Scientific Research Publishing

Journal of Computer and Communications, 2024, 12, 11-25
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

The Deconvolution Method for Obtaining Correspondence in Data-Independent Acquisition Mass Spectrometry Data

Mingming Lyu, Xinye Bian, Changjiu He

College of Computer Science and Technology, Shandong University of Technology, Zibo, China
Email: 17861418173@163.com

How to cite this paper: Lyu, M.M., Bian, X.Y. and He, C.J. (2024) The Deconvolution Method for Obtaining Correspondence in Data-Independent Acquisition Mass Spectrometry Data. *Journal of Computer and Communications*, **12**, 11-25.
https://doi.org/10.4236/jcc.2024.122002

Received: January 14, 2024
Accepted: February 6, 2024
Published: February 9, 2024

Abstract

Although data-independent acquisition (DIA) shows powerful potential in achieving comprehensive peptide information acquisition, the difficulty in determining the precursor m/z and distinguishing fragment ions has posed challenges in DIA data analysis. To address this challenge, a common approach is to recover the correspondence between precursor ions and fragment ions, followed by peptide identification using traditional data-dependent acquisition (DDA) database searching. In this study, we propose a cosine similarity-based deconvolution method that rapidly establishes the correspondence between chromatographic profiles of precursor ions and fragment ions through matrix calculations. Experimental results demonstrate that our method, referred to as CosDIA, yields a peptide identification count close to that of DIA-umpire. However, compared to DIA-umpire, we can establish the correspondence between original MS/MS spectra and pseudo-MS/MS spectra. Furthermore, compared to the CorrDIA method, our approach achieves higher efficiency in terms of time, reducing the time cost of the analysis process. These results highlight the potential advantages of the CosDIA method in DIA data analysis, providing a powerful tool and method for large-scale proteomics research.

Keywords

MS/MS Spectra, XICs, Correspondence, Matrix, Isotopic Peak Cluster

1. Introduction

The bottom-up proteomic approach based on liquid chromatography-tandem mass spectrometry (LC-MS/MS) is widely acknowledged as the primary method for protein analysis and identification. Analyzing and identifying proteins within

biological systems are crucial for a deeper understanding of disease mechanisms and drug development. The main workflow of proteomic analysis involves several steps: initially, specific enzymes are used to digest proteins within the biological system into a mixture of peptides. Subsequently, the mass spectrometer ionizes these peptide mixtures and introduces them into the instrument, where, influenced by electric fields and gravity, these peptide mixtures generate mass spectrometric data [1] [2]. Analysis of this mass spectrometric data enables the identification of peptides and proteins. During the phase of data acquisition in mass spectrometry, two modes are commonly employed based on the method of acquisition: data-dependent acquisition (DDA) and data-independent acquisition (DIA) [3]. Data-dependent acquisition (DDA) is generally more suitable for small sample sizes and relatively simple protein identification, whereas data-independent acquisition (DIA) is better suited for large sample sizes and complex protein identification.

Currently, traditional protein identification methods primarily utilize mass spectrometric data acquired through data-dependent acquisition (DDA). In the cycle of DDA, for each obtained MS1 spectra, the mass spectrometer selects a narrow isolation window and sequentially isolates and fragments the highest-intensity peaks in the MS1 spectra, generating MS/MS spectra (as shown in **Figure 1**, DDA-MS). Since DDA employs narrow isolation windows, the MS/MS spectra produced have minimal ion interference. However, DDA overlooks lower-intensity peaks in the MS1 spectra as it selects the top N peaks in terms of intensity for fragmentation, which may lead to the omission of some peaks with lower intensities [4]. Furthermore, due to the stochastic nature of liquid chromatography and the variability in the top N peaks in the MS1 spectra, the repeatability of mass spectrometric data obtained under the same experimental conditions is relatively poor. To address the limitations of DDA, data-independent acquisition (DIA) was introduced. In the cyclic data acquisition process of DIA, for each acquired MS1 spectra, the mass spectrometer chooses a larger isolation window and unbiasedly fragments all precursor ions within that window, generating MS/MS spectra [5] (as illustrated in **Figure 1**, DIA-MS). Since DIA uniformly fragments all precursor ions in the MS1 spectra, it theoretically enables the collection of information from all precursor ions. Compared to DDA, DIA can capture information from low-intensity peaks in the MS1 spectra. Additionally, due to the isolation window's inclusive nature, DIA's data acquisition method offers better reproducibility. Unlike DDA, DIA's MS/MS spectra contain the chromatographic profile information of fragment ions. However, due to the broad isolation windows used in DIA, it results in highly complex and co-fragmented MS/MS spectra [6]. This not only disrupts the correspondence between precursor and fragment ions, but also introduces issues of fragment ion interference in DIA's MS/MS spectra, posing significant challenges for the analysis of DIA data.

To address the challenges of determining precursor m/z and distinguishing fragment ions in data-independent acquisition (DIA), the primary approach

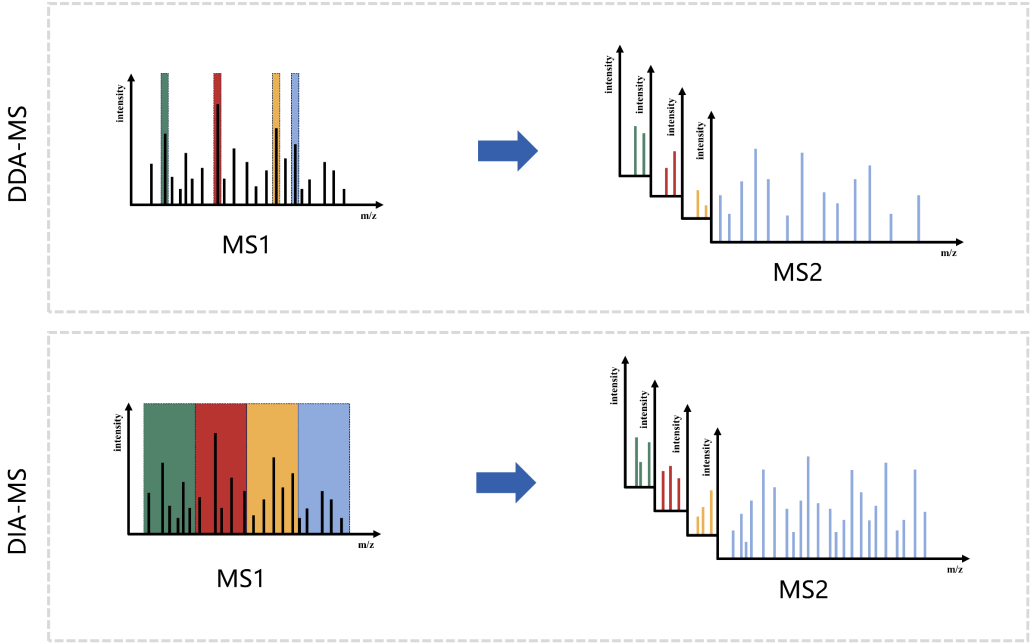

Figure 1. Comparison of DDA and DIA methods.

currently used for DIA data analysis is the pseudo-MS/MS spectra method [7]. The pseudo-MS/MS spectra method involves splitting DIA MS/MS spectra into multiple MS/MS, each containing the fragmented information of a single precursor ion, through spectrum deconvolution algorithms. Subsequently, traditional data-dependent acquisition (DDA) identification tools can be used to identify proteins from these split MS/MS spectra, enabling effective analysis and identification of DIA MS/MS spectra. Currently, there are several tools available for pseudo-MS/MS spectra splitting, including DeMux [8], DIA-Umpire [9], Group-DIA [10], and CorrDec [11], among others. The CorrDec method has already been integrated into the metabolomics analysis platform MS-DIAL [12] [13].

In summary, we propose a spectral deconvolution algorithm that combines cosine similarity calculation with matrix operations, named CosDIA. For the deconvolution of DIA spectra, CosDIA is capable of obtaining the correspondence between pseudo-MS/MS spectra and the scan numbers of the original DIA spectra, facilitating the traceability of the analysis results of the pseudo-MS/MS spectra. Moreover, by storing chromatographic curves in a matrix and performing similarity calculations, CosDIA reduces the time required for spectral deconvolution. These features enable CosDIA to more rapidly deconvolute DIA data and obtain additional corresponding relationships.

2. Workflow and Method

First, based on the information contained in the configuration file, we performed the following steps: extracted the chromatographic profiles of precursor ions and fragment ions, and identified isotope peak clusters in the spectra and

their charge states. Additionally, we identified noise peaks and removed them along with their corresponding chromatographic profiles from the spectra. These chromatographic profiles were organized and stored in a matrix for subsequent rapid similarity calculations. Combining the information about the identified isotope peak clusters, we could determine the charge states of the top N precursor ions. Finally, we applied the generated pseudo-MS/MS spectra to the pFind [14] tool for protein identification. The entire workflow was illustrated in **Figure 2**.

2.1. DIA Mass Spectrometry Data

Data-independent acquisition (DIA) was a method of generating MS/MS spectra by co-fragmenting precursor ions within isolation windows in the MS1 spectrum. Consequently, DIA data acquisition was characterized by the challenges of determining precursor m/z and distinguishing fragment ions. To mitigate the complexity of DIA spectra and facilitate data analysis, DIA data acquisition was categorized into three primary methods. The first method involved full-window fragmentation, exemplified by the MSE technique. The second method employed isolation window fragmentation, subdividing DIA data acquisition into fixed and variable isolation windows. Fixed isolation window methods maintained a constant window width, while variable isolation window methods allowed width adjustments during data acquisition as needed, offering greater flexibility for different sample types or experimental conditions. The third approach was the incorporation of additional data dimensions, known as 4D-DIA methods, such as DIA-PASEF [15] and Scanning SWATH [16]. These methods introduced extra data dimensions, such as ion mobility, to enhance the information richness of DIA data, ultimately improving the accuracy and comprehensiveness of data analysis and identification. These approaches aimed to address the difficulties associated with determining precursor m/z and distinguishing fragment ions in DIA data while enhancing precision and completeness in data analysis.

Due to the non-uniform distribution of precursor ions in the MS1 spectra of DIA data, employing identical isolation windows could lead to increased ion interference in windows with densely distributed parent ions, thus elevating the

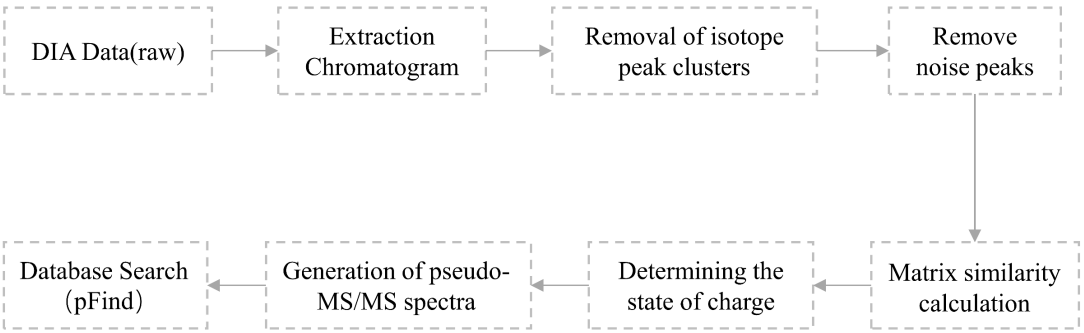

Figure 2. The workflow of the CosDIA method.

complexity of the resulting MS/MS spectra. In our research, we opted for DIA mass spectrometry data with HeLa proteins as the sample, digested with trypsin, and employed variable isolation window fragmentation. Our original data was sourced from ProteomeXchange (PXD005573) and specifically utilized the raw data from a 0.5-hour chromatographic gradient.

2.2. Chromatographic Profile Extraction

In the context of DIA methodology and the acquired mass spectrometry data, the relationship between MS1 and MS/MS spectra could be described as follows: sequentially in time, each MS1 spectrum corresponded to multiple MS/MS spectra. This correspondence could be further explained as multiple MS/MS spectra sequentially aligning with different isolation windows in the MS1 spectrum (as shown in **Figure 3(a)**). To extract the chromatographic profiles of fragment ions, we grouped multiple MS/MS spectra originating from the same isolation window within a specific time range (as illustrated in **Figure 3(b)**). Similarly, to extract the chromatographic profiles of precursor ions, we group MS1 spectra corresponding to each grouped of MS/MS spectra in chronological order. Since the MS/MS spectra within each group all originated from the same isolation window, aligning the peaks within the MS/MS spectra in each group within an error margin allowed them to be connected sequentially in time, resulting in the

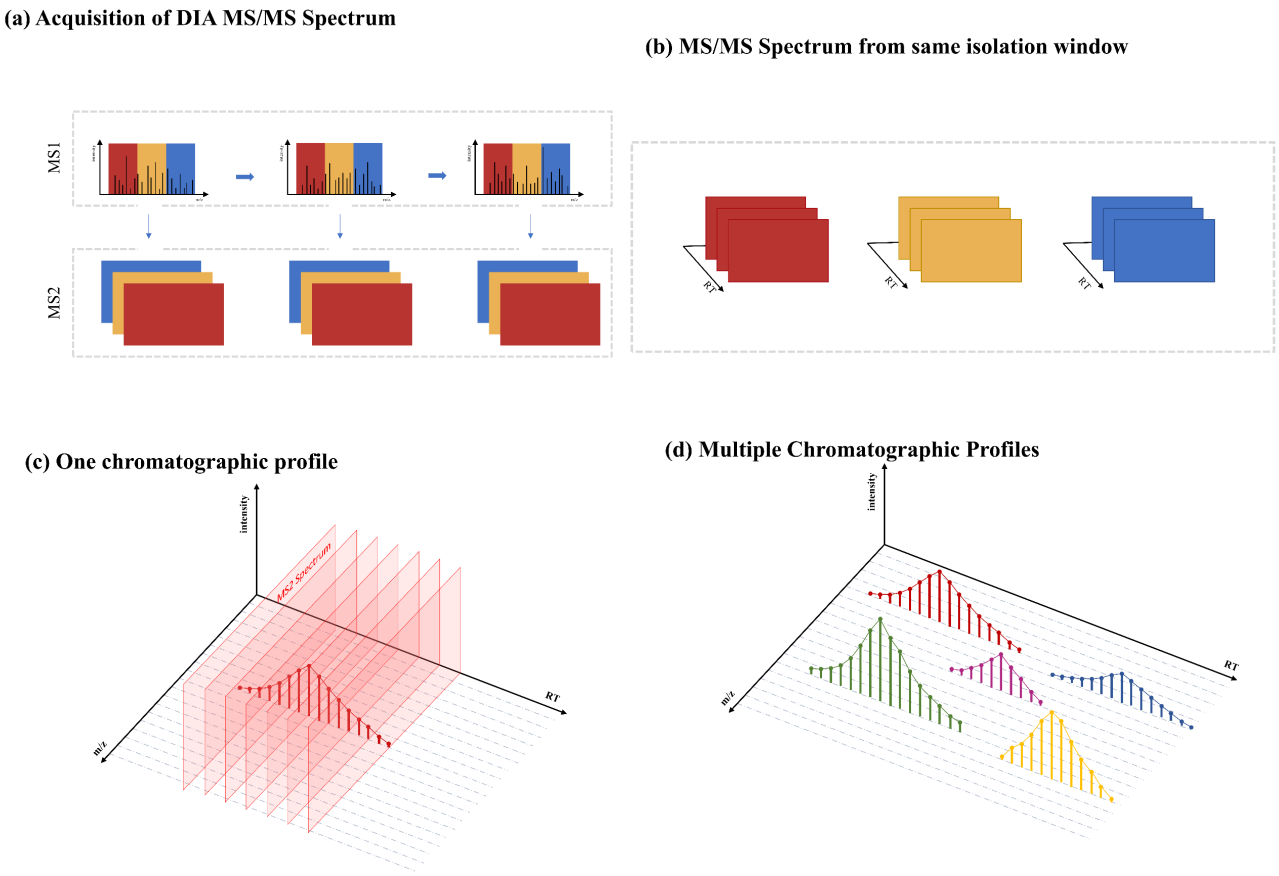

Figure 3. Chromatographic profile extraction.

formation of chromatographic profiles for fragment ions (as depicted in **Figure 3(c)**). This process was repeated multiple times across several groups of MS/MS spectra, ultimately yielding multiple chromatographic profiles (as seen in **Figure 3(d)**), with each profile representing the specific movement trajectory and abundance distribution of particular ions.

In our experiment, we began by identifying the DIA MS/MS spectra to be deconvoluted. Subsequently, within a ±0.1-minute range around each cycle in the DIA MS/MS spectrum, we located the corresponding DIA MS/MS spectra that shared the same positional alignment. Similarly, based on the retention time, we found the corresponding DIA MS1 spectra. For each group of identified DIA MS/MS spectra, we recorded which MS/MS spectrum needed to be deconvoluted, its origin from the isolation window in the MS1 spectrum, as well as the scan numbers and retention times for all MS/MS spectra within that group. Within each set of DIA MS/MS spectra, we aligned the corresponding peaks in the before-and-after MS/MS spectra for each peak in the MS/MS spectrum to be deconvoluted. We set the error tolerance between peaks at 0.02 daltons. By locating the corresponding peaks in the before-and-after MS/MS spectra from a given peak in the MS/MS spectrum to be deconvoluted and connecting these peaks within the retention time range, we obtained a chromatographic profile for the MS/MS spectrum to be deconvoluted. This step ensured the accuracy of the generated chromatographic profiles. Subsequently, we stored each group of DIA MS1 and MS/MS spectra as matrices. In our DIA mass spectrometry data, each DIA MS/MS spectrum contained one-dimensional information, recording the origin isolation window from its MS1 spectrum. Therefore, we could determine which isolation window in the MS1 spectrum the chromatographic profiles of fragment ions originated from. This greatly aided in reducing the search space for similarity calculations between precursor ion and fragment ion chromatographic profiles.

2.3. Chromatographic Profile Similarity Calculation

Due to the characteristics of DIA mass spectrometry, where all precursor ions within an isolation window were co-fragmented to generate MS/MS spectra, it could be challenging to determine the precise source of precursor ions in DIA MS/MS spectra. Additionally, in DIA MS/MS spectra, fragment ions might have originated from multiple precursor ions, making it difficult to establish associations between specific fragment ions and precursor ions. To address this issue, we recovered the correspondence between MS/MS spectra and individual precursor ions by comparing the similarity between the chromatographic profiles of precursor ions and fragment ions. In our experiment, we initially identified other DIA MS/MS spectra originating from the same isolation window within a ±0.1-minute range around the DIA MS/MS spectrum to be deconvoluted. By aligning and connecting the peaks in the preceding and subsequent DIA MS/MS spectra, we created chromatographic profiles for fragment ions (as illustrated in

Figure 4). Similarly, we aligned and connected the corresponding peaks in the DIA MS1 spectra, resulting in chromatographic profiles for precursor ions.

For all chromatographic profiles of precursor ions $P(p_1, p_2, \cdots, p_n)$ and all chromatographic profiles of fragment ions $F(f_1, f_2, \cdots, f_n)$, where n represented the number of ions, the cosine similarity calculation formula for chromatographic profiles was as follows [17]:

$$C = corr(P, F) = \frac{P \cdot F}{\|P\|\|F\|} = \frac{\sum_{i=1}^{n} p_i * f_i}{\sqrt{\sum_1^n (p_i)^2} * \sqrt{\sum_1^n (f_i)^2}} \quad (1)$$

To calculate the similarity between chromatographic profiles of precursor ions and fragment ions more efficiently, we converted the generated chromatographic profiles into matrix form. The number of rows in the matrix represented the quantity of precursor ion chromatographic profiles, while the number of columns represented the number of DIA spectra in a given group (as illustrated in Figure 4). Each element in the matrix stored the corresponding peak intensity values. This means that each row represented a chromatographic profile. Since the chromatographic profiles of fragment ions were formed in the retention time dimension by MS/MS spectra co-fragmented at the same position in the MS1 spectra, this implied that the generated chromatographic profiles of fragment ions all originated from the respective isolation windows in the MS1 spectra. This aided in reducing the search space when calculating the similarity between chromatographic profiles of precursor ions and fragment ions.

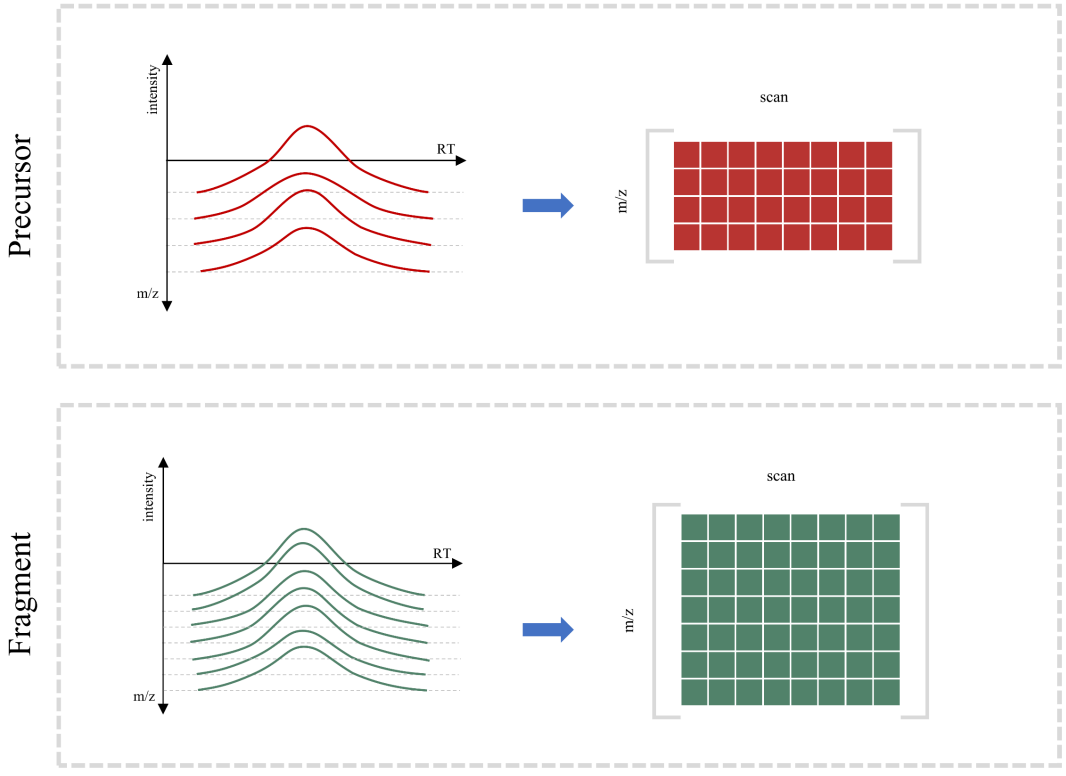

Figure 4. Chromatographic curve storage.

We calculated the cosine similarity between a row in the matrix of chromatographic profiles of precursor ions and every row in the matrix of chromatographic profiles of fragment ions. This provided the similarity between each chromatographic profile of precursor ions and all chromatographic profiles of fragment ions. By performing these similarity calculations row-wise between these two matrices, we obtained a similarity matrix that captured the similarities between all chromatographic profiles of precursor ions and fragment ions. The number of rows in the similarity matrix corresponded to the quantity of precursor ion chromatographic profiles, and the number of columns corresponded to the quantity of fragment ion chromatographic profiles. The values in the matrix represented the similarity scores between chromatographic profiles of precursor ions and fragment ions. In this case, we had 25 chromatographic profiles of precursor ions and 50 chromatographic profiles of fragment ions (as shown in **Figure 5**). Each row in the matrix represented the similarity scores between a single chromatographic profile of a precursor ion and all chromatographic profiles of fragment ions. We computed the number of values in each row greater than 0.6 and used this count as the score for that precursor ion. Consequently, we selected the top 30 precursor ions with the highest scores to serve as the precursor m/z for the pseudo MS/MS spectra.

2.4. Removing Isotopic Clusters and Determining Parent Ion Charges

The presence of isotopic clusters of precursor ions in the MS1 spectra could introduce interference when calculating the similarity between chromatographic profiles of precursor ions and fragment ions. Since the chromatographic profiles

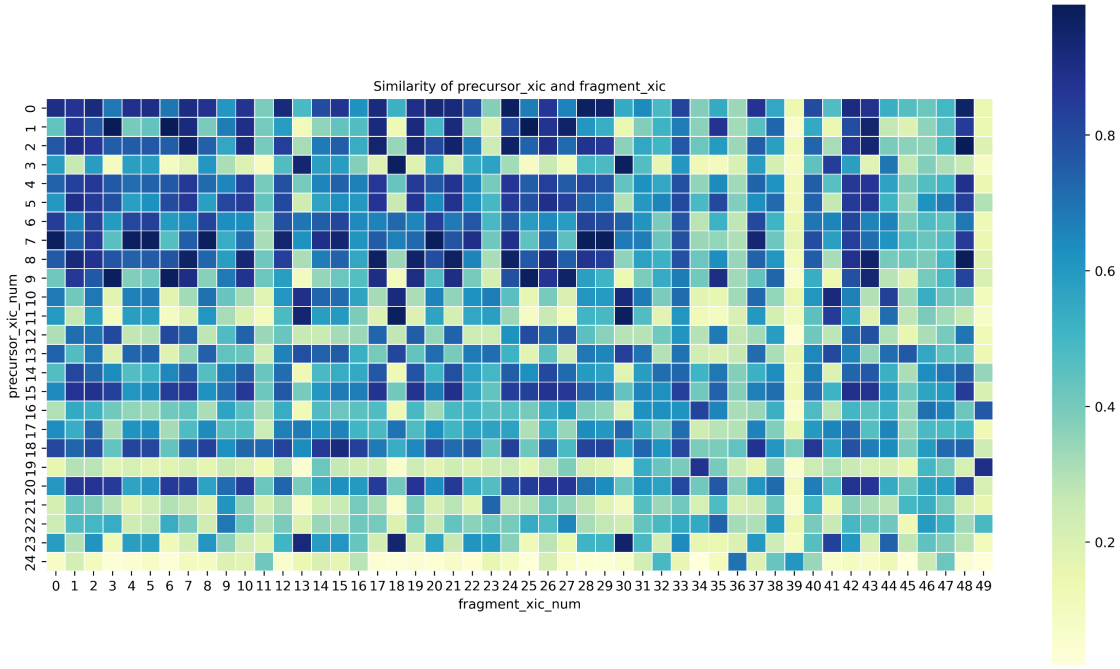

Figure 5. Chromatographic curve similarity calculation.

corresponding to isotopic clusters had similar shapes, this could lead to high-scoring precursor ions and their isotopes being included in the topN list when calculating similarity with fragment ion chromatographic profiles. As a result, this may have caused other genuine precursor ions to go undetected.

To address this issue, we could use the calculation of chromatographic profiles of precursor ions and fragment ions to determine the precursor ion's mass-to-charge ratio (m/z) and charge state. In the MS1 spectra, we could infer the charge state of precursor ions by computing the differences in m/z between pairs of peaks. The difference δ between peaks was approximately equal to Mneutron/c, where c was the charge state of the precursor ion, and Mneutron represented the mass of a neutron. According to the calculation formula, the peak spacing corresponding to precursor ions with different charge states being 1.003 da (+1 charge state), 0.5015 da (+2 charge state), 0.3343 da (+3 charge state), and 0.2508 da (+4 charge state) (as illustrated in Figure 6). For those precursor ions for which isotopic peaks were not found, we defaulted to assuming a +2 charge state. It's worth noting that, for established isotopic clusters, if the peak spacing between peaks in the cluster met the distribution requirements of an isotopic cluster and the peak intensities in the cluster did not exhibit a pattern of alternating high and low, then it could be confirmed as an isotopic cluster. Otherwise, the constructed cluster was not an isotopic cluster.

In our experiment, to improve computational speed, we first identified all isotopic clusters in the MS1 spectra and added all peaks except the first isotopic peak to the isotopic list. Then, we iterated through the precursor ion list, which

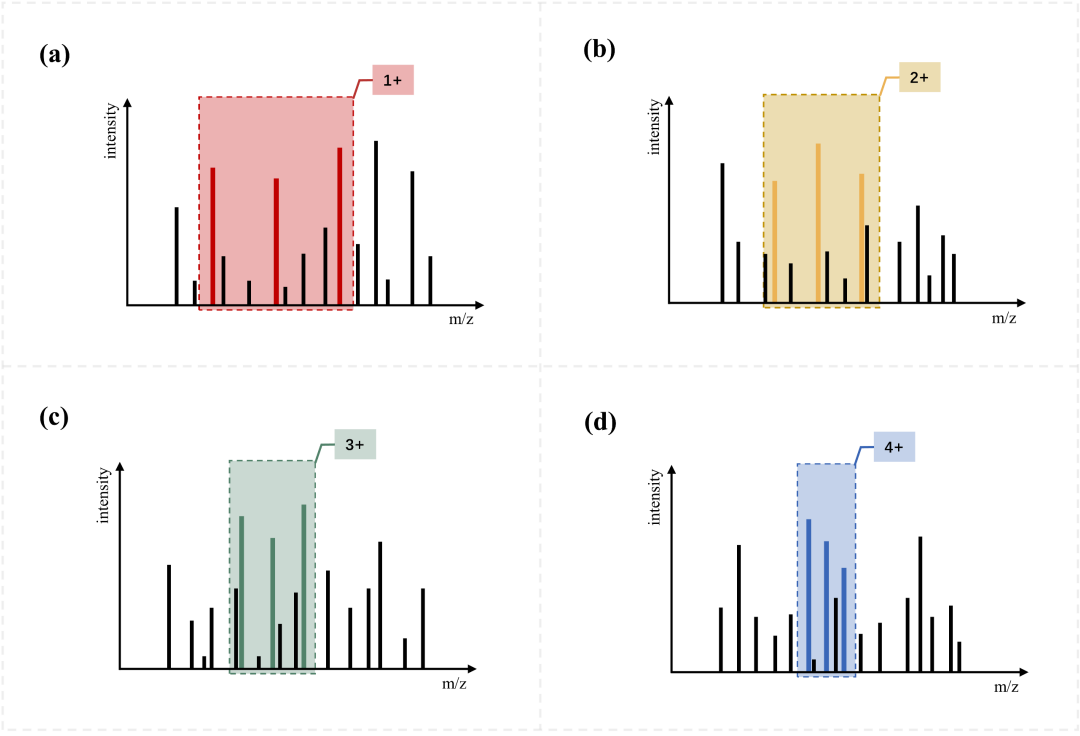

Figure 6. Isotopic peak cluster and its charge.

was sorted in descending order of scores, to identify precursor ions that existed in the isotopic list. We removed these precursor ions from the precursor ion list to avoid the influence of isotopes.

2.5. Identification of Pseudo-MS/MS Spectra

The pseudo-MS/MS spectra generated from the deconvolution results were stored in .mgf format, which was is the same as the traditional data storage format used in DDA. It's important to note that the scan numbers in the deconvoluted results were obtained from the original DIA MS/MS spectra. As a result, traditional database search software could be used for the identification of deconvoluted results, such as pFind [18], DIAmeter [19], de novo [20], and other similar software. In our experiment, we chose pFind to identify our deconvoluted results and DIA-umpire's deconvoluted results. When identifying these two sets of deconvoluted results, we used the same parameter settings. As shown in Table 1.

3. Results

3.1. Experimental Parameters

When considering the reproducibility of the experiment and the stability of the results, the choice of parameters such as retention time, chromatographic curve similarity threshold, and the number of top N pseudo-MS/MS spectra to extract becomes crucial. To assess the impact of different parameters on the experimental results and find the optimal parameter combinations, we employed a rigorous controlled variable method. In our experiment, we carefully selected 1500 DIA MS/MS spectra for testing and used the pFind software for the identification of pseudo-MS/MS spectra while maintaining a false discovery rate (FDR) of 1%. Through statistical analysis of the results obtained from pFind, we were able to clearly determine the optimal parameter combination. When considering the retention time of chromatographic curves, we found that setting it to 0.2 minutes resulted in the best identification outcomes. Additionally, we also established that a chromatographic curve similarity threshold of 0.6 produced the optimal identification performance (as shown in Figure 7).

3.2. Isotope Peak Cluster Calculation

In the field of protein mass spectrometry, the presence of isotope clusters can

Table 1. Database search parameters of pFind.

Parameter	Value
Precursor Tolerance	±20 ppm
Fragment Tolerance	±20 ppm
FDR	Less than 1% at Peptides Level
Fixed Modifications	Carbamidomethyl (C)
Peptide Length	[6, 100]

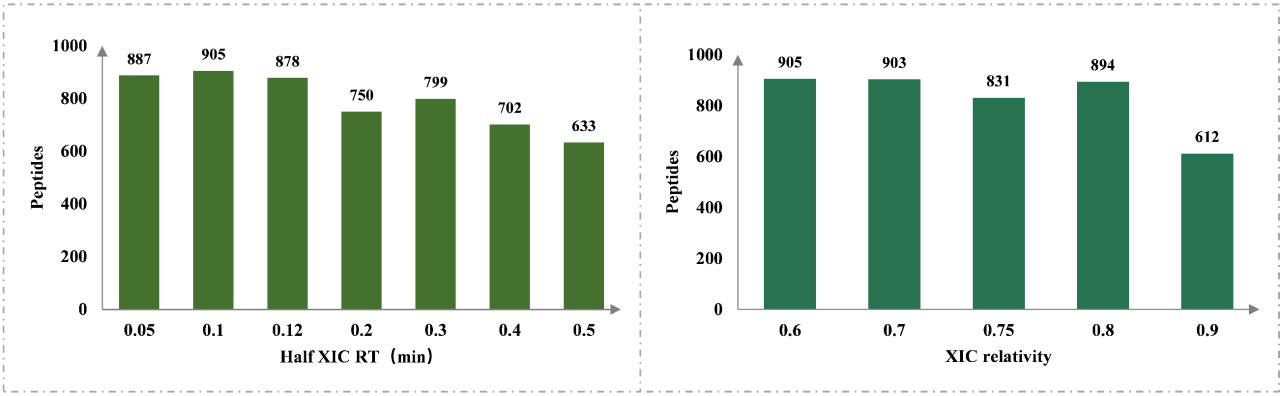

Figure 7. Effect of different parameters on experimental results.

interfere with the similarity calculations of chromatographic curves for both precursor ions and fragment ions. This interference occurs because the chromatographic curves of isotope clusters typically exhibit highly similar characteristics. This similarity can lead to the inclusion of isotope clusters as high-scoring peaks when calculating similarity with fragment ion chromatographic curves, causing pFind to identify redundant peptides. To reduce the interference of isotope clusters in chromatographic curve similarity calculations, we first identify isotope clusters in the MS1 spectra. Then, before performing chromatographic curve similarity calculations, we remove all peaks except the monoisotopic peak. By doing so, we only use data from monoisotopic peaks in the similarity calculations for precursor ions and fragment ions, reducing the influence of isotope clusters. As an example, in a specific MS1 spectrum, we identified four isotope clusters and determined the corresponding monoisotopic m/z using our isotope algorithm. In subsequent chromatographic curve similarity calculations, we only utilized the data from these monoisotopic peaks to obtain more accurate results (as shown in **Figure 8**).

3.3. Identification of Results and Time Spent

In previous studies, we primarily used DIA-Umpire as the deconvolution tool. However, to compare the performance of different deconvolution tools, we applied both CosDIA and DIA-Umpire to perform deconvolution on the same HeLa dataset. The output results from these two tools were saved in the .mgf file format, allowing their outputs to be used with traditional data-dependent acquisition (DDA) search software for mass spectrometry identification. In our experiments, we utilized the database search software pFind to identify the pseudo-MS/MS spectra generated from the deconvolution results. We set the false discovery rate (FDR) at 1% as a condition to filter the peptide identifications.

First, we collected and compared the peptide identification results obtained by pFind. For DIA-Umpire's deconvolution results, we used the best deconvolution file, <filename>_Q1.mgf, for database searching. To validate the accuracy of the deconvolution results, we also compared the peptide identification results obtained by pFind with those from DIA-NN [21]. It was observed that in CosDIA's

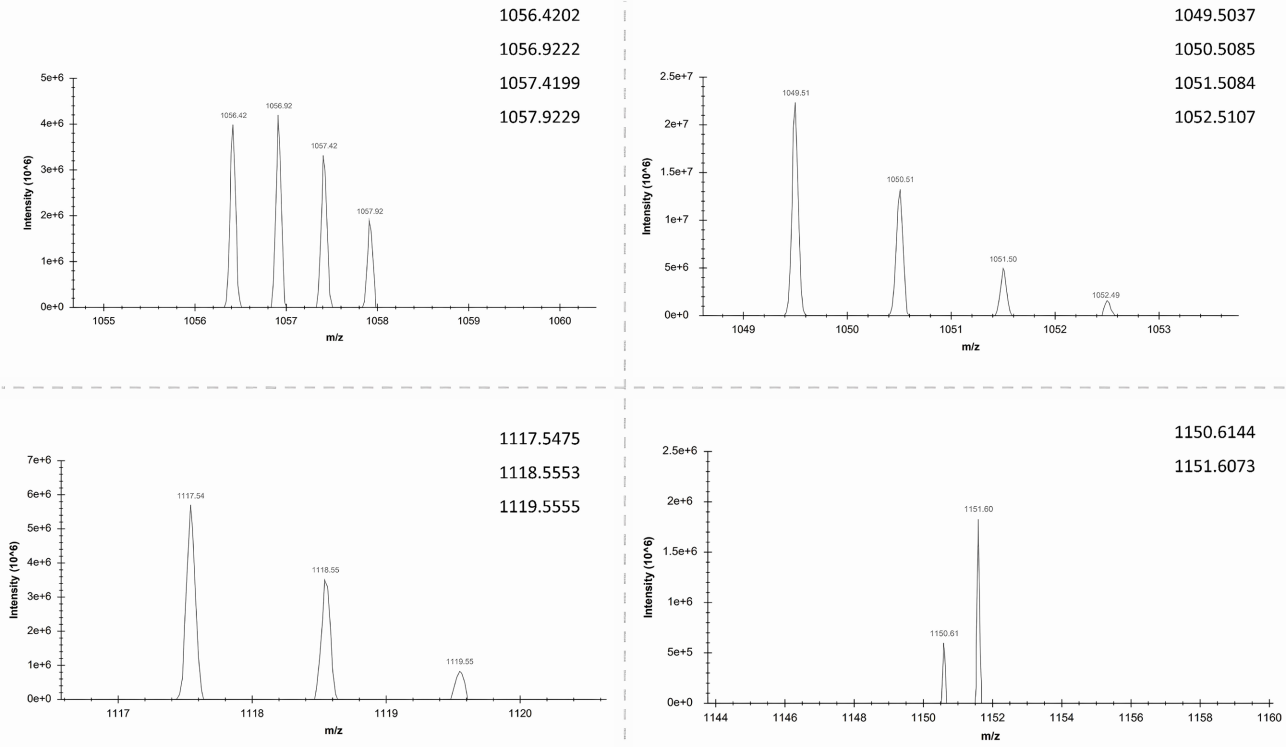

Figure 8. Precursor ion isotopic peak cluster.

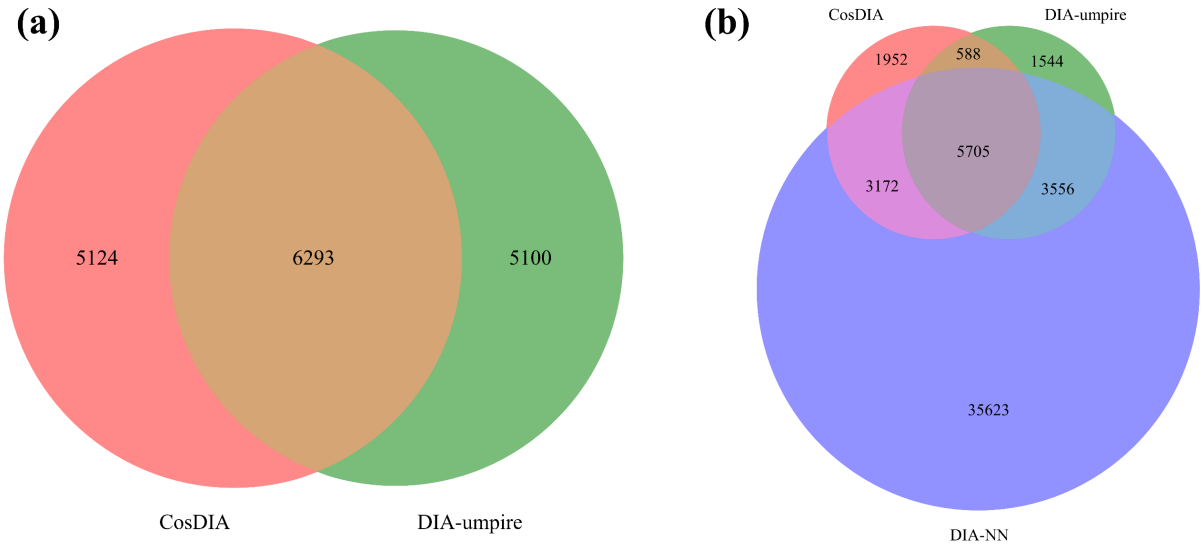

Figure 9. Venn diagram of pFind identified peptides.

identification results, 44.9% of the peptides were not identified by DIA-Umpire, while in DIA-Umpire's identification results, 44.8% of the peptides were not identified by CosDIA (as shown in **Figure 9(a)**). Furthermore, in CosDIA's identification results, 22.2% of the peptides were not identified by DIA-NN, while in DIA-Umpire's identification results, 18.7% of the peptides were not identified by DIA-NN (as shown in **Figure 9(b)**).

In comparison to DIA-Umpire, our approach, based on a spectrum-centric

Table 2. Hardware parameters, data and time spending.

Parameter	Value
Raw	HeLa-0.5h.raw
CPU	Intel(R) Core(TM) i7-11800H @ 2.30 GHz
Memory	32.0 GB
Operating System	Windows 11 64 bit
CorrDIA Time Spending(s)	1,814,400
CosDIA Time Spending(s)	3772.3

deconvolution strategy, allowed us to establish a correspondence between pseudo-MS/MS spectra and original MS/MS spectra. Furthermore, we also recorded the computer hardware configuration, data, and the corresponding time consumption, as detailed in Table 2.

4. Conclusions

One method of data analysis in data-independent acquisition (DIA) is the pseudo-MS/MS spectra approach, which primarily aims to establish the correspondence between individual precursor ions and their fragment ions. The core of this method involves computing the chromatographic similarity of precursor and fragment ion chromatograms to resolve this correspondence. Due to DIA data involving the simultaneous fragmentation of all precursor ions within isolation windows, the MS/MS spectra in DIA contain a large number of peaks, leading to a significant increase in computational complexity when comparing chromatographic similarities. To address this issue, we store precursor and fragment ion chromatograms from within the isolation windows in matrices, facilitating efficient similarity calculations. Additionally, to establish the association between pseudo-MS/MS spectra and the original MS/MS spectra, we extract chromatograms centered around the original MS/MS spectra and perform similarity calculations with the fragment ion chromatograms to determine the correspondence.

We compared two different methods, CosDIA and DIA-Umpire, in our experiments. The results indicate that they tend to identify a similar number of peptide segments effectively. However, our method offers the additional advantage of establishing the correspondence between original MS/MS spectra and pseudo-MS/MS spectra. In terms of time efficiency, we compared our method with CorrDIA [22]. Due to our utilization of matrix operations, our approach proves to be more computationally efficient.

Since we extract chromatographic curves centered around original MS/MS spectra, there may be some redundancy in the number of identified peptide segments. DIA MS/MS spectra are generated through shared fragmentation, which could introduce some interference in the chromatographic curves, potentially leading to lower identification rates for pseudo-MS/MS spectra. The chro-

matographic curve similarity determination method relies on cosine similarity, which might have limitations in sensitivity. The pseudo-MS/MS spectra method primarily relies on the consistency of chromatographic curves for the dissection of original MS/MS spectra. With the rapid development of deep learning, one potential approach is to initially obtain data identification results through database search software. Then, based on these identification results and their corresponding chromatographic curves, you can create positive and negative instances [23]. Using these instances, you can train a model, and then employ the trained model to compute the similarity of chromatographic curves between precursor ions and fragment ions [24]. The multi-parameter nature of deep learning may offer better results. As mass spectrometry instruments continue to evolve, mass spectrometers increase their ability to distinguish chromatographic curves by collecting additional new dimensions of information, thereby enhancing their ability to deconvolute spectra. For example, ion mobility and sliding quadrupole technologies are introduced during the collection process. Furthermore, in the data preprocessing stage, deep learning or machine learning methods can be utilized to differentiate and remove isotope peak clusters and their corresponding chromatographic curves.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Steen, H. and Mann, M. (2004) The ABC's (and XYZ's) of Peptide Sequencing. *Nature Reviews Molecular Cell Biology*, **5**, 699-711. https://doi.org/10.1038/nrm1468

[2] Bilbao, A., Varesio, E., *et al.* (2015) Processing Strategies and Software Solutions for Data-Independent Acquisition in Mass Spectrometry. *Proteomics*, **15**, 964-80. https://doi.org/10.1002/pmic.201400323

[3] Aebersold, R. and Mann, M. (2016) Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature*, **537**, 347-55. https://doi.org/10.1038/nature19949

[4] Li, X., Wang, W.Q. and Chen, J.J. (2017) Recent Progress in Mass Spectrometry Proteomics for Biomedical Research. *Science China Life Sciences*, **60**, 1093-1113. https://doi.org/10.1007/s11427-017-9175-2

[5] Ludwig, C., *et al.* (2018) Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial. *Molecular Systems Biology*, **14**, e8126. https://doi.org/10.15252/msb.20178126

[6] Röst, H.L., *et al.* (2014) OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nature Biotechnology*, **32**, 219-23. https://doi.org/10.1038/nbt.2841

[7] Hou, X.-H., *et al.* (2022) Progress in Data Analysis Methods for Proteome Mass Spectrometry Based on Data-Independent Acquisition. *Progress in Biochemistry and Biophysics*, **49**, 2364-2386.

[8] Bern, M., Finney, G., *et al.* (2010) Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Analytical Chemistry*,

82, 833-41. https://doi.org/10.1021/ac901801b

[9] Tsou, C.-C., *et al.* (2015) DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nature Methods*, **12**, 258-264. https://doi.org/10.1038/nmeth.3255

[10] Li, Y.Y., *et al.* (2015) Group-DIA: Analyzing Multiple Data-Independent Acquisition Mass Spectrometry Data Files. *Nature Methods*, **12**, 1105-1106. https://doi.org/10.1038/nmeth.3593

[11] Tada, I., *et al.* (2020) Correlation-Based Deconvolution (CorrDec) to Generate High-Quality MS2 Spectra from Data-Independent Acquisition in Multisample Studies. *Analytical Chemistry*, **92**, 11310-11317. https://doi.org/10.1021/acs.analchem.0c01980

[12] Tsugawa, H., *et al.* (2015) MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nature Methods*, **12**, 523-526. https://doi.org/10.1038/nmeth.3393

[13] Tsugawa, H., *et al.* (2020) A Lipidome Atlas in MS-DIAL 4. *Nature Biotechnology*, **38**, 1159-1163. https://doi.org/10.1038/s41587-020-0531-2

[14] Shao, G.C., *et al.* (2021) How to Use Open-pFind in Deep Proteomics Data Analysis?—A Protocol for Rigorous Identification and Quantitation of Peptides and Proteins from Mass Spectrometry Data. *Biophysics Reports*, **7**, 207-226. https://doi.org/10.52601/bpr.2021.210004

[15] Meier, F., *et al.* (2020) diaPASEF: Parallel Accumulation-Serial Fragmentation Combined with Data-Independent Acquisition. *Nature Methods*, **17**, 1229-1236. https://doi.org/10.1038/s41592-020-00998-0

[16] Messner, C.B., *et al.* (2021) Ultra-Fast Proteomics with Scanning SWATH. *Nature Biotechnology*, **39**, 846-854. https://doi.org/10.1038/s41587-021-00860-4

[17] Chowdhury, G.G. (2010) Introduction to Modern Information Retrieval. Facet Publishing, London.

[18] Hao, C., *et al.* (2018) Open-pFind Enables Precise, Comprehensive and Rapid Peptide Identification in Shotgun Proteomics. *BioRxiv*, Article ID: 285395.

[19] Lu, Y.Y., *et al.* (2021) DIAmeter: Matching Peptides to Data-Independent Acquisition Mass Spectrometry Data. *Bioinformatics*, **37**, i434-i442. https://doi.org/10.1093/bioinformatics/btab284

[20] Tran, N.H., *et al.* (2019) Deep Learning Enables de Novo Peptide Sequencing from Data-Independent-Acquisition Mass Spectrometry. *Nature Methods*, **16**, 63-66. https://doi.org/10.1038/s41592-018-0260-3

[21] Demichev, V., *et al.* (2020) DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nature Methods*, **17**, 41-44. https://doi.org/10.1038/s41592-019-0638-x

[22] Zhang, X., Wu, R.T. and Qu, Z.J. (2023) A Cosine-Similarity-Based Deconvolution Method for Analyzing Data-Independent Acquisition Mass Spectrometry Data. *Applied Sciences*, **13**, Article 5969. https://doi.org/10.3390/app13105969

[23] Meyer, J.G. (2021) Deep Learning Neural Network Tools for Proteomics. *Cell Reports Methods*, **1**, Article ID: 100003. https://doi.org/10.1016/j.crmeth.2021.100003

[24] He, Q.Z., *et al.* (2023) Dear-DIA^{XMBD}: Deep Autoencoder Enables Deconvolution of Data-Independent Acquisition Proteomics. *Research* (*Washington, D.C.*), **6**, Article ID: 0179. https://doi.org/10.34133/research.0179