Non-negative matrix factorization of SWATH DIA data improves global metabolite identification

Diana Karaki, Sylvain Dechaumet, Annelaure Damont, Benoit Colsch, François Fenaille, Antoine Souloumiac, Etienne A Thévenot

Non-Negative Matrix Factorization of SWATH DIA Data Improves Global Metabolite Identification

Diana Karaki*, Sylvain Dechaumet*, Annelaure Damont*, Benoit Colsch*, François Fenaille*,
Antoine Souloumiac†, and Etienne A. Thévenot*
*Département Médicaments et Technologies pour la Santé (DMTS), MetaboHUB
Université Paris-Saclay, CEA, INRAE, 91191 Gif-sur-Yvette, France
†CEA-List, Université Paris-Saclay, 91120 Palaiseau, France
firstname.lastname@cea.fr

***Abstract*—Liquid chromatography coupled to High-Resolution Mass Spectrometry (LC-HRMS) is the most widely used approach for the global detection of small molecules in biological samples (metabolomics). In complement to such MS1 data, structural identification of metabolites implies the acquisition of fragmentation spectra by performing tandem mass spectrometry (MS2) experiments. To achieve both global detection and identification in a single run, the recently introduced acquisition mode called Sequential Window Acquisition of all THeoretical fragment ions (SWATH-type) Data Independent Acquisition (DIA) alternates MS1 detection and MS2 analysis of large and continuous m/z windows. The resulting MS2 data, however, contain a mixture of fragment ions originating from different precursor ions. To deconvolve these data and reconstruct pure individual MS2 spectra, the few existing software rely on determining a peak shape for each precursor ion. Such a strategy, however, may fail to separate co-eluting compounds. Here, we show how sparse non-negative matrix factorization (NMF) can separate pure spectral components successfully. We developed an end-to-end workflow called DIA-NMF to process SWATH DIA files, identify the detected compounds, and showed that it outperforms the reference algorithms MS-DIAL and DecoMetDIA, especially in the case of low-intensity or co-eluting compounds. Importantly, the reconstructed spectra include all the MS1 and MS2 ions related to the sought compounds and thus provide enriched chemical information that facilitates interpretation and identification.**

***Index Terms*—non-negative matrix factorization, sparsity, mass spectrometry, SWATH DIA, metabolomics**

I. Introduction

SWATH DIA for comprehensive metabolite detection and identification

Metabolomics (the large-scale analysis of small molecules in a biological sample) is a powerful approach to discover biomarkers for precision medicine [1]. Liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) is the most sensitive technique for the global detection of small compounds [2]. The sample first enters the LC, where the molecules are separated according to their physico-chemical properties. After desorption and ionization of the molecules, the mass-to-charge ratios (*m/z*) of the ionized molecules are then measured, generally in a few hundred milliseconds (depending on the mass analyzer). The LC-MS file, therefore, consists of a succession of mass spectra (MS1), each being a list of (*m/z*,*intensity*) ordered pairs, acquired at specific retention times (*rt*). Aligning and concatenating the spectra enables building each detected ion's elution profile (i.e., its Extracted Ion Chromatogram, or EIC).

Structural identification of the compounds detected by LC-HRMS remains a challenge in metabolomics due to the huge chemical diversity of metabolites. Tandem mass spectrometry (MS/MS or MS2), which analyzes the fragmentation pattern of a molecule, is the method of choice to gain insight into the compound structure: in the most common conditions, ions selected by the first mass analyzer are fragmented in a collision cell, and the resulting product ions are analyzed in the second mass analyzer.

Multi-event MS1 and MS2 acquisition protocols have been developed to achieve the most comprehensive detection and identification of metabolites. In the most common *data-dependent* acquisition mode (DDA), the MS instrument automatically switches from MS1 to MS2 when ions satisfy a predefined rule (e.g., the 10 most intense ions). In this mode, precursor ions are selected using a small isolation window (typically $<$1 Da wide), which leads to high-quality and high-purity MS2 spectra for selected precursor ions. In contrast, the sequential window acquisition of all theoretical spectra *data-independent* acquisition mode (SWATH DIA) aims to select and fragment simultaneously all MS1 signals (precursor ions) within large and contiguous *m/z* isolation windows (typically 10-50 Da) which cover the whole mass range [3]. As a result, each DIA MS2 spectrum is a hybrid spectrum resulting from the fragmentation of many precursor ions within a selected m/z range. It is, therefore, mandatory to precisely extract the fragments originating from each precursor (by using the property that all fragments from one MS1 precursor have the same elution profile), i.e., to reconstruct the pure fragmentation spectrum, before it can be used for compound identification.

Motivation

The two existing approaches that address the issue of DIA data without relying on predefined spectral libraries for unmixing, namely MS-DIAL and DecoMetDIA, are based on the determination of peak models for each precursor: the mixed elution profiles are then decomposed as linear

This project is funded by the Focus Biomarqueurs programme from CEA.

combinations of these peaks [4, 5]. Such a strategy, however, is not appropriate for small peaks whose retention time is close to more intense ones (co-eluting compounds). In such cases, the peak shape defined by the algorithm may in fact encompass several analytical peaks and may not allow their proper deconvolution [6].

As an alternative to peak models, we take advantage of the blind source separation paradigm and develop here for the first time a non-negative matrix factorization (NMF) approach to unmix the MS2 spectra. NMF is widely used for analyzing high-dimensional data and feature extraction [7]. It automatically extracts meaningful features from a non-negative linear mixture. It has been recently applied to LC-MS [8] and has been shown superior to model peak algorithms for gas chromatography (GC)-MS data [6].

In this paper, we developed DIA-NMF as an end-to-end workflow for SWATH DIA data analysis, which starts with the raw data and outputs the identified molecules, and relies on NMF for mixed data factorization.

II. Non-negative matrix factorization (NMF)

Given a mixed matrix $X \in \mathbf{R}^{m\times n}$, the components' sources are mixed up in an unknown but linear way. The un-mixing model can be compactly written in this matrix form:

$$X \approx WH, \tag{1}$$

where $W \in \mathbf{R}^{m\times r}$ is the basis matrix and $H \in \mathbf{R}^{r\times n}$ is the coefficients matrix. Each column of W is the unknown spectrum/source that is not negative, whereas each row of H represents an elution profile that determines the contribution of each source, which is also non-negative. Thus, n is the number of measurements, m is the number of source samples, and r is the number of pure sources. Solving problem (1) can be written under the constrained form:

$$\underset{W,H\geqslant 0}{\operatorname{argmin}}\, \mathcal{D}(X \parallel WH) + J(W). \tag{2}$$

$\mathcal{D}$ is a divergence function, as the Euclidean distance (l_2), it measures the discrepancy between the data X and it's factorization WH. J is an optional regularization function providing prior information about the spectra.

NMF is NP-Hard [9] and can present numerous local minima. For this reason, additional constraints or prior information can help recover the sought sources. Here, we impose a sparse on the sources and their non-negativity. The numerical minimization of (2) is non-convex in both W and H, but the sub-problems are convex. Thus, the minimization of this cost function is generally solved by alternately updating W and H.

The first multiplicative NMF algorithm originated by Lee and Seung updates W and H with a weighted gradient descent [10]. The weights ensure that the gradient steps do not increase the cost function in (2) and keep W and H non-negative.

Nevertheless, the non-negativity constraint is not always sufficient to recover the sources and mixing matrix. Non-negativity and sparsity of the sources are naturally inherent in many applications, such as those using MS or Nuclear Magnetic Resonance (NMR). In the context of Blind Source Separation (BSS), sparsity has been shown to increase the diversity between the sources which greatly helps their separation [11, 12].

Puscual-Montano *et al.* introduced the non-smooth NMF (nsNMF) algorithm [13]. They claimed its superiority over previous sparse NMF variants for synthetic and real datasets.

Rapin *et al.* introduced the nGMCA^s algorithm, which aims to solve the sparse non-negative BSS [8]. This algorithm minimizes the following optimization problem:

$$\underset{W,H}{\operatorname{argmin}}\, \frac{1}{2}||X - WH||_2^2 + \lambda||W||_1 + i^+(W) + i^+(H), \tag{3}$$

where i^+ is the characteristic function of the non-negative orthant that enforces the non-negative constraints; it is applied point-wise on every entry of W and H:

$$i^+(w_{i,j}) = \begin{cases} 0 & \text{if} \quad w_{i,j} \geq 0. \\ +\infty & \text{otherwise} \end{cases} \tag{4}$$

nGMCA^s alternatively minimizes the constrained sub-problems to obtain stable solutions with the sought structure:

- Fix H, sub-problem in W is:

$$\underset{W}{\operatorname{argmin}}\, \frac{1}{2}||X - WH||_2^2 + \lambda||W||_1 + i^+(W). \tag{5}$$

  Let, $f(W) = \frac{1}{2}||X - WH||_2^2$ and $g(W) = \lambda||W||_1 + i^+(W)$. f is differentiable, convex, and its gradiant, ∇f is $L = ||HH^t||_s$ Lipschitz while g is convex, proper, and lower semi-continuous. This sub-problem can be solved by the forward-backward splitting algorithm (FBS) [14] from proximal splitting methods.
  More precisely, f is smooth; the gradient descent step is employed. However, g is not, but its proximal operator can be defined point-wise as:

$$prox_{\lambda||.||+i^+(.)}(w_{i,j}) = [Soft_\lambda(w_{i,j})]_+. \tag{6}$$

  Where $Soft$ is the soft thresholding operator and is defined as:

$$Soft_\lambda(w_{i,j}) = sign(w_{i,j})[|w_{i,j}| - \lambda]_+ \tag{7}$$

  Then, the update of W is made by:

$$W_{k+1} = prox_{\frac{\lambda}{L}} g(W_k + \frac{1}{L}\nabla f(W_k)). \tag{8}$$

- Similarly, the sub-problem of H can be solved as (3), with $g(H) = i^+(H)$. The proximal operator of this function is the projection on the positive orthant $[.]_+$. Thus, the FBS is reduced to the projected gradient algorithm.

III. Materials and Methods

A. DIA-NMF workflow

The DIA-NMF workflow processes SWATH DIA raw data files to generate a table of all detected compounds, including their intensity and identity. It is implemented in R and includes the following steps, which will be further detailed:

1) Detect and quantify all MS1 ions

2) For each detected MS1 ion p:
   a) Build the matrix X_p of the elution profiles from all candidate MS1 precursors (isotopes, adducts) and MS2 fragments of p.
   b) Factorize X_p to obtain the pure MS2 spectrum s_p.
   c) Identify the compound p (e.g. by matching s_p to a reference spectral library).

1) MS1 data processing for ion detection and quantification: The DIA raw files were converted to the *.mzML* open format using the MSConvert software [15], and then processed with the XCMS software to detect and quantify all MS1 ions (as in DecoMetDIA). At this step, each detected ion is characterized by its *m/z*, *rt*, and *intensity*.

2) MS2 data processing for ion identification: For each detected (precursor) ion p, the matrix of the elution profiles from all putative parent and fragment ions is built, and subsequently factorized by NMF to obtain the pure MS2 spectrum, which is finally matched to a reference library to identify the compound:

a) Extraction of the matrix $X^p_{m,t}$ containing the elution profiles from a candidate precursor p and its putative parent and fragment ions: The MS1 spectra acquired within a retention time window similar to p are aligned in the *m/z* dimension, and concatenated column-wise to obtain the $X^{1p}_{m,t}$ matrix. Rows (i.e., elution profiles) with constant or noise-only profiles are discarded. Similarly, the MS2 elution profiles with non-constant intensities from the *m/z* isolation window including p and within a retention time window similar to p are concatenated row-wise to obtain the $X^{2p}_{m,t}$ matrix. Finally, the $X^{1p}_{m,t}$ and $X^{2p}_{m,t}$ are concatenated row-wise to generate $X^p_{m,t}$.

Significantly, our strategy to build the *mixed* matrix $X^p_{m,t}$ differs from the DecoMetDIA and MS-DIAL approaches, which consider only the elution profiles from the MS2 fragments and the precursor ion p (i.e., not the parent ions of p). However, all parent ions and their fragments have identical elution profiles since they originate from the same compound. It is, therefore, impossible to distinguish them in a DIA acquisition. In contrast, our approach enables to comprehensively group all MS1 and MS2 ions related to the same molecule into a single spectrum to facilitate its interpretation and the identification of the metabolite.

b) Non-negative matrix factorization: To select the expected number of pure components in the mixed matrix (the factorization rank r), the *concordance* metric from [16] was used. The concordance exploits the stochastic nature of the NMF estimates, examining their stability relative to reference estimates obtained by non-negative double singular value decomposition (NNDSVD) up to some permutation.

The mixed matrix $X^p_{m,t}$ is then factorized using nGMCA[s] algorithm [8] and approximated by the product of the non-negative matrices $W^p_{m,r} \times H^p_{r,t}$. The column of $W^p_{m,r}$ with the highest intensity at the *m/z* value of the precursor p is selected as the spectrum of interest which groups all MS1 and MS2 ions related to p (i.e., all parent ions and fragments originating from the same molecule).

c) Spectral matching: The pure spectrum is restricted to the precursor and MS2 ions and matched against our in-house database of reference spectra using the mean of two classical scores [4]: 1) the inverse dot product (i.e., the dot product $\sum(I_{measured} \times I_{library})^2 / (\sum I^2_{measured} \times \sum I^2_{library})$ restricted to the fragments common to the query and reference spectra), and 2) the percentage of reference peaks found in the query spectrum (*# matched fragments / # reference fragments*). The compound from the database with the highest matching mean score above 0.3 was assigned to the MS1 precursor.

B. Experimental DIA dataset

Human plasma samples spiked with a pool of 47 chemical compounds at 7 known concentrations (from 0 to 10 ng/mL for each metabolite) were analyzed in triplicate by SWATH DIA, as described in [17]. A full scan MS event was followed by ten MS2 spectra collected from consecutive precursor ion isolation windows (20 to 50 Da each) on an Orbitrap Fusion instrument operated in the positive ionization mode. A stepped Normalized Collision Energy (NCE) was used to optimize the fragmentation: each MS2 spectrum is the mean of spectra acquired at 30%±20% NCE. MS1 and MS2 spectra were recorded at a resolution of 120,000 and 15,000 (at *m/z* 200), respectively.

C. Reference MS2 database

Our comprehensive in-house spectral library comprises 11,417 MS2 spectra from 853 pure compounds. This reference library includes spectra from the 47 metabolites spiked in the experimental dataset. It was generated, however, with collision energies slightly distinct from those used in the DIA dataset, which may result in some differences in the fragmentation patterns.

IV. Results and Discussion

To process DIA data efficiently and achieve global detection, quantification, and identification of the compounds in a biological sample, we developed a workflow called DIA-NMF based on NMF to unmix MS2 spectra and link precursor ions with their corresponding fragment ions. To evaluate the performance of DIA-NMF, we used a ground truth DIA dataset consisting of 24 files obtained from human plasma spiked with 47 compounds at 7 known concentrations in triplicate.

A. DIA-NMF successfully identify compounds, even at low concentrations

We first assessed the number of correct identifications among the spiked molecules (Fig. 1). DIA-NMF achieved 83% of correct identification at the highest 10 ng/ml concentration (30 correctly identified compounds out of 36 detected precursors) and still 71.43% at the lowest 0.05 ng/ml spiking concentration (10 correct identifications out of 14 precursors). Some missing identifications result from false positive detections of MS1 ions by the *centWave* algorithm at step 1, such

as curcumin and cholic acid (in such cases, the mixed matrices are merely noise). Importantly, DIA-NMF outperforms the peak model approaches at all concentrations (except at 1 ng/ml where DIA-NMF and MS-DIAL achieve similar results).

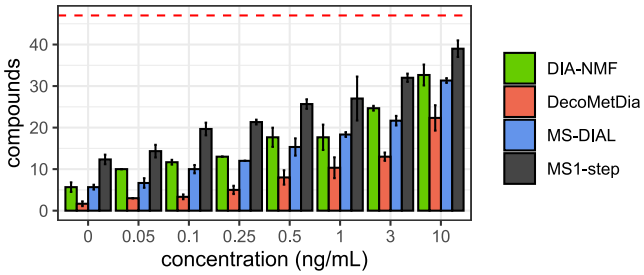

Fig. 1: Identification of spiked compounds in plasma by DIA-NMF, DecoMetDIA, and MS-DIAL. The 24 files from the DIA dataset were processed by the three software. The error bars correspond to the mean +/- standard deviation of the three replicates, whereas the red dashed line indicates the total number of spiked compounds (47). Note that at 0 ng/ml, the compounds were not spiked. Hence, the signal comes from the endogenous plasma compounds.

B. *DIA-NMF outperforms the model peak approach*

For a finer evaluation of the quality of the reconstructed MS2 spectra, we inspected the value of the matching score for all identified compounds (Fig. 2). At high spiking concentration (10 ng/mL, replicate 1), DIA-NMF, DecoMetDIA, and MS-DIAL identify 30, 23, and 32 compounds, respectively (Fig. 2 *left*). Notably, the 2 compounds specifically detected by MS-DIAL in this replicate (riboflavin and D-sphingosine) are also detected by DIA-NMF in the 2 other replicates from the same sample, suggesting that the absence of identification by DIA-NMF in the first replicate may result from an absence of detection of the precursor ion by the *centWave* algorithm which is used in step 1 in both DIA-NMF and DecoMetDIA pipelines. At a lower concentration (0.5 ng/mL, replicate 1), DIA-NMF still identifies 19 spiked molecules, while DecoMetDIA and MS-DIAL only identify 9 and 11 compounds, respectively (Fig. 2 *right*).

The NMF approach outperforms the model peak methods for compounds with close retention times. In the case of atropine (Fig. 3), DIA-NMF successfully unmixed the single peak EIC (Fig. 3 *left*) into two components (Fig. 3 *middle*) and correctly assigned component 1 to the pure MS2 spectrum of atropine (Fig. 3 *right*) with a score of 0.775 at 0.5ng/mL. The latter was superior to those from DecoMetDIA and MS-DIAL (0.61 and 0.66, respectively).

C. *DIA-NMF spectra are enriched in chemical information*

An advantage of our approach is that it groups all related MS1 and MS2 ions from the same molecule within a single component. In particular, adducts and isotopes of the precursor and their corresponding fragments are included in the generated spectrum (Fig. 4). Such additional fragments are of high interest for the chemical interpretation of the spectrum and the elucidation of the compound structure: the presence of an isotope (e.g., ^{34}S) may be used to confirm the molecular formula; additionally, the presence of adducts gives insights

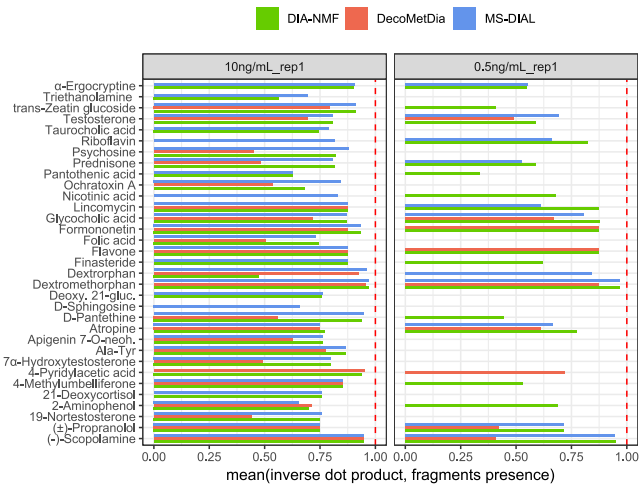

Fig. 2: Matching scores for the spiked compounds identified. The matching scores of the pure MS2 spectra resulting from the processing by DIA-NMF, DecoMetDIA, and MS-DIAL of the acquisition files at 0.5 ng/mL and 10 ng/mL spiking concentration (first replicate) are displayed.

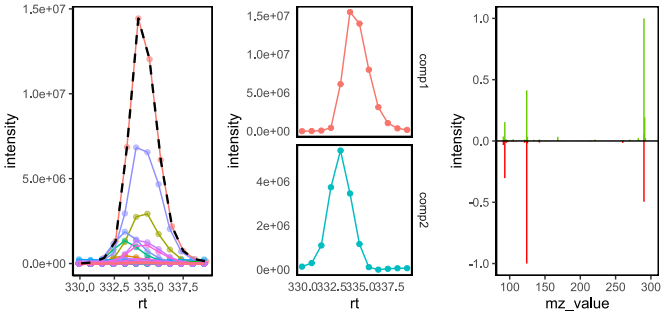

Fig. 3: Successful factorization by DIA-NMF of atropine from a co-eluting precursor of the same m/z 290.1751. *Left*: The EIC of the atropine precursor seems to contain a single peak (black dashed line), but the EICs from the candidate fragments (colored lines) suggest that it results from the contribution of two distinct precursors. *Middle*: The two components are factorized by NMF (the 2 rows of the $H^p_{1:2,t}$ matrix are shown), and component 1 is assigned to the atropine precursor. *Right*: The pure MS2 spectrum $W^p_{mz,1}$ (*top*) matches the reference spectrum from atropine (*bottom*) with an inverse dot product and a percentage of common fragments of 0.82 and 0.73, respectively. Retention times are expressed in seconds.

about the ionization of the molecule, and hence its structure. These additional peaks, however, are absent from the reference spectra in classical databases, where only a single ion species is selected as precursor for fragmentation (DDA approach). We, therefore, acquired reference spectra in the DIA mode (similar to the one used to analyze spiked plasma samples) for some of the compounds. As expected, the spectra generated by our DIA-NMF approach achieve the highest match to these DIA reference spectra compared to the pure spectra from DecoMetDIA and MS-DIAL (Fig. 4).

V. Conclusions

SWATH DIA is a promising approach for the high-throughput and comprehensive annotation of metabolites in biological samples. However, processing multiplexed data and

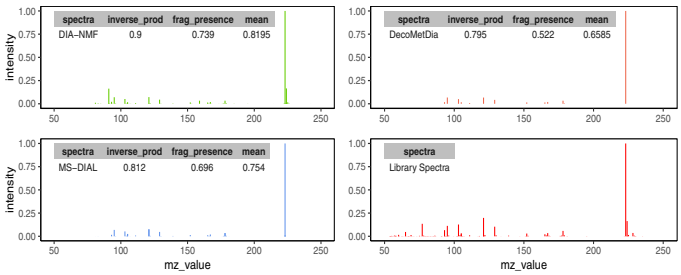

Fig. 4: DIA-NMF MS2 spectra contain additional relevant peak information: example of the flavone compound. Pure MS2 spectra were obtained with DIA-NMF, DecoMetDIA, and MS-DIAL compared to the MS2 reference spectrum obtained in the DIA mode.

extracting a pure MS2 spectrum for each precursor is challenging. Here, we present the DIA-NMF framework, which relies on NMF to efficiently unmix the fragments and extract pure MS2 spectra. DIA-NMF is an end-to-end workflow for SWATH DIA data analysis, consisting of several modules to process the raw data, factorize the multiplexed MS2 data, extract the pure MS2 spectra, and compute the matching scores to a reference database.

By applying DIA-NMF to a real dataset of human plasma spiked with 47 known compounds, we show that identification recalls up to 83% are achieved, outperforming existing approaches based on model peaks, especially at low concentrations and for co-eluting compounds.

Since the existing peak detection software (such as *centWave* used in DIA-NMF) is known to generate false positives and negatives, especially at low compound concentrations, we are currently developing alternative MS1 detection algorithms to increase the identification rate. In addition, to further improve the separation between the components (i.e., the purity of the MS2 spectra), we are developing new NMF algorithms that include sparsity on the elution profiles and use alternatives to the l_1 norm.

Interestingly, our method to build and factorize the matrix of elution profiles for each precursor p generates a spectrum that gathers all the MS1 and MS2 fragments related to the sought compound. This differs from existing approaches that only relate all the selected MS2 fragments to p. Our strategy is, therefore, not only more rigorous (because the detected MS2 ions in DIA may originate from adducts or isotopes of p since they all have similar elution profiles), but it also provides new insights into chemical interpretation by gathering the full MS and MS2 information about the compound. The comparison with reference spectra acquired in the same DIA conditions confirms that the DIA-NMF spectra contain additional peaks from isotopes and adducts. We are currently working on new algorithms to automatically annotate the spectra and facilitate their interpretation by the chemists.

In conclusion, DIA-NMF provides a new approach and workflow that will be of value for the global detection and identification of metabolomics.

References

[1] D. S. Wishart, "Metabolomics for investigating physiological and pathophysiological processes," *Physiological Reviews*, vol. 99, no. 4, pp. 1819–1875, 2019.

[2] C. Junot, F. Fenaille, B. Colsch, and F. Becher, "High resolution mass spectrometry based techniques at the crossroads of metabolic pathways," *Mass Spectrometry Reviews*, vol. 33, no. 6, pp. 471–500, 2014.

[3] R. Bonner and G. Hopfgartner, "SWATH data independent acquisition mass spectrometry for metabolomics," *TrAC Trends in Analytical Chemistry*, vol. 120, p. 115278, Nov. 2019.

[4] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, and M. Arita, "MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis," *Nature Methods*, vol. 12, no. 6, pp. 523–526, Jun. 2015.

[5] Y. Yin, R. Wang, Y. Cai, Z. Wang, and Z.-J. Zhu, "DecoMetDIA: Deconvolution of Multiplexed MS/MS Spectra for Metabolite Identification in SWATH-MS-Based Untargeted Metabolomics," *Anal. Chem.*, vol. 91, no. 18, pp. 11 897–11 904, 2019.

[6] A. Smirnov, Y. Qiu, W. Jia, D. I. Walker, D. P. Jones, and X. Du, "ADAP-GC 4.0: Application of Clustering-Assisted Multivariate Curve Resolution to Spectral Deconvolution of Gas Chromatography–Mass Spectrometry Metabolomics Data," *Anal. Chem.*, vol. 91, no. 14, pp. 9069–9077, Jul. 2019.

[7] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, "Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability, Algorithms, and Applications," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 59–80, Mar. 2019.

[8] J. Rapin, A. Souloumiac, J. Bobin, A. Larue, C. Junot, M. Ouethrani, and J.-L. Starck, "Application of non-negative matrix factorization to LC/MS data," *Signal Processing*, vol. 123, pp. 75–83, 2016.

[9] S. A. VAVASIS, "On the complexity of nonnegative matrix factorization," *SIAM J. OPTIM.*, vol. 20, no. 3, pp. 1364–1377, 2009.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[11] M. Zibulevsky and B. A. Pearlmutter, "Blind Source Separation by Sparse Decomposition in a Signal Dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, Apr. 2001.

[12] Y. Li, S.-i. Amari, S. Shishkin, J. Cao, F. Gu, and A. Cichocki, "Sparse Representation and Its Applications in Blind Source Separation," in *Advances in Neural Information Processing Systems*, vol. 16. MIT Press, 2003.

[13] A. Pascual-Montano, J. Carazo, K. Kochi, D. Lehmann, and R. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, Mar. 2006.

[14] P. L. Combettes and V. R. Wajs, "Signal Recovery by Proximal Forward-Backward Splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Jan. 2005.

[15] R. B. Matthew C Chambers, Brendan Maclean, "A cross-platform toolkit for mass spectrometry and proteomics," *Nature Biotechnology*, vol. 30, no. 10, pp. 918–920, 2012.

[16] P. Fogel, C. Geissler, N. Morizet, and G. Luta, "On Rank Selection in Non-Negative Matrix Factorization Using Concordance," *Mathematics*, vol. 11, no. 22, p. 4611, Jan. 2023.

[17] P. Barbier Saint Hilaire, K. Rousseau, A. Seyer, S. Dechaumet, A. Damont, C. Junot, and F. Fenaille, "Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics," *Metabolites*, vol. 10, no. 4, p. 158, 2020.