Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-020-02789-0

RESEARCH PAPER

MCR-ALS analysis of ^{1}H NMR spectra by segments to study the zebrafish exposure to acrylamide

Yolanda Pérez[1] · Marta Casado[2] · Demetrio Raldúa[2] · Eva Prats[2] · Benjamín Piña[2] · Romà Tauler[2] · Ignacio Alfonso[3] · Francesc Puig-Castellví[2,4]

Received: 11 April 2020 / Revised: 31 May 2020 / Accepted: 24 June 2020

Abstract

Metabolomics is currently an important field within bioanalytical science and NMR has become a key technique for drawing the full metabolic picture. However, the analysis of ^{1}H NMR spectra of metabolomics samples is often very challenging, as resonances usually overlap in crowded regions, hindering the steps of metabolite profiling and resonance integration. In this context, a pre-processing method for the analysis of 1D ^{1}H NMR data from metabolomics samples is proposed, consisting of the blind resolution and integration of all resonances of the spectral dataset by multivariate curve resolution-alternating least squares (MCR-ALS). The resulting concentration estimates can then be examined with traditional chemometric methods such as principal component analysis (PCA), ANOVA-simultaneous component analysis (ASCA), and partial least squares-discriminant analysis (PLS-DA). Since MCR-ALS does not require the use of spectral templates, the concentration estimates for all resonances are obtained even before being assigned. Consequently, the metabolomics study can be performed without neglecting any relevant resonance. In this work, the proposed pipeline performance was validated with 1D ^{1}H NMR spectra from a metabolomics study of zebrafish upon acrylamide (ACR) exposure. Remarkably, this method represents a framework for the high-throughput analysis of NMR metabolomics data that opens the way for truly untargeted NMR metabolomics analyses.

Keywords MCR-ALS · NMR · Chemometrics · Metabolite discovery · 1H · Metabolomics

Abbreviations

1D ^{1}H NMR	One-dimensional proton nuclear magnetic resonance
AAMA	*N*-Acetyl-S-(carbamoylethyl)-L-cysteine
ACR	Acrylamide
ANOVA	Analysis of variance
ASCA	ANOVA-simultaneous component analysis
AXP	Adenosine nucleotides
BATMAN	Bayesian automated metabolite analyzer for NMR
C	Concentrations matrix in the MCR-ALS analysis
D	Input matrix in the MCR-ALS analysis
DSS	2,2-Dimethyl-2-silapentane-5-sulfonate
GABA	Gamma-aminobutyric acid
GUI	Graphical user interface
MCR-ALS	Multivariate curve resolution-alternating least squares
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect SpectroscopY
PC1	First principal component
PC2	Second principal component
PCA	Principal component analysis
PLS-DA	Partial least squares-discriminant analysis
PQN	Probabilistic quotient normalization
SCA	Simultaneous component analysis
S^{T}	Spectrum matrix in the MCR-ALS analysis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-020-02789-0) contains supplementary material, which is available to authorized users.

✉ Francesc Puig-Castellví
francesc.puigcastellvi@agroparistech.fr

[1] NMR Facility, IQAC-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

[2] Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

[3] Department of Biological Chemistry, IQAC-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

[4] Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 75005 Paris, France

Published online: 02 July 2020

SVD	Single value decomposition
UDP	Uridine diphosphate

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most important analytical techniques used for metabolite discovery [1, 2]. In the biological field, NMR is also employed in metabolomics studies to characterize the metabolites in cell extracts, tissues, and living organisms for disease diagnosis and biomarker discovery [3]. The interpretation of NMR spectra from biological samples is not easy due to their complex composition. In a typical metabolomics NMR spectrum, hundreds of resonances from several dozens of metabolites are obtained [4].

Since resonances cannot be directly assigned because of their strong overlap, the NMR spectra are usually investigated using multivariate data analysis approaches [5–10], such as principal component analysis (PCA) [11], ANOVA-simultaneous component analysis (ASCA) [12], or partial least squares-discriminant analysis (PLS-DA) [13]. When these approaches are used for the analysis of NMR spectra, the chemical shifts of the resonances from the altered metabolites can be identified. However, the other spectroscopic parameters (*J*, multiplicity, proton number) are not directly retrieved from these analyses and further work needs to be done by the NMR spectroscopist to successfully characterize the metabolites of interest. Moreover, to obtain quantitative information, these chemometric analyses should be followed by the integration of the relevant resonances on the original NMR spectra using deconvolution approaches.

To improve the spectroscopic interpretation of these chemometric analyses of NMR data and to avoid the integration and deconvolution steps, we have designed a new metabolomics workflow that introduces as a preliminary step the application of the multivariate curve resolution-alternating least squares (MCR-ALS) [14] method. MCR-ALS allows resolving, for every metabolite, their resonances and their corresponding integrals (or relative concentrations) [15]. Thus, after obtaining the relative concentrations from this step, chemometric (e.g., PCA, PLS-DA) and statistical (e.g., *t* test) analysis can be performed on them.

MCR-ALS performance to resolve all the pure spectral resonances (and pure integrals) of the metabolites in the analyzed sample depends on the variance of its associated dataset [14–17]. As a matter of fact, the resolved components obtained in the MCR-ALS analysis of 1D ^{1}H NMR metabolomics datasets usually contain mixed contributions of several metabolites and they are descriptive of metabolic or physiologic processes [18–20]. Having said that, it is possible to achieve an improved performance and get the pure resonance profiles of the metabolites in the analyzed samples by using the DTC-MCR-ALS methodology, which constrains the search and resolution of resonances in the narrow spectral region where they are located [18]. The DTC-MCR-ALS approach works in two steps. In the first step, the NMR data is split into smaller datasets (or segments) comprising each one a spectral segment of the original dataset and then each one is analyzed by MCR-ALS. In a second step, the original full-length NMR data is analyzed with MCR-ALS in combination with a selectivity constraint built after examining the data from the first step using a decision tree approach [18]. Since the smaller segments analyzed in the first step of the DTC-MCR-ALS contain resonances from a lower number of metabolites, the complexity of each data segment would be substantially lower than in the original raw data, and therefore, a better MCR-ALS resolution of the pure resonances using a lower number of components is feasible.

In this work, we have investigated the use of the first step of the DTC-MCR-ALS as an independent data pre-processing method to retrieve the concentration and clean spectral profiles of the resonances present in the analyzed ^{1}H NMR dataset. The principle of using MCR-ALS as a pre-processing tool for ^{1}H NMR data was first introduced by Ebrahimi et al. in 2016 [15]. In this paper, MCR-ALS was applied on 7 segments from a 1D ^{1}H NMR dataset comprising 1632 samples to obtain the resonance integrals from 7 known metabolites. This study can be regarded as targeted, although it showed it is possible to obtain the concentration profiles of NMR resonances without needing to assume their peak shape as with deconvolution methods [7, 21–24], revealing a high potential for being applied in untargeted studies.

Contemporary to our work, Khakimov showed that MCR-ALS analysis on the NMR segments can be used to examine metabolomics data from urine samples [25]. By using this approach, they were able to recover the concentration profiles from 209 resonances. In addition, due to the complexity or urine samples, 94 NMR segments containing the most shifted resonances were analyzed with binning instead of MCR-ALS. Their analyses revealed that the pre-processing based on MCR-ALS produced the lowest variability within group and the largest between groups if compared with other NMR pre-processing methods. However, due to the methodological purpose of the study, the differences between groups at a metabolic level were not examined.

In this regard, in the present study, we have investigated a ^{1}H NMR dataset of metabolic extracts by windowing and MCR-ALS analysis. Specifically, we have evaluated the metabolic alterations in zebrafish muscle and brain upon acrylamide (ACR) exposure. Zebrafish is a vertebrate model used to study ACR neurotoxicity [26]. ACR is a neurotoxic, mutagenic, and reprotoxic compound that can be formed in cooking at elevated temperatures [27, 28]. Major ACR symptoms related include lethargy, ataxia, skeletal muscle weakness, and

numbness of the extremities [29]. Due to its public health risk, ACR consumption levels are regulated in some countries [30, 31].

Samples included in the dataset comprised an experimental design of 2 factors and 2 levels, which were investigated by PCA, ASCA, and PLS-DA. Opposed to Khakimov's study [25], we were able to analyze all the spectral segments with MCR-ALS since they do not contain strong peak shifts as in the urine samples. Moreover, the PLS-DA analyses revealed metabolomics differences between groups that could be linked to the system of study.

Experimental section

Biological experiment and dissection

Wild-type zebrafish were obtained from Piscicultura Superior (Barcelona, Spain) and maintained in fish water (reverse osmosis purified water containing 90 mg mL/Instant Ocean® (Aquarium Systems, Sarrebourg, France), 0.58 mM $CaSO_4 \cdot 2H_2O$) at 28 ± 1 °C under a 12L:12D photoperiod. Adult zebrafish (≈ 50:50 male:female ratio) were exposed for 72 h to 0.75 mM ACR (Sigma-Aldrich, St. Louis, MO) in fish water. Control fish were maintained in fish water under identical conditions. Experimental solutions were renewed after 48 h of exposure.

For sample collection, fish were euthanized by inducing hypothermic shock in ice-chilled water (2° to 4 °C). Brain and muscle tissue were immediately excised and pooled (4 brains or muscle/sample) and stored at − 80 °C for further analyses. All procedures were approved by the Institutional Animal Care and Use Committees at the CID-CSIC and conducted in accordance with the institutional guidelines under a license from the local government (agreement number 9027).

Metabolite extraction and sample preparation

The extraction method protocol from Raldúa et al. [32] was used. Tissue samples were freeze-dried and homogenized in 1500 μl of a solution of cold $CHCl_3$:MeOH (2:1) using a TissueLyser (Qiagen, 2 stainless steel beads, 50 Hz, 2 min for brain, 4 min for muscle). Next, 350 μl of cold Milli-Q water was added and the biphasic system was mixed at 4 °C for 20 min by orbital shaking. After centrifugation (13,000 rpm, 4 °C, 20 min), the aqueous phase was collected and transferred to a 2-ml Eppendorf tube. The extraction process was repeated once. The two supernatants were combined and freeze-dried, and the resulting dried residue was stored at − 80 °C until analyzed.

The extracts were dissolved in 700 μl of deuterated phosphate buffer (Na_2DPO_4 25 mM, pH 7.0 in D_2O) with DSS 0.2 mM as internal standard. After centrifugation at 13,000 rpm for 5 min at 4 °C, 650 μl of the supernatant was collected and used for the NMR analysis. In total, 20 samples were prepared (5 per every individual condition).

NMR spectroscopy

1D NOESY-presat NMR spectra (Bruker library—noesygppr1d) were recorded at 298 K on a Bruker Avance III console (Bruker Biospin, Germany) combined with a 9.7 T Oxford Instruments magnet (^{1}H 500 MHz), equipped with a 5-mm inverse cryogenically cooled triple resonance TCI (^{1}H, ^{13}C, ^{15}N, and ^{2}H lock) cryoprobe with a z-axis magnetic field gradient capability.

For every tube, the sample was locked to the solvent, tuned and matched, and shimmed, and the optimal 90° pulse was calculated using the automated routine. The 90° length value and the saturation frequency and power (presaturation power corresponding to a 25-Hz field) for each sample and the rest of fixed acquisition parameters were used for the 1D NOESY acquisition. After spectral recording, for every sample, the optimal water saturation (the residual water peak should be small and to allow good phase correction) and the line width at half height of the DSS (which should be less than 0.7 Hz with no line broadening) were also checked [32].

In addition, for one of the samples of each group (untreated/exposed to ACR), an ensemble of 2D NMR experiments was acquired (^{1}H J-resolved, COSY and TOCSY, and ^{1}H-^{13}C HSQC; acquisition parameters in the Electronic Supplementary Material, ESM).

Pre-processing of NMR spectra

^{1}H (1D NOESY-presat) NMR spectra were automatically referenced, phased, and baseline corrected using TopSpin (Bruker BioSpin GmbH, Billerica, MA, USA) routines. Moreover, the 1D ^{1}H NMR spectra were apodized with a weighting exponential function of 0.3 Hz with MestreNova v.11.0 (Mestrelab Research, Spain), and imported to MATLAB R2016a (The MathWorks Inc., Natick, MA, USA) as a spectral data matrix. Then, regions of 4.68–5.14 ppm (water), 7.65–7.68 ppm (chloroform), below 0.7 ppm (DSS), and above 9.5 ppm (empty) were removed [32]. After the removal of these spectral regions, the number of data points for each ^{1}H NMR spectra was 28,913. Minor resonance misalignments were corrected with *icoshift* [33] algorithm. The processed ^{1}H NMR spectra were normalized using the probabilistic quotient normalization (PQN) [34] method. An adaptive, intelligent bucketing [35] was applied on the pre-processed ^{1}H NMR spectra using the desktop NMRProcFlow tool [36]. Buckets with a SNR above 3 were selected for further analysis. The final matrix of buckets contained 109 buckets.

Multivariate curve resolution-alternating least squares

In the application of the multivariate curve resolution-alternating least squares (MCR-ALS) method, a data matrix, **D**, is decomposed by a bilinear model into the product of two factor matrices which describe the set of concentrations, **C**, and spectral profiles, $\mathbf{S}^{\mathbf{T}}$, of the constituents of the analyzed samples (Fig. 1).

For 1D ^{1}H NMR data, matrix **C** contains the set of relative concentrations corresponding to the $\mathbf{S}^{\mathbf{T}}$ proton resonances in the 1D ^{1}H NMR spectra of the analyzed samples (Fig. 1). The number of resolved profiles in **C** and $\mathbf{S}^{\mathbf{T}}$ (or number of components) needs to be initially determined. In this case, the optimal number of components was assessed by singular value decomposition analysis (see ESM) [37]. During the MCR-ALS bilinear decomposition, constraints should be proposed to drive the resolution of **C** and $\mathbf{S}^{\mathbf{T}}$ factor matrices to be in agreement with the previous physical knowledge about the physical nature of the profiles of the chemical constituents of the studied system. In this study, only non-negativity constraints were applied for the resolution of both **C** and $\mathbf{S}^{\mathbf{T}}$ matrices.

The dataset consisted of 20 1D ^{1}H NMR spectra (in rows) with 28,913 δ_H values (in columns). This dataset was split into 98 spectral segments (step B1 in Fig. 2). Every spectral window was set to contain between one and four resonances, and limits (beginning and end) of these spectral segments were chosen at chemical shifts whose intensities were at the noise level with the aim not to split any resonance between two segments.

Every window was then examined separately with MCR-ALS (step B2 in Fig. 2). MCR-ALS analyses were performed using the MCR-ALS GUI 2.0 [38] toolbox under MATLAB 2016b (The MathWorks Inc. Natick, MA, USA) environment.

A total number of 98 spectral segments were resolved by MCR-ALS analysis using between 1 and 4 components for each window. Thus, from these analyses, 98 **C** and 98 $\mathbf{S}^{\mathbf{T}}$ matrices were obtained, and a total number of 222 components were obtained in the analysis of all the spectral segments of the 1D ^{1}H NMR dataset. The 98 **C** matrices were then combined row-wisely into a super-augmented **C** matrix (from now on, this is called the "**superC**" matrix) with 20 rows (one per sample) and 222 columns (one per component) (step B3 in Fig. 2).

The arrangement and indexing of these components in the **superC** matrix was performed using the same convention as when plotting NMR spectra: lower column indexes corresponded to MCR-ALS components associated with spectral features at lower chemical shifts, while components with spectral features at higher chemical shifts were given high column indexes. The **superC** matrix was then further investigated with other multivariate data analysis chemometric methods, such as PCA or PLS-DA (step B4 in Fig. 2), and the relevant components were assigned by the inspection of their corresponding spectral features (step B5 in Fig. 2).

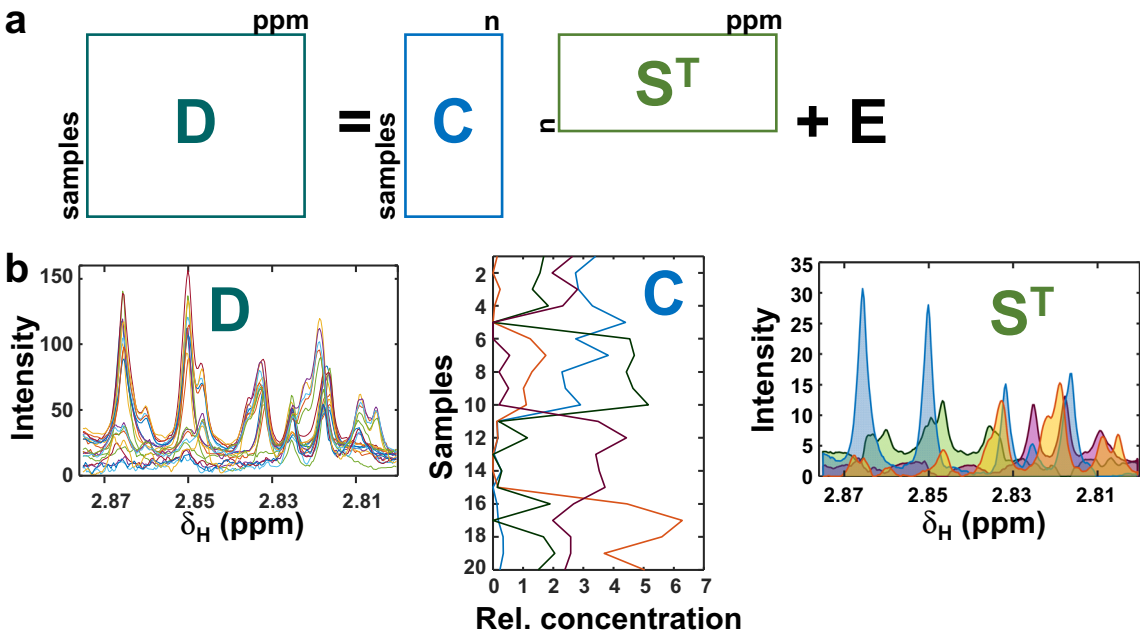

Fig. 1 MCR-ALS. (**A**) Matrix representation of the MCR-ALS bilinear decomposition. (**B**) The 1D ^{1}H NMR dataset, **D**, can be decomposed into a set of concentrations, **C**, and spectral resonances, $\mathbf{S}^{\mathbf{T}}$, from *n* metabolites. The plot of **C** profiles shows the evolution of the resolved concentration profiles for the 3 metabolites ($n = 4$) over the 20 analyzed samples. The plot of $\mathbf{S}^{\mathbf{T}}$ spectra shows the resonances from the 4 metabolites

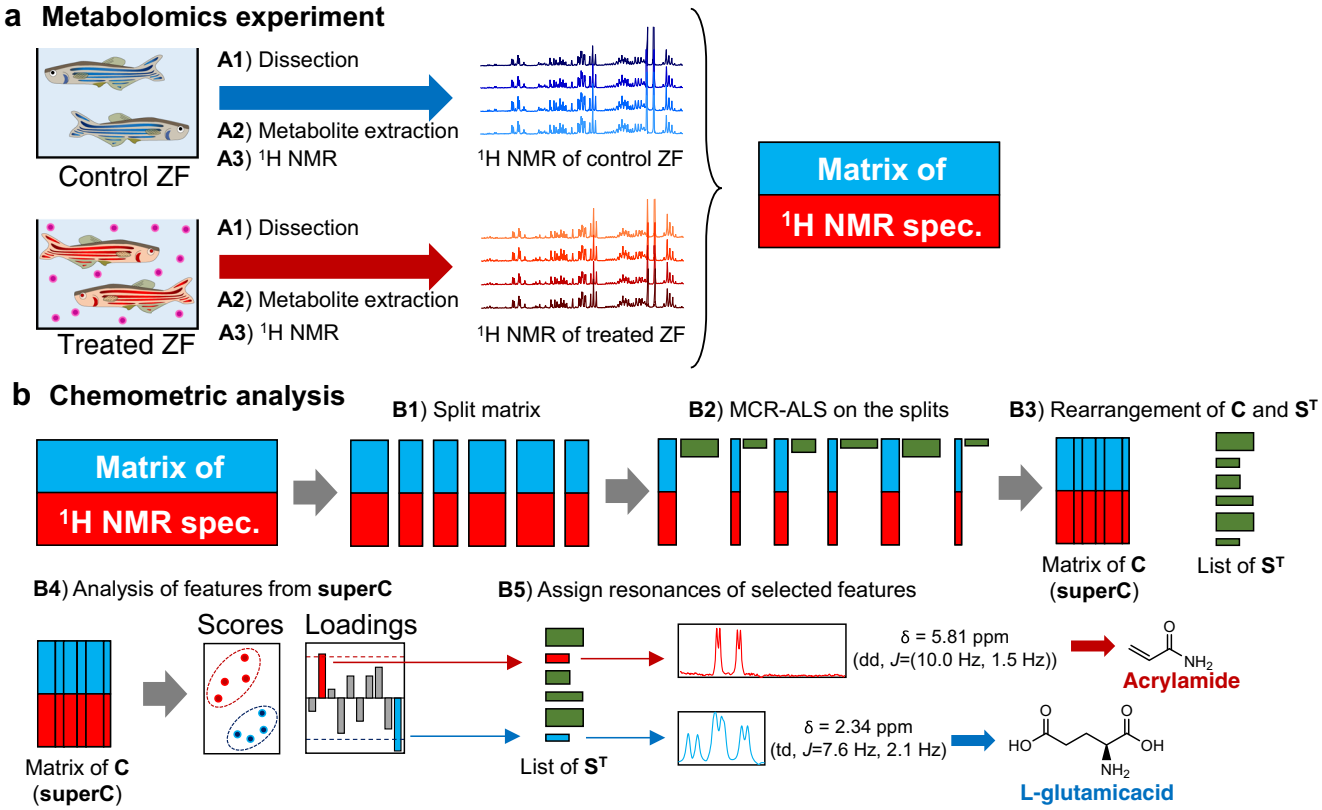

Fig. 2 Proposed workflow for the NMR metabolomics experiment

Principal component analysis

PCA [11] was applied to the raw 1D ^{1}H NMR spectral data, to the bucketed data, and to the **superC** data matrices. Prior to PCA analyses, all datasets were mean-centered. PCA results were validated using a venetian blinds cross-validation method. The analyses were carried out using the PLS toolbox 7.8.0 (Eigenvector Research Inc.).

ANOVA-simultaneous component analysis

ANOVA-simultaneous component analysis (ASCA) is a multivariate data analysis method that combines the power of ANOVA to separate variance sources with the advantages of simultaneous component analysis (SCA) to the modeling of the individual separate effect matrices [12]. In this study, ASCA was employed to untangle the metabolic response linked to two effects of factors, the type of tissue and the ACR exposure (treatment), according to the model defined in Eq. 1:

$$\mathbf{X} = \mathbf{X}_{\mathbf{Tissue}} + \mathbf{X}_{\mathbf{treatment}} + \mathbf{X}_{\mathbf{Treatment\_Tissue}} + \mathbf{E} \quad (1)$$

where $\mathbf{X}$ corresponds to the **superC** matrix, $\mathbf{X}_{\mathbf{Tissue}}$ and $\mathbf{X}_{\mathbf{treatment}}$ are descriptive of the separated metabolic fingerprints that are either determined by the two types of tissue or by the ACR exposure treatment, and $\mathbf{X}_{\mathbf{Treatment\_Tissue}}$ is giving the metabolic fingerprint showing the possible specific interaction between these two factors (type of tissue and ACR exposure) for each set of samples. From the analysis of each of the three decomposed matrices, it was possible to investigate whether significant differences existed due to the type of tissue, to the ACR treatment only, or to the two factors at the same time (e.g., if ACR exposure causes different metabolic response depending on the type of tissue). The statistical significance of the effects of each factor and their interaction was evaluated using a permutation test (with 1000 permutations) [39]. Before the ASCA analysis, the **superC** matrix was mean-centered. ASCA analysis was performed using the PLS toolbox 7.8.0 (Eigenvector Research Inc.).

Partial least squares-discriminant analysis

In this study, 3 different PLS-DA analyses were performed. One first PLS-DA analysis (PLS-DA #1 in Fig. 3) was used to discriminate the two analyzed tissues (brain and muscle) at control conditions and identify the metabolites responsible of the tissue discrimination. The two other PLS-DA analyses (PLS-DA #2 and #3 in Fig. 3) were applied to discriminate between samples from control and exposed ZF for each of the two tissues, respectively, and to identify the metabolites associated to their separation. For each of the 3 PLS-DA

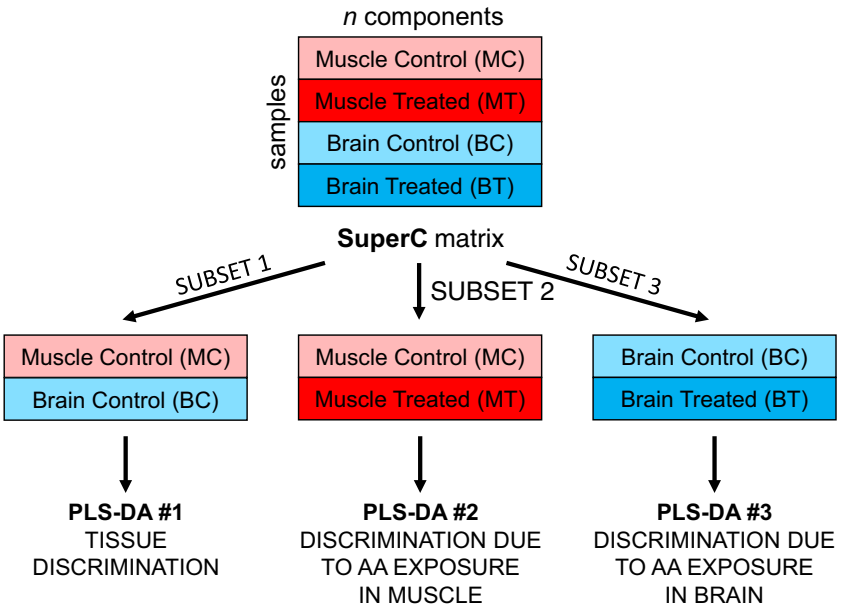

Fig. 3 PLS-DA on subsets from **SuperC** matrix

analyses, the studied dataset had the concentration profiles from the **superC** matrix corresponding to the subgroup of samples to be compared. Before each PLS-DA analysis, the data were auto-scaled. PLS-DA analyses were validated with a venetian blind type of cross-validation. In these PLS-DA analyses, the variables with VIP values ≥ 1 were selected as discriminant. The statistical significance of the selected variables was confirmed with one-way ANOVA. The PLS-DA analyses were carried out using the PLS toolbox 7.8.0 (Eigenvector Research Inc.).

Spectral resonances integration with BATMAN

Results obtained in the PLS-DA analyses were verified by manual integration of the individual resonances using the BATMAN R package [24] and by the posterior assessment of their statistical significance using one-way ANOVA.

Results and discussion

Exploratory analysis of the superC matrix by PCA

The PCA analysis of the **superC** matrix containing the relative concentration values of the 222 resolved features in the 20 zebrafish (muscle or brain) samples revealed an orthogonal distribution of the metabolism of this organism according to the type of tissue and to the ACR exposure (Fig. 4A). Specifically, the samples were separated according to the two types of tissues (muscle and brain) in PC1 (63.43% of the explained variance), while PC2 (9.99%) allowed the separation according to the ACR exposure.

Whereas this orthogonal score discrimination was clearly detected in the PCA of the **superC** matrix, it could not be observed when the same analysis was performed on the original 1D ^{1}H NMR dataset (Fig. 4B, C) nor on the analysis of the bucketed data (Fig. 4D, E). In fact, results from the PCA analyses in Fig. 4B (spectral data) and Fig. 4C (bucket data) erroneously suggested that the metabolic response of the ACR treatment could be neglected when compared with the metabolic variance observed between the two tissues. In Fig. 4B, the component that separates between the two tissues (PC1) accounted for the 97.12% of the explained variance of the dataset. In Fig. 4D, the same separation was achieved with 99.74% of the explained variance. Thus, in both cases, the remaining variance that could potentially describe the tissue effect was much lower.

In order to separate samples due to the ACR treatment in the analysis of the spectral data, it is needed to examine the dataset at the fifth PC (PC5, Fig. 4C), which only accounted for the 0.18% of the explained variance of the dataset. For the bucketed data, this is achieved in PC3 (Fig. 4E, 0.06% of the explained variance). However, the cluster separations were not perfect either.

We attribute the divergence in the results shown in Fig. 4 to the different reliability of the examined data in the two cases. While each of the variables from the **superC** matrix is descriptive of pure resolved resonances, every single variable from the 1D ^{1}H NMR spectral matrix can be representative of multiple resonances at the same time (if signal overlap occurs) in addition to the instrumental noise. Due to the high signal overlap and instrumental noise, the contribution of the individual resonances within the corresponding variables is diluted. For this reason, sample differences would be less efficiently distinguished when resonances are severely overlapped, as in 1D ^{1}H NMR raw data, than when each variable is representing

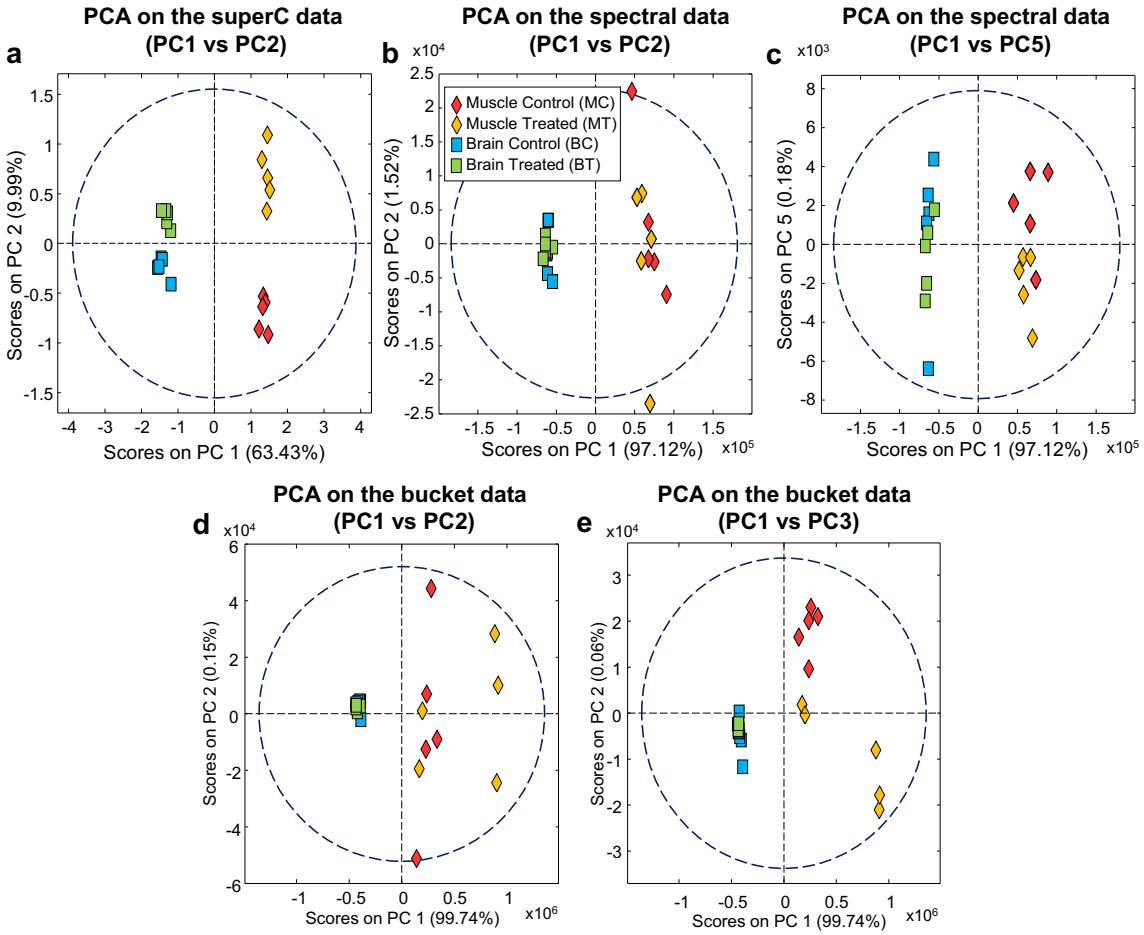

Fig. 4 (**A**) PCA scores (PC1 vs PC2) on the **superC** matrix. (**B**, **C**) PCA scores on the 1D ^{1}H NMR spectral matrix. (**B**) PC1 vs PC2. **C.** PC1 vs PC5. (**D**, **E**) PCA scores on the bucket matrix. (**D**) PC1 vs PC2. (**E**) PC1 vs PC3

one single resolved resonance, as in the **superC** matrix. For the case of the bucketed data, despite the pre-processing included a step for filtering the variables' descriptive of noise, the sample distribution did not improve, pointing out that untangling the overlapped resonances is required to obtain the best performance. Our results agree with [25], where it was proved that MCR-ALS used as a pre-processing method instead of other approaches resulted on the lowest variability within group and the largest between groups, therefore maximizing the sample clustering in the PCA.

On the other hand, the 1D ^{1}H NMR data contains a significant amount of variables only descriptive of instrumental noise. This is not the case for the **superC** matrix, since most of the instrumental noise is set apart in the residual matrix (**E** in Fig. 1) during the MCR-ALS analysis. This is the reason why the PCA analysis of the **superC** matrix was much less affected by the noise contributions than the analysis of raw 1D ^{1}H NMR data.

Therefore, for a comprehensive untargeted NMR metabolomics analysis, we propose complementing the traditional chemometric analyses of the 1D ^{1}H NMR spectral raw data or buckets with the more refined analysis of the MCR-ALS resolved concentrations.

Tissue specificity of ACR exposure in ZF

As shown above, the PCA analysis of the **superC** matrix detected two main sources of data variance, one related to the tissue (PC1 in Fig. 4B) and another to the ACR exposure (PC2 in Fig. 4B). In order to determine the significance of these two factors, and to evaluate whether there exists any interaction between the two factors, the ASCA method was applied and the results are shown in Fig. 5.

The effects of the two studied factors are clearly visualized in Fig. 5. As in the PCA analysis, metabolic differences where more significant for the tissue type (60.74%, Fig. 5A) rather than for the ACR exposure itself (8.05%, Fig. 5B). In addition, the possible interaction between these two factors (type of tissue and ACR exposure) can also be detected accounting for a 4.56% of the total data variance (Fig. 5C).

These results were confirmed after the calculation of the statistical significances of the effects for the two factors and of their interaction using a permutation test [39]. In all the cases, the significance of the tests were high (p values = 0.001).

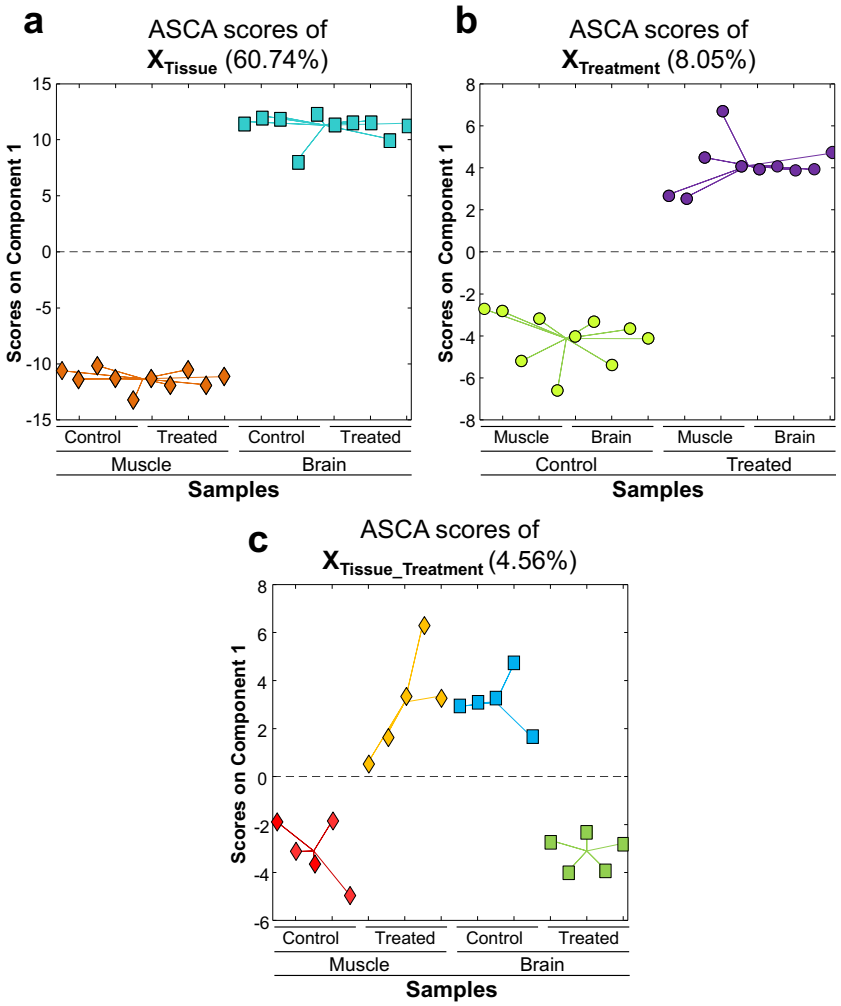

Fig. 5 ASCA scores of (**A**) $\mathbf{X}_{\mathbf{Tissue}}$, (**B**) $\mathbf{X}_{\mathbf{treatment}}$, and (**C**) $\mathbf{X}_{\mathbf{Treatment\_Tissue}}$. In the legends, MC and MT are muscle extracts from control and ACR-treated ZF; BC and BT are the corresponding brain extract counterparts

Metabolic discrimination of the tissues

In order to select the metabolite characteristic for the two tissues, a PLS-DA analysis was performed comparing the 222 features for the two types of tissues (muscle and brain) of control samples (PLS-DA #1 in Fig. 3). As a result, 151 features were found to discriminate between muscle and brain. By looking at the spectral profiles associated with these features ($\mathbf{S}^{\mathbf{T}}$ in Fig. 1, and step B5 in Fig. 2) studying the spectroscopic parameters (δ_H, *J*, multiplicity) that define these resonances, and after consulting these parameters on a NMR spectral library (such as HMDB [40] or BRMB [41]), it was possible to assign some of them to metabolites (see the list of assigned features and Table S1 in ESM). It must be noted that this level of spectroscopic investigation cannot be performed by more conventional approaches. In these approaches, resonances are assigned according to their chemical shifts values (δ), whereas the other associated spectroscopic parameters cannot be easily determined due to the resonance overlapping. Moreover, since the data were auto-scaled, even the slightest variations among tissues can be detected. The resonances most distinctive for brain extracts were for γ-aminobutyric acid (GABA), acetic acid, *N*-acetyl-L-aspartic acid, L-glutamic acid, citric acid, carnosine, phosphothreonine, and some other resonances from lipidic compounds. On the other hand, creatine, L-histidine, taurine, glycerophosphorylcholine, L-alanine, fumaric acid, L-tyrosine, D-glucose, glycogen, UDP, and AXP were significantly more concentrated in muscle extracts than in brain extracts. In other words, brain sample characteristic compounds (compared with muscle samples) include neurotransmitters, neuronal osmolytes, and fats; in muscle, the characteristic metabolites (compared with brain samples) were those involved in its functioning, structure, redox, and energy metabolism [42].

Metabolic fingerprint due to ACR treatment

The metabolic biomarkers of ACR exposure in muscle and in brain were determined by performing a PLS-DA analysis on

each separated data subset of the **superC** matrix by comparing the control and treated samples related to the same tissue (PLS-DA #2 and #3 in Fig. 3). Resulting from these two PLS-DA analyses, several features were detected to be significant in each of the two tissues: 82 features in the muscle subset and 45 features in the brain subset. After the spectral investigation of these significant features, we concluded that in the muscle tissue samples, ACR exposure caused an increase of the pools of the following metabolites: ACR, L-leucine, L-isoleucine, L-valine, L-lysine, GABA, phosphothreonine, methionine sulfoxide, D-glucose, UDP, L-tyrosine, and L-phenylalanine (Fig. 6). In contrast, the pools of L-alanine, L-glutamic acid, glutathione, and glycerol were diminished due to ACR exposure. In the brain tissue samples, the ACR exposure increased significantly the pools of free ACR, *N*-acetyl-S-(carbamoylethyl)-L-cysteine (AAMA), and methionine sulfoxide. Conversely, the pools of carnosine, L-glutamic acid, glutathione, and betaine were reduced due to this ACR treatment (Fig. 6).

All these metabolic alterations reflect that ACR depleted the pools of intracellular glutathione in brain and muscle by reacting with it. As a product of this reaction, in brain, ACR was accumulated as *N*-acetyl-S-(carbamoylethyl)-L-cysteine (AAMA). In addition to AAMA formation, the substantial decrease of the glutathione levels was a direct consequence of the inhibition by ACR of the gamma-glutamylcysteine ligase, which is involved in the synthesis of de novo glutathione [43]. Glutathione and carnosine depletion caused a reduction of the organism oxidative stress defenses, which elicited the appearance of new oxidized species, such as methionine sulfoxide pools. The accumulation of ACR and AAMA, together with the prevailing oxidative stress, could explain the neurotoxicity of ACR at a metabolic level in the brain.

In muscle, this enhanced oxidative stress may have caused the decrease of the amino acid pools as well, probably to be used for protein synthesis, in agreement with the activation of protein synthesis of thioredoxins by ACR exposure [32]. Other metabolic affectations observed in muscle are the modification of the lipid membranes and the activation of pathways linked to energy metabolism. Related to the latter, the consumption of amino acids convertible to TCA intermediates and the accumulation of D-glucose and UDP (involved in glycogen metabolism) were detected.

Biomarker discovery

Spectral features $\mathbf{S}^{\mathbf{T}}$ associated with the same metabolite are strongly correlated, as they can be observed for instance in the heatmap plots shown in Fig. S2 from ESM. Thus, this fact can be used as a guide during the resonance assignment step. However, not all the selected features could be assigned to metabolites because their resonances are from compounds not included in metabolomics libraries [44, 45]. In order to provide a more comprehensive description of the effects of ACR exposition on zebrafish, we have investigated unassigned features that increased due to the ACR exposure. In particular, we focused on the features 78, 79, and 96 shown in Fig. 7C. These 3 features were significantly important in muscle extracts of ACR-treated zebrafish (Fig. 7D–F). The reconstructed 1D ^{1}H NMR spectrum with these 3 features revealed two symmetric multiplets, centered at 2.60 and at 2.86 ppm, that resemble two double triplets (Fig. 7C). This observed coupling is in agreement with the spin system of a $RSCH_2CH_2R'$ moiety. Since this spin system is also found in AAMA, we propose that this molecule is a reaction product from ACR. In humans, several ACR-derived compounds in addition to AAMA have been detected after oral exposition to ACR [46].

Hence, in this work, we detected two metabolites resulting from the ACR transformation inside the organism, AAMA, and the "ACR-derivative." AAMA was found more prevalent in the brain of treated ZF, while the "ACR-derivative" was more prominent in the corresponding muscle extracts. This tissue specificity may reflect the high reactivity of ACR in the two metabolically different tissues.

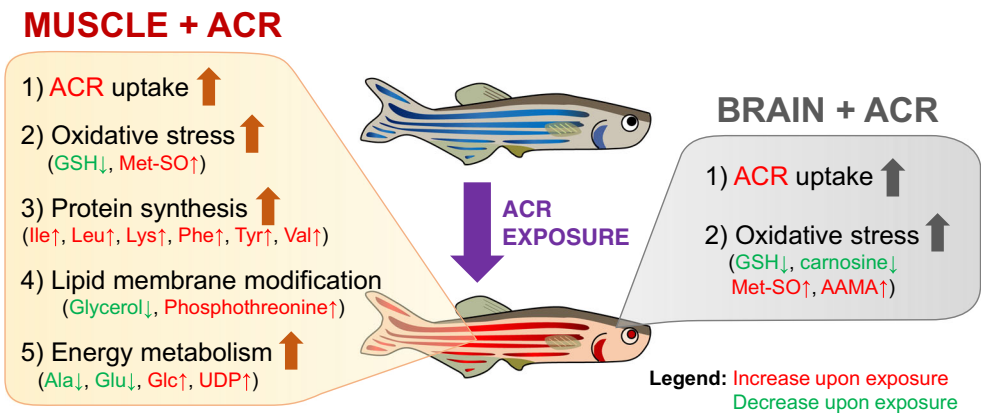

Fig. 6 Schematic representation of the effects of ACR on zebrafish muscle and brain metabolomes

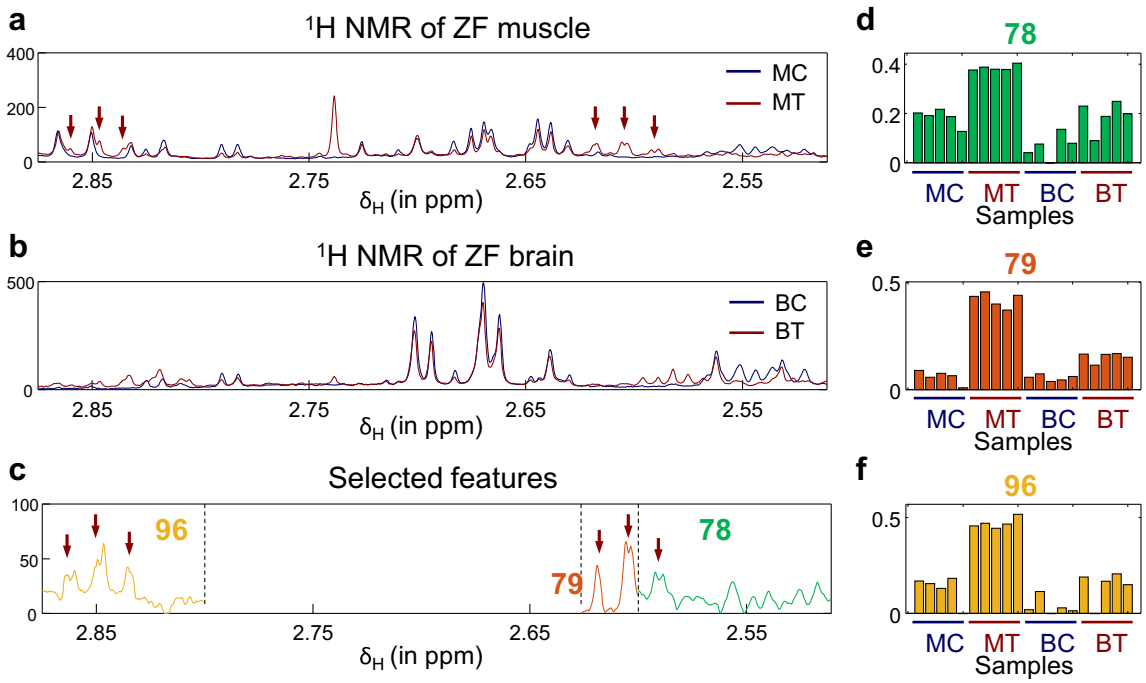

Fig. 7 (**A**, **B**) Representative 1D ^{1}H NMR spectra of muscle (**A**) and (**B**) brain extracts. 1D ^{1}H NMR spectra from control ZF are in blue, while spectra from treated ZF are colored in red. (**C**) Reconstructed 1D ^{1}H NMR with features 78, 79, and 96. (**D**–**F**) Concentration profiles for features 78, 79, and 96. Arrows point the resonances from the two triple doublets centered at 2.60 and 2.86 ppm

Method considerations

The proposed MCR-ALS-based chemometric methodological pipeline has been proven to be very useful for both metabolic fingerprint and biomarker discovery. Noteworthy, some considerations regarding this pipeline need to be made.

First, one of the most important steps in this pipeline is to set the limits of the NMR resonances in every window to be analyzed by MCR-ALS. In this study, we have chosen the window limits by visual inspection (see methods), relying on the user's spectroscopic knowledge avoiding to split resonances between two segments. However, more automatable alternative approaches could be designed. For instance, the 1D ^{1}H NMR dataset could be bucketed with a procedure that does not split resonances with complex splitting patterns in different buckets, and MCR-ALS can be applied separately then to every bucket.

Second consideration is that MCR-ALS is a robust chemometric method which estimates the concentrations (**C**) associated with the same resonances ($\mathbf{S}^{\mathbf{T}}$) in all the investigated samples. However, this chemometric approach should not be used with an NMR dataset presenting inconsistencies between spectra (e.g., derived from incomplete shimming or phasing). Apart from that, the MCR-ALS can be used with 1D ^{1}H NMR data from any type of samples (e.g., extracts, serum, CSF) as long as the resonances are aligned across all the spectra. In metabolomics studies in the lab, where robust sample preparation methods are used [47], chemical shifting variations among the spectra can be usually corrected with simple resonance alignment methods, such as *icoshift* [33]. Also, for samples with complex background signals, such as serum samples, it may be recommended to add extra components in the MCR-ALS analyses to account for these background signals.

And as a third consideration, resonances presenting the same variance across samples and that are found within the same window will be resolved together, even if they are from different metabolites. Although this particularity can be seen as a limitation of the method, the resolved components will still be very useful and informative. The resolved spectral data will be much simpler than the original data (as all the other resonances presenting different variance will appear in other components) facilitating the resonance assignment. Moreover, the grouping of resonances in the same component suggests that the corresponding compounds are metabolically related (since they present the same metabolic variance), resulting very useful for the biological interpretation.

Conclusions

The chemometric methodological pipeline presented here takes advantage of the rapid and powerful performance of the MCR-ALS chemometrics method to blindly resolve the concentrations and resonances from untangled metabolites, which allows the identification of the spectral features characteristic of the metabolomics system under study.

In this work, we have proved that this methodology simplifies the metabolic analysis of 1D ^{1}H NMR spectra from complex biological extracts and that it can even be used for biomarker discovery. Moreover, the presented methodology is more sensitive to small metabolic variations than other chemometric methods, which do not take into account the prominent resonance overlapping in 1D ^{1}H NMR data.

In consequence, the proposed methodology can accelerate the discovery of unknown metabolites, and it has the potential to facilitate the study of large NMR datasets, such as those analyzed in cohort metabolomics studies [48].

Code availability The MCR-ALS GUI can be downloaded from www.mcrals.info.

Authors' contribution B Piña and D Raldúa designed the experiments. M Casado and Y Pérez performed the experiments. F Puig-Castellví performed the chemometric analysis. All authors have contributed in the manuscript writing and have given approval to its final version.

Funding information The research leading to these results has received funding from the NATO SfP project MD.SFPP 984777.

Data availability The data is available upon request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval All procedures were approved by the Institutional Animal Care and Use Committees at the CID-CSIC and conducted in accordance with the institutional guidelines under a license from the local government (agreement number 9027).

Consent to participate The authors gave their consent to participate in the research.

Consent for publication The authors give formal consent for publication of the work.

References

1. Boiteau RM, Hoyt DW, Nicora CD, Kinmonth-Schultz HA, Ward JK, Bingol K. Structure elucidation of unknown metabolites in metabolomics by combined NMR and MS/MS prediction. Metabolites. 2018;8(1):8.
2. Dona AC, Kyriakides M, Scott F, Shephard EA, Varshavi D, Veselkov K, et al. A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Comput Struct Biotechnol. 2016;14:135–53.
3. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17:451.
4. Bingol K. Recent advances in targeted and untargeted metabolomics by NMR and MS/NMR methods. High-Throughput. 2018;7(2):9.
5. Cappello T, Maisano M, Mauceri A, Fasulo S. ^{1}H NMR-based metabolomics investigation on the effects of petrochemical contamination in posterior adductor muscles of caged mussel *Mytilus galloprovincialis*. Ecotoxicol Environ Saf. 2017;142(Supplement C):417–22.
6. Nagato EG, D'eon JC, Lankadurai BP, Poirier DG, Reiner EJ, Simpson AJ, et al. ^{1}H NMR-based metabolomics investigation of Daphnia magna responses to sub-lethal exposure to arsenic, copper and lithium. Chemosphere. 2013;93(2):331–7.
7. Puig-Castellví F, Pérez Y, Piña B, Tauler R, Alfonso I. Comparative analysis of ^{1}H NMR and ^{1}H–^{13}C HSQC NMR metabolomics to understand the effects of medium composition in yeast growth. Anal Chem. 2018;90(21):12422–30.
8. Tomassini A, Vitalone A, Marini F, Praticò G, Sciubba F, Bevilacqua M, et al. ^{1}H NMR-based urinary metabolic profiling reveals changes in nicotinamide pathway intermediates due to postnatal stress model in rat. J Proteome Res. 2014;13(12):5848–59.
9. Ruan LY, Fan JT, Hong W, Zhao H, Li MH, Jiang L, et al. Isoniazid-induced hepatotoxicity and neurotoxicity in rats investigated by ^{1}H NMR based metabolomics approach. Toxicol Lett. 2018;295:256–69.
10. Fathi F, Oskouie AA, Tafazzoli M, Naderi N, Sohrabzedeh K, Fathi S, et al. Metabonomics based NMR in Crohn's disease applying PLS-DA. Gastroenterol Hepatol Bed Bench. 2013;6(Suppl 1):S82–6.
11. Bro R, Smilde AK. Principal component analysis. Anal Methods. 2014;6(9):2812–31.
12. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers RJAN, van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics. 2005;21(13):3043–8.
13. Barker M, Rayens W. Partial least squares for discrimination. J Chemom. 2003;17(3):166–73.
14. Tauler R, Kowalski B, Fleming S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial-process. Anal Chem. 1993;65.
15. Ebrahimi P, Larsen FH, Jensen HM, Vogensen FK, Engelsen SB. Real-time metabolomic analysis of lactic acid bacteria as monitored by in vitro NMR and chemometrics. Metabolomics. 2016;12(4):77.
16. Winning H, Larsen FH, Bro R, Engelsen SB. Quantitative analysis of NMR spectra with chemometrics. J Magn Reson. 2008;1:26–32.
17. Abdollahi H, Tauler R. Uniqueness and rotation ambiguities in multivariate curve resolution methods. Chemom Intell Lab Syst. 2011;108(2):100–11.
18. Puig-Castellví F, Alfonso I, Tauler R. Untargeted assignment and automatic integration of ^{1}H NMR metabolomic datasets using a multivariate curve resolution approach. Anal Chim Acta. 2017;964(Supplement C):55–66.
19. Karakach TK, Knight R, Lenz EM, Viant MR, Walter JA. Analysis of time course ^{1}H NMR metabolomics data by multivariate curve resolution. Magn Reson Chem. 2009;47(S1):S105–17.
20. Montoliu I, Martin FPJ, Collino S, Rezzi S, Kochhar S. Multivariate modeling strategy for intercompartmental analysis of tissue and plasma ^{1}H NMR spectrotypes. J Proteome Res. 2009;8(5):2397–406.
21. Röhnisch HE, Eriksson J, Müllner E, Agback P, Sandström C, Moazzami AA. AQuA: an automated quantification algorithm for high-throughput NMR-based metabolomics and its application in human plasma. Anal Chem. 2018;90(3):2095–102.
22. Tardivel PJC, Canlet C, Lefort G, Tremblay-Franco M, Debrauwer L, Concordet D, et al. ASICS: an automatic method for identification and quantification of metabolites in complex 1D ^{1}H NMR spectra. Metabolomics. 2017;13(10):109.
23. Cañueto D, Gómez J, Salek RM, Correig X, Cañellas N. rDolphin: a GUI R package for proficient automatic profiling of 1D ^{1}H-NMR spectra of study datasets. Metabolomics. 2018;14(3):24.

24. Hao J, Liebeke M, Astle W, De Iorio M, Bundy JG, Ebbels TMD. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc. 2014;9(6):1416–27.
25. Khakimov B, Mobaraki N, Trimigno A, Aru V, Engelsen SB. Signature Mapping (SigMa): an efficient approach for processing complex human urine 1H NMR metabolomics data. Anal. Chim. Acta, 2020, In press, 1–10. https://doi.org/10.1016/j.aca.2020.02.025.
26. Faria M, Ziv T, Gómez-Canela C, Ben-Lulu S, Prats E, Novoa-Luna KA, et al. Acrylamide acute neurotoxicity in adult zebrafish. Sci Rep. 2018;8:1–14.
27. Dearfield KL, Abernathy CO, Ottley MS, Brantner JH, Hayes PF. Acrylamide: its metabolism, developmental and reproductive effects, genotoxicity, and carcinogenicity. Mutat Res-Rev Genet. 1988;195(1):45–77.
28. Tareke E, Rydberg P, Karlsson P, Eriksson S, Törnqvist M. Analysis of acrylamide, a carcinogen formed in heated foodstuffs. J Agric Food Chem. 2002;50(17):4998–5006.
29. Garland TO, Patterson MWH. Six cases of acrylamide poisoning. Br Med J. 1967;4:134–8.
30. Tepe Y, Çebi A. Acrylamide in environmental water: a review on sources, exposure, and public health risks. Expos Health. 2017.
31. Duke TJ, Ruestow PS, Marsh GM. The influence of demographic, physical, behavioral, and dietary factors on hemoglobin adduct levels of acrylamide and glycidamide in the general U.S. population. Crit Rev Food Sci Nutr. 2018;58(5):700–10.
32. Raldúa D, Casado M, Prats E, Faria M, Puig-Castellví F, Pérez Y, et al. Targeting redox metabolism: the perfect storm induced by acrylamide poisoning in the brain. Sci Rep. 2020;10:312.
33. Savorani F, Tomasi G, Engelsen SB. Icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson. 2010;202(2):190–202.
34. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR Metabonomics. Anal Chem. 2006;78(13):4281–90.
35. de Meyer T, Sinnaeve D, van Gasse B, Tsiporkova E, Rietzschel E, de Buyzere M, et al. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. Anal Chem. 2008;80(10):3783–90.
36. Jacob D, Deborde C, Lefebvre M, Maucourt M, Moing A. NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. Metabolomics. 2017;13:36.
37. Abdi H. Singular Value decomposition (SVD) and generalized singular value decomposition (GSVD). In: Salkind NJ, editor. Encyclopedia of measurement and statistics. SAGE Publications: 2007;907–912.
38. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: new features and applications. Chemom Intell Lab Syst. 2015;140:1–12.
39. Zwanenburg G, Hoefsloot HCJ, Westerhuis JA, Jansen JJ, Smilde AK. ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison. J Chemom. 2011;25:561–7.
40. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0—the human metabolome database in 2013. Nucleic Acids Res. 2013;41(D1):D801–7.
41. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, et al. BioMagResBank. Nucleic Acids Res. 2008;36(suppl_1):D402–8.
42. Shi M, Ellingsen Ø, Bathen TF, Høydal MA, Koch LG, Britton SL, et al. Skeletal muscle metabolism in rats with low and high intrinsic aerobic capacity: effect of aging and exercise training. PLoS One. 2018;13(12):e0208703.
43. Faria M, Prats E, Gómez-Canela C, Hsu C, Arick MA II, Bedrossiantz J, et al. Therapeutic potential of N-acetylcysteine in acrylamide acute neurotoxicity in adult zebrafish. Sci Rep. 2019;9:16467.
44. da Silva RR, Dorrestein PC, Quinn RA. Illuminating the dark matter in metabolomics. PNAS. 2015;112(41):12549–50.
45. Jones OAH. Illuminating the dark metabolome to advance the molecular characterisation of biological systems. Metabolomics. 2018;14(8):101.
46. Kopp EK, Dekant W. Toxicokinetics of acrylamide in rats and humans following single oral administration of low doses. Toxicol Appl Pharmacol. 2009;235(2):135–42.
47. McHugh CE, Flott TL, Schooff CR, Smiley Z, Puskarich MA, Myers DD, et al. Rapid, reproducible, quantifiable NMR metabolomics: methanol and methanol: chloroform precipitation for removal of macromolecules in serum and whole blood. Metabolites. 2018;8(4):93.
48. Trivedi DK, Hollywood KA, Goodacre R. Metabolomics for the masses: the future of metabolomics in a personalized world. New Horiz Trans Med. 2017;3(6):294–305.