app as a single chromatographic run (single sample option) while the data from the 7 replicate fish embryo samples were imported as multiple chromatographic runs (using the multi-sample option) arranged as a column-wise augmented data matrix (see Fig. 1).

### Regions of interest (ROI)

Before applying the ROI compression procedure to the NetCDF imported datasets, the MS1 and MS2 scans in positive (+) and negative (-) modes were separated, either for the analysis of the single sample (amino acid standards) or for the simultaneous analysis of the multiple fish embryo samples. After the data separation, ROI compression was performed separately for MS1(+), MS1(-), MS2(+), and MS2(-) datasets. Additionally, small parts of the chromatograms were cut by selecting only the time intervals of interest, disregarding the first 0.3 min (corresponding to the column dead volume) of the chromatographic run.

The ROI analysis generates a data matrix with a number of rows equal to the number of elution times and a number of columns equal to the number of MSROIs. This data matrix has the filtered MSROI data (following the parameters previously mentioned for each MS1 and MS2 dataset). The data matrices resulting from the ROI compression can be augmented row-wisely (horizontal concatenation) and be analyzed simultaneously in the two acquisition modes, the MS1 and the MS2 DIA data, at the same chromatographic retention times. This row-wise data matrix augmentation can be extended to include MS1 and MS2 DIA modes in both positive and negative ionization modes. An additional step for filtering those MSROIs with good elution profiles (through graphical visualization) was performed to improve the analysis before performing multivariate curve resolution, excluding those MSROIs with poor resolution, usually related to noise, also present in blank samples.

It is important to highlight that the row-wise matrix augmentation to analyze MS1 and MS2 DIA in positive and negative ionization modes requires that they are aligned to the same retention times. In other words, even if the acquisitions were performed in different chromatographic runs (as it occurred in the current study), the retention times between these datasets should be aligned as much as possible before the horizontal concatenation of the corresponding individual data matrices in the different acquisition modes. Therefore, MS signals were interpolated to match the chromatographic

retention times of the two acquisition modes before the horizontal concatenation of positive and negative MS signals.

### Multivariate curve resolution-alternating least squares (MCR-ALS)

$$D_{MS} = C\,S_{MS}^{T} + E \tag{1}$$

Previously arranged MSROI data matrices, $D_{MS}$, were decomposed according to a factor decomposition bilinear model using MCR-ALS [18, 30] into the product of two-factor matrices with the elution profiles, $C$, giving the relative concentrations of the compounds, and the MS spectra profiles, $S^{T}$, giving the spectra with the MS signals and their m/z values, for the $N$ selected components. Not explained variance is stored in the residuals/error data matrix, $E$ (Eq. 1).

Components' spectra normalization was carried out as a constraint by MCR-ALS (spectra equal height) during the resolution of the components of the system. The mass spectra of every resolved component (potentially representing a chemical compound) have the MS1 and MS2 signals of the precursor and fragment ions, respectively, for each acquisition and ionization mode.

MCR can be performed separately for each chromatographic run using one acquisition and ionization mode or simultaneously using multiple chromatographic runs and acquisition, as described by Pérez-López et al. [20], and ionization positive and negative modes, as shown in the following Eqs. 2, 3, 4, and 5.

In the individual MCR-ALS analysis of a single chromatographic run by DIA MS1 and MS2 (Eq. 2), the horizontally concatenated data matrix $D_{MS1,MS2}$ is decomposed in the product of two-factor matrices, $C$ and $S_{MS1,MS2}^{T}$, related respectively with the concentration (elution) and with the horizontally concatenated (linked) MS1 and MS2 profiles of the chemical constituents present in the analyzed samples.

$$D_{MS1,MS2} = CS_{MS1,MS2}^{T} + E_{MS1,MS2} \tag{2}$$

When multiple chromatographic runs are analyzed simultaneously, the different individual data matrices, $D_{MS}$, can be concatenated vertically sharing the same spectral vector space, as shown in Eq. 3 for the case of concatenating chromatographic multi-run data matrices acquired in one of the two acquisition modes, either in positive or negative ionization.

$$D_{MS1,MS2,aug} = \begin{bmatrix} D_{MS1,MS2} \\ D_{MS1,MS,2} \\ \cdots \\ D_{MS1,MS2,K} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \cdots \\ C_K \end{bmatrix} S_{MS1,MS2}^{T} \begin{bmatrix} E_1 \\ E_2 \\ \cdots \\ E_K \end{bmatrix}$$
$$= C_{caug} S_{MS1,MS2}^{T} + E_{MS1,MS2,aug} \tag{3}$$

where now $D_{MS1,MS2,aug}$, $C_{aug}$, and $E_{MS1,MS2,aug}$ refer to the augmented vertically concatenated, concentration (elution

profiles), and residual/error matrices, when multiple MSROI data matrices coming from different chromatographic runs are simultaneously analyzed. And $\mathbf{S^T_{MS1, MS2}}$ has as before the horizontally concatenated (linked) MS1 and MS2 profiles of the chemical constituents present in the analyzed samples.

These horizontal and vertical concatenations and MCR-ALS analysis can be also extended (introduced for the first time in this work) to the two possible ionization modes as shown in the following Eqs. 4 and 5

$$\mathbf{D_{MS1+,MS1-MS2-,MS2-}} = \mathbf{C}\, \mathbf{S^T_{MS1+,MS1-,MS2-,MS2-}} \\ + \mathbf{E_{MS1+,MS2+,MS1-,MS2-}} \tag{4}$$

$$\mathbf{D_{MS1+,MS1+MS2-,MS2-,aug}} = \mathbf{C_{aug}}\mathbf{S^T_{MS1+,MS1-,MS2-,MS2-}} \\ + \mathbf{E_{MS1+,MS2+,MS1-,MS2-,aug}} \tag{5}$$

In all these two equations, the $\mathbf{C}$ and $\mathbf{C_{aug}}$ concentration matrices, have the elution profiles of every component resolved by MCR-ALS in one of or in multiple chromatographic runs. From the peak areas or heights of the elution profiles resolved in $\mathbf{C}$ or $\mathbf{C_{aug}}$, relative quantitative information on every resolved component can be obtained. This is especially useful in the case of multiple samples (chromatographic runs) simultaneously analyzed (in $\mathbf{C_{aug}}$), for calibration purposes or semiquantitative estimations, in case of the absence of standards for quantification. With this information, the variation of the concentrations of the metabolites among different samples can be studied.

On the other hand, from the ion signals in the resolved mass spectra, $\mathbf{S^T_{MS1+,MS1-, MS2-,MS2}}$ horizontally concatenated and linked, the compound annotation or identification step is improved.

The complete data analysis workflow and the different types of data matrices concatenations and augmentations are displayed in Figs. 1 (augmentation of multiple acquisition and ionization modes of the amino acids' standard sample) and 2 (augmentation of multiple chromatographic runs at different acquisition and ionization modes for the fish embryo replicate samples).

The steps of the MCR-ALS analysis of the different MSROI data matrices are also given in the Supporting Information and have been described in previous papers [17, 18, 20, 22, 30].

### Identification of the amino acids

The identification of the amino acids from the mass spectra profiles resolved by the MCR-ALS, both in the standard solution and in the fish embryo samples, was initially performed by directly checking if their m/z ROI values for precursor (MS1) and fragment (MS2) ions present in the MCR-ALS resolved mass spectra, $\mathbf{S^T_{MS1,MS2}}$, were available

in public databases (Mass Bank, PubChem [31], accessed Nov-Dec/2022). To confirm the amino acid identity, a minimum mass error of 0.01 Da both for MS1 and MS2 values was considered. Mass measurement accuracy (average of individual mass errors) of 0.0017 Da was obtained for the positive acquisition mode and 0.0034 Da for the negative acquisition mode, and 0.0025 Da for the simultaneous analysis.

In addition, the identity of the amino acids verified by the ROIMCR approach was further confirmed using the Data Analysis from Bruker's MS software, and also the MS-DIAL software [14] by checking the reference-matched (MS1 and MS2 spectra were confirmed with literature databases if a total score > 850 in compound search information was found) and suggested ions (only MS1 spectra is confirmed). To do this, the raw MS data NetCDF files were converted to ABF format and the parameters were set with similar conditions to those used for the ROI analysis, with few modifications (intensity threshold of 1000, mass error of 0.01 Da, and other details given in the Table S3 in the Supporting Information) to increase the sensitivity of the analysis and consider more potential compounds, with milder filtering conditions.

## Results and discussion

### ROIMCR results and parameter selection

Significant reduction of the data dimensionality of the amino acid standard sample was achieved by selecting the following ROI parameters (Table 1): a signal intensity threshold of 5000 (the equivalent of 0.1% of maximum intensity), a mass error tolerance of 0.005 Da, and minimum occurrence of 5 scans. Similar parameters were used for the ROI analysis of fish embryo samples, except for the intensity threshold which was set to a higher value of 10,000 (0.5%) due to the increased size of the data for biological samples.

For the analysis of the amino acids' standards, 30 components were resolved by MCR-ALS (Supporting Table S4), while for the fish embryo samples, 100 components were resolved considering the elevated variety of features present in biological samples. Non-negativity constraints were applied to the elution and mass spectra profiles. All MCR-ALS analyses provided more than 99.5% of explained variance for both the amino acid standard and the fish embryo samples (see Supporting Table S4).

### ROIMCR results in the analysis of the amino acid standard

The LC/Q-TOF method with the HILIC column used in this study represented a fast, simple, and effective strategy for separating and detecting the amino acids (Supporting Figure S1),
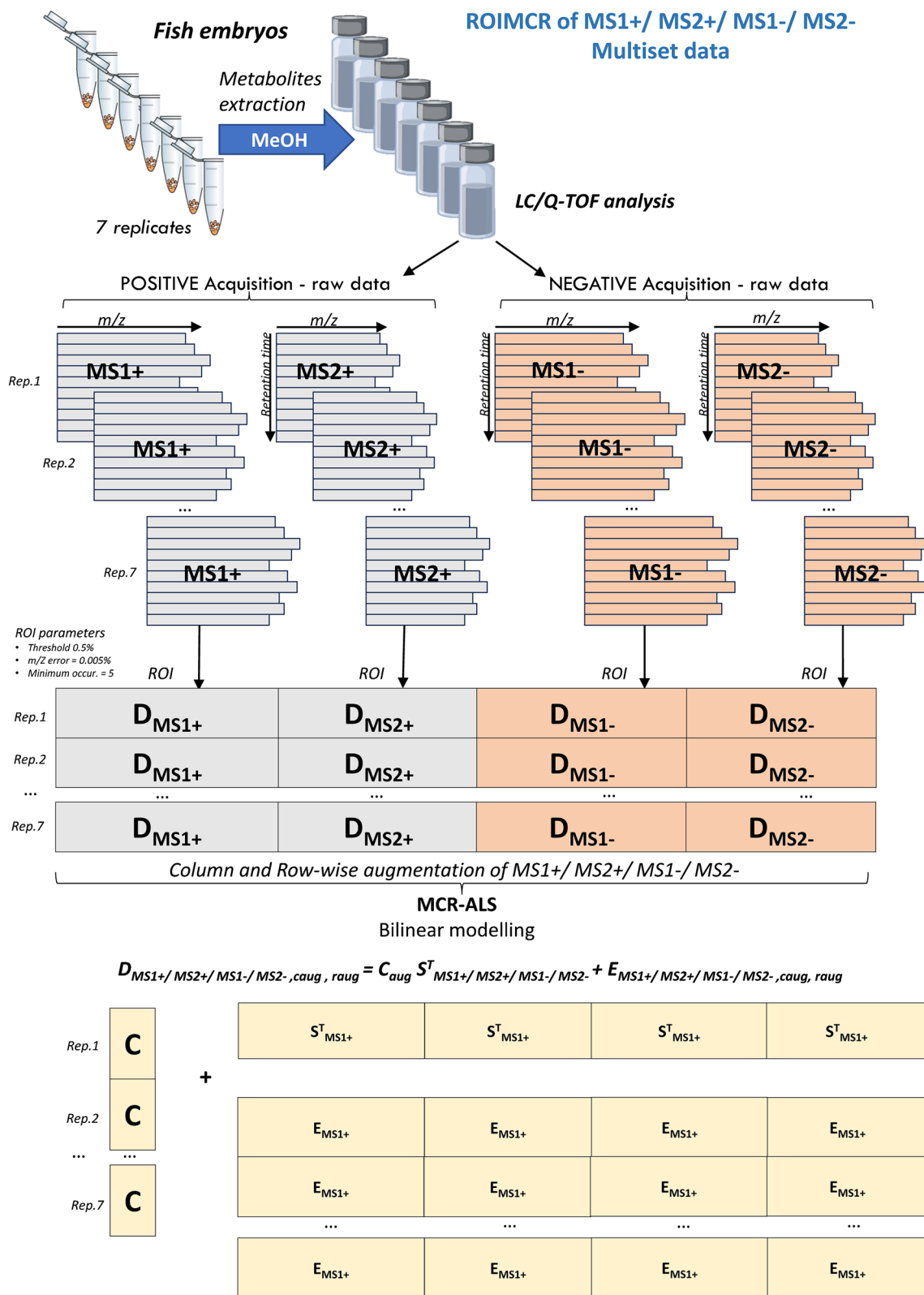
**Fig. 2** Column- and row-wise matrix augmentation and MCR bilinear model factor decomposition of MS1(+), MS2(+), MS1(-), and MS2(-) data acquired in the analysis of multiple fish embryo sample replicates

**Table 1** Regions of interest (ROI) parameters

| Regions of interest (ROI) algorithm parameters | | |
| --- | --- | --- |
| | Aa standard mix | Fish embryo samples |
| Signal intensity threshold | 5000 (~0.1%) | 10,000 (~0.5%) |
| Factor signal (× threshold) | 1 | 1 |
| Mass error deviation | 0.005 | 0.005 |
| Mass units | Daltons | Daltons |
| m/z values calculation | Median | Median |
| Minimum occurrence | 5 | 5 |

as reported by [32]. In addition, the ROIMCR approach solves the mixture analysis problem even in situations of strong compounds' coelutions, by improving the resolution of their elution profiles. Due to the selectivity of high-resolution MS signals, ROIMCR can separate the elution and spectra profiles of a large number of possible chemical constituents in complex biological and environmental samples [7, 19, 27, 29].

ROIMCR analysis allowed identifying all seventeen amino acids present in the standard mixture sample (see

Table 2 and Supporting Figure S2), as well as the two internal standards present in the sample. As expected, the most intense signals were observed at higher m/z values for MS1 when compared to MS2 for both acquisition modes.

Different signal intensities were obtained for the acquisitions in positive and negative modes for the different amino acids. When MS1 and MS2 DIA data in positive and negative ionization modes were analyzed separately (Eq. 2), distinct sets of fifteen amino acids were detected, with the L-proline and L-valine being detected only in the positive mode and the L-glycine and L-cystine only in the negative mode. In general, better ionization efficiency (higher intensity values) was obtained for the amino acids detected at the positive mode (average signal intensities of 928 and 117 units with 10 amino acids with the highest intensities, Table 2), than at the negative mode (average of signal intensities of 395and 887). This is also illustrated in Supporting Figure S2 where higher peaks are observed for the elution profiles in positive mode. On the other hand,

**Table 2** Parameters used for the detection of the seventeen amino acid standard solution and the two additional internal standards analyzed. Exact[1] and experimental[2] exact mass and their maximum height signal intensity values (arbitrary units)[3] obtained at each ionization mode (positive and negative), and their chromatographic retention times[4]. The last column[5] confirms (x) the detection of the amino acid (or internal standard) in the simultaneous data analysis of the positive and negative modes

| Amino acids | | Positive ionization | | | Negative ionization | | | RT[4] (min) | Pos/neg[5] |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | [H⁺] exact mass[1] | [H⁺] exper. Mass[2] | Intensity[3] | [H⁻] exact mass[1] | [H⁻] exper. mass[2] | Intensity[3] | | |
| 1 | L-Phenylalanine | 166.0868 | 166.0846 | **1,091,678** | 164.0712 | 164.0671 | 584,228 | 2.82 | x |
| 2 | L-Isoleucine | 132.1025 | 132.1007 | **1,562,956** | 130.0868 | 130.0837 | 223,230 | 2.89 | x |
| 3 | L-Leucine | 132.1025 | 132.1007 | **1,562,956** | 130.0868 | 130.0837 | 223,230 | 2.89 | x |
| 4 | L-Tyrosine | 182.0817 | 182.0793 | 441,574 | 180.0661 | 180.0616 | **677,328** | 2.99 | x |
| 5 | L-Methionine | 150.0589 | 150.0569 | **946,572** | 148.0432 | 148.0395 | 816,872 | 3.03 | x |
| 6 | L-Valine | 118.0868 | 118.0853 | **775,806** | 116.0712 | ND | ND | 3.24 | x |
| 7 | *Methionine sulfone* | 182.0487 | 182.0463 | 120,358 | 180.03305 | 180.0284 | **808,366** | 3.34 | x |
| 8 | Glycine | 76.03986 | ND | ND | 74.0242 | 74.0223 | **286,062** | 3.58 | x |
| 9 | L-Threonine | 120.0661 | 120.0645 | 95,528 | 118.0504 | 118.0474 | **919,562** | 3.58 | x |
| 10 | *PIPES* | 303.0684 | 303.0643 | 82,214 | 301.05281 | 301.0448 | **611,436** | 3.58 | x |
| 11 | L-Proline | 116.0712 | 116.0697 | **4,609,182** | 114.0555 | ND | ND | 3.61 | x |
| 12 | L-Serine | 106.0504 | 106.0492 | 38,116 | 104.0348 | 104.0322 | **813,598** | 3.72 | x |
| 14 | L-Alanine | 90.05551 | 90.0543 | **70,432** | 88.0399 | 88.0375 | 16,284 | 5.49 | x |
| 13 | L-Glutamic acid | 148.061 | 148.0592 | 96,378 | 146.0453 | 146.0415 | **617,628** | 4.00 | x |
| 15 | L-Aspartic acid | 134.0453 | 134.0438 | 8814 | 132.0297 | 132.0262 | **57,966** | 5.49 | x |
| 16 | L-Arginine | 175.1195 | 175.1174 | **2,055,464** | 173.1039 | 173.0992 | 195,606 | 6.14 | x |
| 17 | L-Lysine | 147.1134 | 147.1116 | **420,304** | 145.0977 | 145.0939 | 37,522 | 6.39 | x |
| 18 | Cystine | 241.0316 | ND | ND | 239.0160 | 239.0098 | **8722** | 6.49 | * |
| 19 | L-Histidine | 156.0773 | 156.0752 | **146,004** | 154.0616 | 154.0577 | 129,352 | 7.08 | x |

*ND*, non-detected amino acid in one mode of acquisition

*Non-detected amino acid due to its poor ionization in both modes of acquisition under tested conditions

Italic font - Metabolites used as internal standards

Bold values - Highest intensity values for one of the acquisition modes

both internal standards showed higher signal intensities in negative mode.

According to the literature, the ionization of amino acids is mostly operated in the positive mode due to their higher ionization efficiency [11, 32, 33]. But alternatively, in some cases, it gives also good results in negative acquisition [34]. Liigand et al. [10] reported better ionization energy in the negative mode for 33 compounds that ionize in both modes, including amino acids. The amino acids' inherent amphoteric property allows functional groups to either receive or donate protons, therefore having a zwitterionic form. Indeed, their different physiochemical properties determine their different ionization efficiencies [32]. When MS1 and MS2 DIA data in both positive and negative ionization modes were simultaneously analyzed (Eq. 4) by the ROIMCR, all amino acids (except for cysteine due to its poor ionization in both modes under tested conditions) present in the amino acid standard mixture were properly detected and resolved.

Even so, this evidences the relevance and advantages of this method, since for instance, the single analysis in positive mode could have missed the detection of the glycine, while in negative mode could have missed the identification of valine and proline.

Figure 3 gives, as an example, the elution and spectra profiles of two different amino acids (L-phenylalanine and L-proline) detected in MS1 and MS2 DIA in simultaneous data analysis of positive and negative acquisition modes (TIC, total ion chromatograms are shown in Figure S1 in the Supporting Information). Apart from their molecular formulas, MS1 and MS2 major ions in positive and negative modes obtained from their respective ROIMCR resolved spectra in the four modes, MS1(+)/MS2(+)/MS1(-)/MS2(-), are given, as well as their elution profiles plotted for the number of scans or with respect to the retention times in the $x$-axis. In this case, as expected, L-phenylalanine elutes earlier (169.3 s) than L-proline (217.1 s). This is explained by the lower polarity of the former when compared with the latter. For both amino acids, higher ionization efficiencies were observed in the positive acquisition mode as reflected in the different scales of the $y$-axis of the resolved mass spectra. In the four cases (MS1(+), MS2(+), MS1(-), MS2(-) signals), their values were scaled to their maximum signal intensity value equal to 1, which was achieved for MS1 in positive mode for both L-phenylalanine and L-proline. In fact, the latter did not give a significant MS- signal. The annotation of the different amino acids can be potentially enhanced using the correspondent link between the main precursor ions in positive (protonated ion [H$^+$]) and negative (deprotonated ion [H$^-$]) modes, differing approximately 2.01 m/z units. Thus, the proposed ROIMCR allows the direct linking and correspondence between the MS signals in the four acquisition modes (MS1(+)/MS2(+)/MS1(-)/MS2(-)), for the same chemical species (see Tables 2 and S5), when performing their simultaneous analysis. Linking the spectra of the different amino acids in their four acquisition modes facilitates their

coverage, identification, and possible quantitation [20]. Table S5 (in Supplementary Information) describes the list of MS2 fragment ions obtained in the ROIMCR analysis, confirmed by the literature database, that facilitated the annotation and identification of the amino acids in the standard sample.

## ROIMCR results in the analysis of the fish embryo samples

After confirming the detection of the amino acids in their standard mixtures, their presence in the fish embryo samples was analyzed using the ROIMCR approach. Similarly to the pattern observed for the amino acids' standard, it was possible to observe a better ionization efficiency in the positive mode, with more ROI signals of higher intensity being obtained in this acquisition mode when compared to the negative mode (Table 2, Supporting Figure S3). All seven replicates in each acquisition mode gave relatively similar elution profiles. Higher intensity signals were obtained for MS1 ions in comparison to MS2 ion fragments.

In Fig. 4, ROIMCR results are shown for the resolution of L-threonine (up) and L-valine (down) amino acids in the analysis of the seven different embryo samples. As in Fig. 3, m/z values of the major MS signals for the two amino acids are given for the different acquisition and ionization modes, which agree with those expected for these two amino acids (see Table 2). Whereas L-threonine gave significant signals at the four acquisition and ionization modes, L-valine was giving MS signal only in positive ionization, for both MS1 and MS2 DIA acquisition modes.

In the upper subplot for L-threonine and the lower subplot for L-valine, the seven elution profiles obtained by ROIMCR in the resolution of the seven embryo replicate samples are shown giving good peak shapes. The reproducibility of these peaks indicates the adequacy of the analytical procedure and the correct ROIMCR resolution of these two amino acids in the different fish embryo samples analyzed with only some small differences in the height and peak areas (also in the two subplots below). In the overlayed elution profiles plot also given in the two subplots, proper peak alignment and shapes among the elution profiles of both amino acids were achieved. Finally, on the right side of the two subplots, the spectra profiles of the amino acid show the MS1 and MS2 DIA signals for the two ionization modes. Whereas for L-threonine, MS1 and MS2 DIA signals were recorded in the four acquisition and ionization modes, and no MS signal was obtained for valine in the negative mode due to its poorer ionization at the conditions of this work.

In traditional analytical work, a broad assessment of multiple classes of compounds usually requires two separate analyses in positive and negative modes, and the optimization of chromatographic conditions [34, 35]. Even when both ionization modes are measured simultaneously (using switching acquisition), the data analysis is usually performed separately
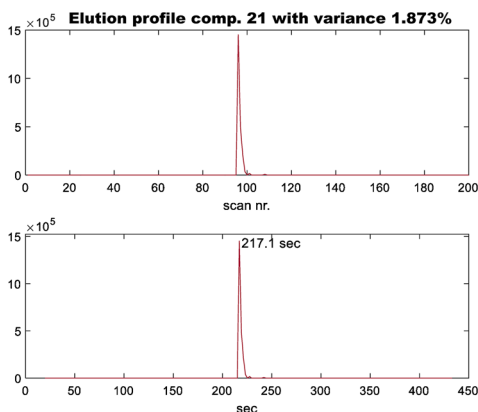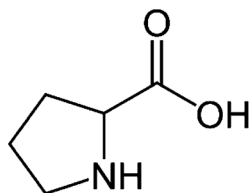
# L-Phenylalanine
## (*Aromatic side-chains*)
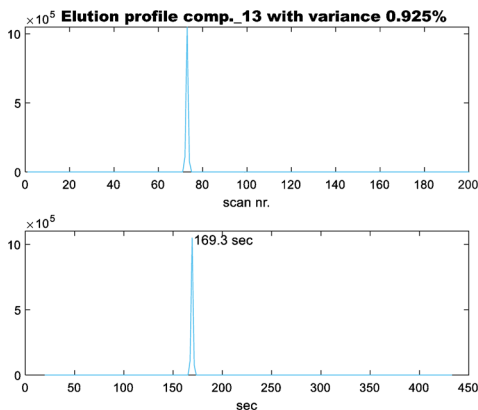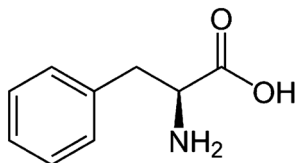
| m/z | Positive | Negative |
|---|---|---|
| MS1 | 166.08680 | 164.0671 |
| MS2 | 120.0796/ 103.0533 | 147.041 |
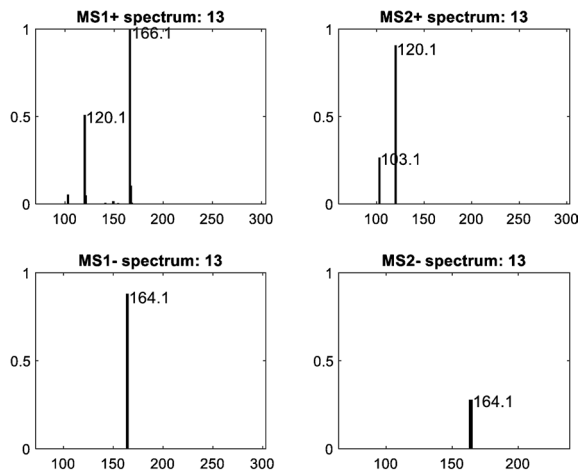


# L- Proline
## (*Aliphatic side-chains*)

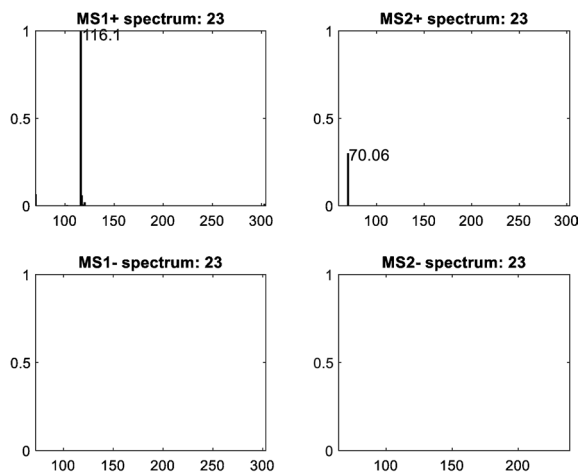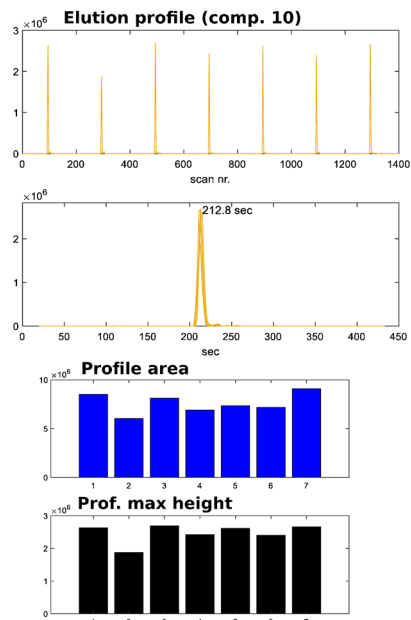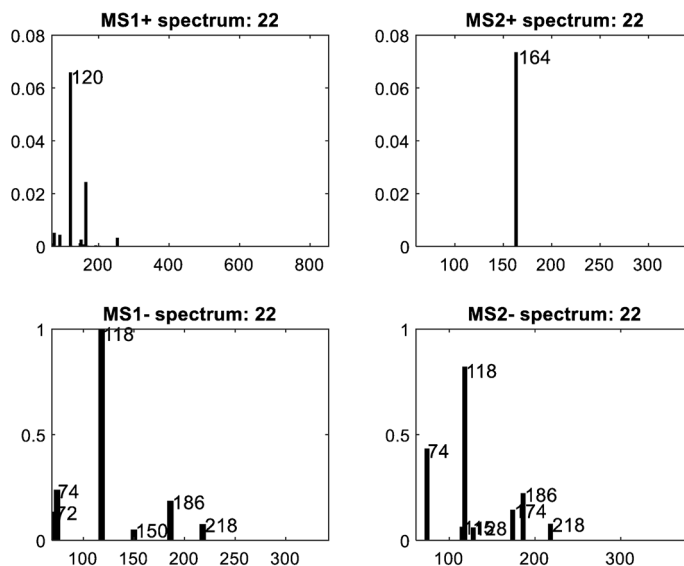| m/z | Positive | Negative |
|---|---|---|
| MS1 | 116.0711 | - |
| MS2 | 70.065 | - |



**Fig. 3** Elution and spectra profiles of two different amino acids (L-phenylalanine and L-proline) resolved by ROIMCR simultaneous analysis of MS and MS/MS DIA data in positive and negative acquisition modes. Left panels: molecular formula of the identified amino acid (upper), ROIMCR elution profiles of the amino acid vs scan number (middle), and vs retention time (lower). Right panels: list of m/z values with more significative MS1 and MS2 signals (upper), and MS1(+), MS2(+), MS1(-), and MS2(-) ROIMCR resolved spectra profiles with the m/z values of their more intense signals (lower)
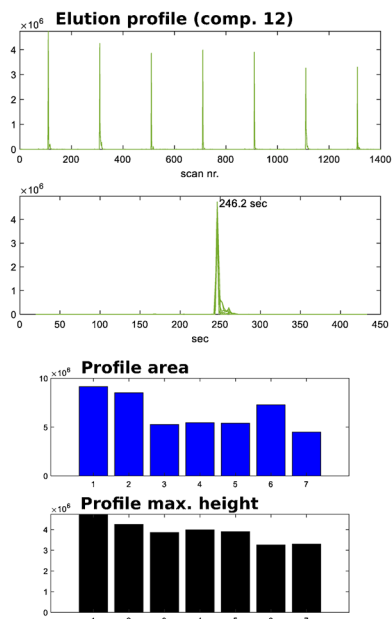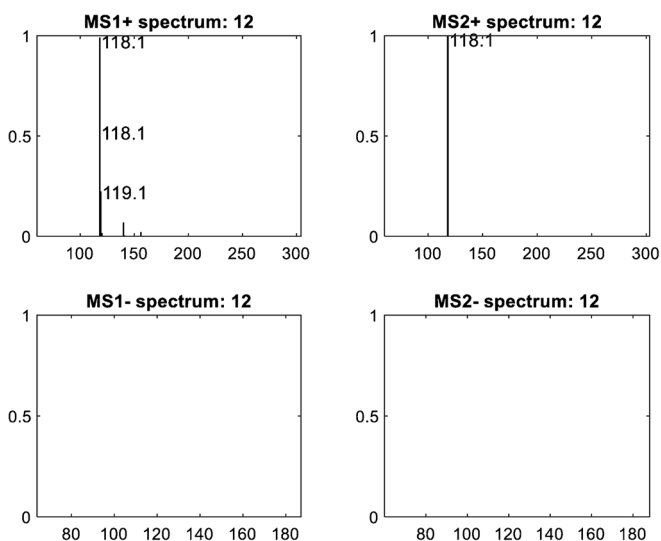
**Fig. 4** Elution and spectra profiles of two different amino acids (ʟ-threonine and ʟ-valine) detected in the fish embryo samples, resolved by the ROIMCR simultaneous analysis of MS1 and MS2 DIA data in positive and negative acquisition modes. Left panels: ROIMCR elution profiles of the amino acid vs scan number (upper), vs retention time (middle), and their peak areas and maximum heights for each sample analyzed. Right panels: MS1 and MS2 signals (upper) and MS1(+), MS2(+), MS1(-), and MS2(-) (lower) signals of the ROIMCR resolved spectra profiles giving the m/z values of their more intense signals

for some specific ions and m/z values [36, 37]. In contrast, the possible simultaneous analysis of MS1 and MS2 DIA signals in positive and negative modes, as demonstrated in this work, can increase the number of simultaneously detected compounds with broader chemical coverage and diversity, optimizing efforts in a shorter time. Moreover, the identification steps are extremely facilitated with the direct links established between precursor ions (in MS1) and fragment ions (in MS2 DIA) and with different signals in the two ionization modes. The suitability of the ROIMCR concept associated with the bilinear factor decomposition of both MS1 and MS2 signals arranged in a single matrix has been recently shown by Pérez-López et al. [20], which is extended now in this work with the inclusion of MS1 and MS2 signals in both ionization modes.

In addition, the comparison of ROIMCR results with those obtained using the MS-DIAL method [14, 38, 39] has confirmed the identity of the seventeen amino acids and the two internal standards, with their corresponding m/z values in MS1 and MS2 DIA and with matched retention times. Since MS-DIAL does not allow performing the simultaneous analysis of the positive and negative ionization modes, it was necessary to perform the ion search separately. For the positive mode, MS-DIAL identified the presence of thirteen amino acids detected in the fish embryo samples (reference matched, total score > 850) and the presence of L-threonine and L-aspartate as suggested ions (only MS1 spectra confirmed). On the negative mode, fewer amino acids were identified as reference-matched ions (both MS1 and MS2 spectra confirmed with databank). MS-DIAL only detected five amino acids with a good reference match, and other eight ions were suggested. For both positive and negative modes, the surrogate methionine sulfone was also found as a reference-matched metabolite as well as the internal standard PIPES as a suggested ion. Although MS-DIAL is a well-recognized method that greatly facilitates compound searching, it requires the separated analysis of the MS signals in the two ionization modes.

## Conclusions

Results obtained in this study showed for the first time the suitability of the simultaneous data analysis of MS1(+)/MS2(+)/MS1(-)/MS2(-) signals using the ROIMCR approach in the analysis of a standard mixture of amino acids in solution and fish embryo extracts. Comparison of results obtained in the simultaneous analysis of all acquisition (MS1 and MS2) and ionization (positive and negative) data modes with the results obtained in the individual separate analysis of the different acquisition and ionization modes showed that time efforts and metabolite coverage can be improved in the simultaneous analysis. The link established between the different MS1+, MS2+, MS1, and MS2- acquisitions and ionizations for the same

chemical compound greatly facilitates the identification of the metabolites present in the analyzed samples. The identification of the amino acids in the fish embryo samples by the proposed ROIMCR method has confirmed the reliability of the proposed strategy for untargeted metabolomics analysis. The approach proposed in this work is an important improvement in the context of the rapid growth of untargeted types of analysis using data-independent acquisition modes. This study encourages further metabolomic investigations using the ROIMCR-described approach combining different MS ionization and acquisition modes.

## Declarations

**Ethics approval** This study had the approval from the Ethics Committee for the use of animals from the São Paulo State University (Protocol 12/2018-CEAUA) and the experimental conditions complied with the National Council for the Control of Animal Experimentation (CONCEA, law 11.794/2008). The biological material was obtained by the Panama Fish Farm (Paulo Lopes, Brazil) https://www.pisciculturapanama.com.br/.

**Competing interests** The authors declare no competing interests.

## References

1. Debnath M, Prasad GBKS, Bisen PS. Omics technology. Molec Diagn: Promises Possibilities. 2010:11–31. https://doi.org/10.1007/978-90-481-3261-4_2.

2. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;2017(18):1–15. https://doi.org/10.1186/S13059-017-1215-1.

3. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: technologies and their applications. J Chromatogr Sci. 2017;55:182–96. https://doi.org/10.1093/chromsci/bmw167.

4. Bouhifd M, Hartung T, Hogberg HT, Kleensang A, Zhao L. Review: Toxicometabolomics. J Appl Toxicol. 2013;33:1365–83.

5. Koelmel JP, Napolitano MP, Ulmer CZ, Vasiliou V, Garrett TJ, Yost RA, Prasad MNV, Godri Pollitt KJ, Bowden JA. Environmental lipidomics: understanding the response of organisms and ecosystems to a changing world. Metabolomics. 2020;16:1–22. https://doi.org/10.1007/s11306-020-01665-3.

6. Martín-Blázquez A, Díaz C, González-Flores E, Franco-Rivas D, Jiménez-Luna C, Melguizo C, Prados J, Genilloud O, Vicente F, Caba O, Pérez del Palacio J. Untargeted LC-HRMS-based metabolomics to identify novel biomarkers of metastatic colorectal cancer. Sci Rep. 2019;9:1–9. https://doi.org/10.1038/s41598-019-55952-8.

7. Ortiz-Villanueva E, Benavente F, Piña B, Sanz-Nebot V, Tauler R, Jaumot J. Knowledge integration strategies for untargeted metabolomics based on MCR-ALS analysis of CE-MS and LC-MS data. Anal Chim Acta. 2017;978:10–23. https://doi.org/10.1016/j.aca.2017.04.049.

8. Rochat B, Mohamed R, Sottas PE. LC-HRMS metabolomics for untargeted diagnostic screening in clinical laboratories: a feasibility study. Metabolites. 2018;8:1–17. https://doi.org/10.3390/metabo8020039.

9. Salihovic S, Broeckling CD, Ganna A, Prenni JE, Sundström J, Berne C, Lind L, Ingelsson E, Fall T, Ärnlöv J. Nowak C (2020) Non-targeted urine metabolomics and associations with prevalent and incident type 2 diabetes. Sci Rep. 2020;10:1–9. https://doi.org/10.1038/s41598-020-72456-y.

10. Liigand P, Kaupmees K, Haav K, Liigand J, Leito I, Girod M, Antoine R, Kruve A. Think negative: finding the best electrospray ionization/MS mode for your analyte. Anal Chem. 2017;89:5665–8. https://doi.org/10.1021/ACS.ANALCHEM.7B00096/ASSET/IMAGES/LARGE/AC-2017-000969_0003.JPEG.

11. Mak J, Cowan TM, Le A. Quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry. Methods Mol Biol. 2019;2030:85–109. https://doi.org/10.1007/978-1-4939-9639-1_8.

12. Verhoeven A, Giera M, Mayboroda OA. Scientific workflow managers in metabolomics: an overview. Analyst. 2020;145:3801–8. https://doi.org/10.1039/D0AN00272K.

13. Tsugawa H, Ikeda K, Takahashi M, Satoh A, Mori Y, Uchino H, Okahashi N, Yamada Y, Tada I, Bonini P, Higashi Y, Okazaki Y, Zhou Z, Zhu ZJ, Koelmel J, Cajka T, Fiehn O, Saito K, Arita M. Arita M (2020) A lipidome atlas in MS-DIAL 4. Nat Biotechnol. 2020;38(10):1159–63. https://doi.org/10.1038/s41587-020-0531-2.

14. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, Vandergheynst J, Fiehn O. Arita M (2015) MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods. 2015;12(6):523–6. https://doi.org/10.1038/nmeth.3393.

15. Lynn KS, Cheng ML, Chen YR, Hsu C, Chen A, Lih TM, Chang HY, Huang CJ, Shiao MS, Pan WH, Sung TY, Hsu WL. Metabolite identification for mass spectrometry-based metabolomics using multiple types of correlated ion information. Anal Chem. 2015;87:2143–51. https://doi.org/10.1021/ac503325c.

16. Xiao JF, Zhou B, Ressom HW. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. Trends Analyt Chem. 2012;32:1. https://doi.org/10.1016/J.TRAC.2011.08.009.

17. Gorrochategui E, Jaumot J, Tauler R. ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets. BMC Bioinformatics. 2019;20:1–17. https://doi.org/10.1186/S12859-019-2848-8/FIGURES/6.

18. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: new features and applications. Chemom Intell Lab Syst. 2015;140:1–12. https://doi.org/10.1016/j.chemolab.2014.10.003.

19. Bayat M, Marín-García M, Ghasemi JB, Tauler R. Application of the area correlation constraint in the MCR-ALS quantitative analysis of complex mixture samples. Anal Chim Acta. 2020;1113:52–65. https://doi.org/10.1016/j.aca.2020.03.057.

20. Pérez-López C, Oró-Nolla B, Lacorte S, Tauler R. Regions of interest multivariate curve resolution liquid chromatography with data-independent acquisition tandem mass spectrometry. Anal Chem. 2023. https://doi.org/10.1021/ACS.ANALCHEM.2C05704.

21. Pérez-Cova M, Bedia C, Stoll DR, Tauler R, Jaumot J. MSroi: a pre-processing tool for mass spectrometry-based studies. Chemometr Intell Lab Syst. 2021;215. https://doi.org/10.1016/j.chemolab.2021.104333.

22. De Juan A, Jaumot J, Tauler R. Multivariate curve resolution (MCR). Solving the mixture analysis problem. Anal Methods. 2014;6:4964–76. https://doi.org/10.1039/c4ay00571f.

23. Março PH, Valderrama P, Alexandrino GL, Poppi RJ, Tauler R. Resolução multivariada de curvas com mínimos quadrados alternantes: Descrição, funcionamento e aplicações. Quim Nova. 2014;37:1525–32.

24. Tauler R, Gorrochategui E, Jaumot J, Tauler R. A protocol for LC-MS metabolomic data processing using chemometric tools. Protoc Exch. 2015;1–28. https://doi.org/10.1038/protex.2015.102

25. Tauler R. Multivariate curve resolution of multiway data using the multilinearity constraint. J Chemom. 2021;35:1–24. https://doi.org/10.1002/cem.3279.

26. Yamamoto FY, Souza ATC, de Paula V de CS, Beverari I, Garcia JRE, Padial AA, de Souza Abessa DM. From molecular endpoints to modeling longer-term effects in fish embryos exposed to the elutriate from Doce River. Sci Total Environ. 2022;846:157332. https://doi.org/10.1016/j.scitotenv.2022.157332.

27. Ortiz-Villanueva E, Jaumot J, Martínez R, Navarro-Martín L, Piña B, Tauler R. Assessment of endocrine disruptors effects on zebrafish (Danio rerio) embryos by untargeted LC-HRMS metabolomic analysis. Sci Total Environ. 2018;635:156–66. https://doi.org/10.1016/j.scitotenv.2018.03.369.

28. WATERS. Gradient separation of amino acids on ACQUITY UPLC BEH HILIC | Waters. In: Application Note WA60134. 2009. https://www.waters.com/nextgen/en/library/application-notes/2009/gradient-separation-of-amino-acids-on-acquity-uplc-beh-hilic.html. Accessed 10 Jan 2023.

29. Gorrochategui E, Jaumot J, Lacorte S, Tauler R. Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: overview and workflow. TrAC - Trends Anal Chem. 2016;82:425–42. https://doi.org/10.1016/j.trac.2016.07.004.

30. Tauler R. Multivariate curve resolution applied to second order data. Chemom Intell Lab Syst. 1995;30:133–46. https://doi.org/10.1016/0169-7439(95)00047-X.

31. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45:703–14. https://doi.org/10.1002/JMS.1777.

32. Violi JP, Bishop DP, Padula MP, Steele JR, Rodgers KJ. Considerations for amino acid analysis by liquid chromatography-tandem mass spectrometry: a tutorial review. TrAC Trends Anal Chem. 2020;131:116018. https://doi.org/10.1016/J.TRAC.2020.116018.

33. Liu Z, Tu MJ, Zhang C, Jilek JL, Zhang QY, Yu AM. A reliable LC-MS/MS method for the quantification of natural amino acids in mouse plasma: method validation and application to

a study on amino acid dynamics during hepatocellular carcinoma progression. J Chromatogr B Analyt Technol Biomed Life Sci. 2019;1124:72. https://doi.org/10.1016/J.JCHROMB.2019.05.039.

34. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens JP, Morla A, Bouchu D. ESI-MS/MS analysis of underivatised amino acids: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionisation mode. Rapid Commun Mass Spectrom. 2003;17:1297–311. https://doi.org/10.1002/RCM.1054.

35. Bang DY, Lim S, Moon MH. Effect of ionization modifiers on the simultaneous analysis of all classes of phospholipids by nanoflow liquid chromatography/tandem mass spectrometry in negative ion mode. J Chromatogr A. 2012;1240:69–76. https://doi.org/10.1016/J.CHROMA.2012.03.073.

36. Yang Z, Anacleto J, Maljers L. Positive-negative switching LC–MS–MS for quantification of pesticides in juice. Column. 2014;10:9428–35. https://doi.org/10.1021/JF503468T.

37. Dey AS, Ayon NJ, Bhattacharya C, Gutheil WG, Mukherji M. Positive/negative ion-switching-based LC–MS/MS method for quantification of cytosine derivatives produced by the TET-family 5-methylcytosine dioxygenases. Biol Methods Protoc. 2020;5:1–9. https://doi.org/10.1093/BIOMETHODS/BPAA019.

38. Tada I, Tsugawa H, Meister I, Zhang P, Shu R, Katsumi R, Wheelock CE, Arita M, Chaleckis R. Creating a reliable mass spectral–retention time library for all ion fragmentation-based metabolomics. Metabolites 2019;9. https://doi.org/10.3390/metabo9110251.

39. Lai Z, Tsugawa H, Wohlgemuth G, Mehta S, Mueller M, Zheng Y, Ogiwara A, Meissen J, Showalter M, Takeuchi K, Kind T, Beal P, Arita M. Fiehn O (2017) Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. Nat Methods. 2017;15(1):53–6. https://doi.org/10.1038/nmeth.4512.

40. OECD. Test No. 236: Fish embryo acute toxicity (FET) test. OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing 2013;1–22. https://doi.org/10.1787/9789264203709-en