

Research on the Upper Limit of Accuracy for Predicting Theoretical Tandem Mass Spectrometry

Changjiu He, Xiaoyu Wang, Mingming Lyu, Xinye Bian

School of Computer Science and Technology, Shandong University of Technology, Zibo, China

Email: 21505020612@stumail.sdut.edu.cn

How to cite this paper: He, C.J., Wang, X.Y., Lyu, M.M. and Bian, X.Y. (2024) Research on the Upper Limit of Accuracy for Predicting Theoretical Tandem Mass Spectrometry. *Journal of Computer and Communications*, 12, 184-195.

<https://doi.org/10.4236/jcc.2024.123011>

Received: March 4, 2024

Accepted: March 24, 2024

Published: March 27, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In recent years, numerous theoretical tandem mass spectrometry prediction methods have been proposed, yet a systematic study and evaluation of their theoretical accuracy limits have not been conducted. If the accuracy of current methods approaches this limit, further exploration of new prediction techniques may become redundant. Conversely, a need for more precise prediction methods or models may be indicated. In this study, we have experimentally analyzed the limits of accuracy at different numbers of ions and parameters using repeated spectral pairs and integrating various similarity metrics. Results show significant achievements in accuracy for backbone ion methods with room for improvement. In contrast, full-spectrum prediction methods exhibit greater potential relative to the theoretical accuracy limit. Additionally, findings highlight the significant impact of normalized collision energy and instrument type on prediction accuracy, underscoring the importance of considering these factors in future theoretical tandem mass spectrometry predictions.

Keywords

Tandem Mass Spectrometry, Spectral Prediction, Theoretical Limit

1. Introduction

In the field of proteomics research, tandem mass spectrometry has emerged as a key tool for protein identification and structural analysis, attributed to its high resolution and sensitivity. Through the process of ionization, large molecular proteins are subjected to multiple stages of mass spectrometry within the mass spectrometer, resulting in complex and informative experimental spectra, known as tandem mass spectra. These spectra contain fragmented products of proteins,

namely fragment ions, whose characteristics are instrumental in deducing the amino acid sequences, structures, and modifications of proteins.

Interpreting tandem mass spectra to acquire accurate peptide information is a complex task. Each peptide molecule undergoes multiple collisions and fragmentations upon ionization, resulting in intricate fragment ion spectra. Spectra contain two dimensions of information: mass-to-charge ratio (m/z) and intensity. The m/z indicates the ratio of mass to charge of the fragment ions, while the intensity indicates the abundance of ions at a particular m/z . Accurate prediction of theoretical tandem mass spectra is crucial for the correct identification of peptides, posing the challenge of precisely describing the peptide fragmentation patterns and the intensity of the resulting fragment ions. Theoretical tandem mass spectrometry predictions are widely applied in proteomics and biomolecular mass spectrometry for identifying proteins and peptides, analyzing modifications, distinguishing subtypes, and detecting mutations through the simulation of ionization and fragmentation processes.

To date, existing spectrum prediction methods predominantly fall into two categories: those based on statistical physics models and those based on machine learning models. Early statistical physics approaches often utilized the mobile proton model, incorporating assumptions about fragmentation patterns, which led to methods like MS-Simulator [1] [2] and MassAnalyzer [3] [4]. With the growth of data volumes and computational resources, machine learning-based prediction methods have emerged more recently. Among these, PeptideART [5] [6] relies on a two-layer feedforward neural network, while MS2PBPI [7] utilizes gradient boosting regression trees. In contrast, the pDeep [8] [9] [10] series and PredFull [11] employ deep learning network models, with pDeep utilizing bidirectional long short-term memory networks (BiLSTM) for sequence modeling, and PredFull applying convolutional neural networks (CNN). These approaches have explored various techniques and model designs in the realm of theoretical spectrum prediction, significantly contributing to the advancement of the field.

Existing spectrum prediction methods have achieved high accuracy within their respective ion type prediction ranges. For instance, pDeep and Prosit [12] [13] have realized a median Pearson Correlation Coefficient (PCC) greater than 0.980 for b and y backbone ion predictions, while PredFull has achieved a median PCC greater than 0.800 for full spectrum prediction methods. However, a clear standard for defining the upper limit of spectrum prediction accuracy and its potential for improvement remains absent. If current methods are nearing this limit, further research into spectrum prediction may not be necessary. Conversely, this indicates a need for more precise methods and models. Given the close relationship between peptide fragmentation, spectrum detection, mass spectrometry parameters, and instrument types, this study investigates the upper limits of prediction by analyzing backbone and full spectrum ions, taking into account the impact of normalized collisional energy (NCE) and instrument types, thereby unveiling the potential for improvement and key influencing factors in spectrum prediction.

2. Workflow and Method

2.1. Data Preprocessing

This article conducted experiments using publicly available datasets that are widely recognized and employed by existing spectral prediction methods. The instrument setup, mass spectrometry experimental conditions and experimental design in the publicly available data were performed under standardized procedures, and no new mass spectrometry experiments were required to obtain data in order to ensure comparability with existing methods. This article employs five datasets originating from the Kuster, Mann, and Pandey laboratories, each dataset derived from different biological species using various instrument types and analyzed under different NCE settings. Considering that over 80% of tandem mass spectrometry prediction methods target spectra under the High-energy Collisional Dissociation (HCD) [20] fragmentation pattern, this study utilizes spectra data from this specific mode for experimentation. **Table 1** provides detailed information about the datasets, all of which are sourced from the Proteome Xchange mass spectrometry data archive website (<https://proteomecentral.proteomexchange.org/>).

The original raw format files corresponding to the datasets mentioned are downloaded from Proteome Xchange, and the pParse [21] software is employed to extract fragment ion information for each experimental spectrum. During this process, only fragment ion information of each spectrum is retrieved, leaving the peptide identities unknown. To identify peptides, the raw files are searched using a protein search engine. To expedite the search process, identification results provided with the datasets are used for peptide-spectrum matching. For datasets without provided identification results, an open search is conducted using the pFind [22] search engine, with search parameters detailed in the following **Table 2**.

Upon completing peptide-spectrum matching, Xcalibur software is utilized to extract NCE and instrument type information for each spectrum, retaining only those spectra acquired in HCD mode. With the necessary annotation information for the experimental spectra obtained, peptides are used to simulate fragmentation, generating theoretical spectra for the annotation of fragment ion intensities.

Table 1. Dataset information.

Data set	Species	Labs	NCE	instruments
PXD004732 [14]	synthetics	Kuster	20/23/25/28/30/35	Lumos
PXD021013 [15]	HLA	Kuster	25	Lumos
PXD000269 [16]	Yeast	Mann	25	QE
PXD001250 [17]	Mouse	Mann	25/27	QE/QE-HF
PXD000561 [18] [19]	Human	Pandey	27	Elite

Table 2. pFind open search parameter setting.

Parameters	Setting
Precursor tolerance	±20 ppm
Fragment tolerance	±20 ppm
Fixed modification	Carbamidomethyl on C
Variable modification	Oxidation on M
FDR	0.01
Max cleave sites	3

2.2. Ion Statistics and Labelling

This study conducted an analysis of the proportion of backbone ions in the processed experimental spectra, specifically examining the prevalence of a, b, c, x, y, and z ions. According to the findings illustrated in **Figure 1**, y ions were the most prevalent, constituting about half of the total, followed by b ions, which accounted for more than a quarter of the total. Together, these two ion types comprised three-quarters of the total, indicating that b and y ions are the dominant fragment ion types in HCD fragmentation mode. Notably, the proportion of a ions was slightly less than that of b and y ions but was higher than the combined proportions of c, x, and z ions, suggesting that a ions also have a significant abundance in this fragmentation mode.

In addition to the ion types previously mentioned, spectra often contain numerous neutral loss ions, which are typically formed when peptides lose specific groups during fragmentation in mass spectrometers, commonly water (H₂O) or ammonia (NH₃) molecules. These dehydration and deammoniation ions are relatively stable, making them among the most frequent neutral loss ions in spectra. Given their prevalence alongside high-abundance ions, the probability of encountering dehydrated or deammoniated ions is also higher, especially for singly charged ions compared to doubly charged ones. Consequently, the study annotated a total of 18 ion types, including a+, a++, a-H₂O+, a-H₂O++, a-NH₃+, a-NH₃++, b+, b++, b-H₂O+, b-H₂O++, b-NH₃+, b-NH₃++, y+, y++, y-H₂O+, y-H₂O++, y-NH₃+ and y-NH₃++.

2.3. Evaluation Metrics

The evaluation of theoretical tandem mass spectrometry prediction results heavily relies on similarity metrics, which assess the reliability of predictions by comparing the predicted theoretical spectrum ion intensities with those of the experimental spectrum. The PCC is widely recognized for this purpose. In addition to PCC, some methods also use Cosine Similarity (COS), Spearman's Rank Correlation Coefficient (SPC), and other custom criteria to evaluate predictions. Metrics like the mean PCC, median PCC, and the proportions of PCC > 0.75 and PCC > 0.80 are commonly employed to present evaluation results. The formulas for PCC, COS and SPC are listed below.

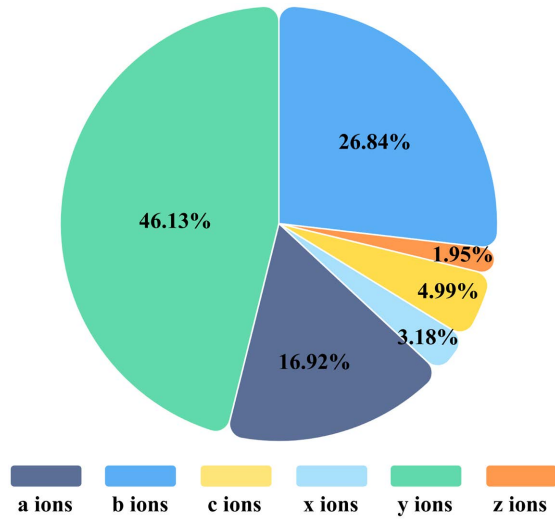


Figure 1. Percentage of 6 backbone ions.

$$PCC = p(\text{Real}, \text{Pred}) = \frac{\sum_{i=1}^n (\text{Real}_i - \overline{\text{Real}})(\text{Pred}_i - \overline{\text{Pred}})}{\sqrt{\sum_{i=1}^n (\text{Real}_i - \overline{\text{Real}})^2} \sqrt{\sum_{i=1}^n (\text{Pred}_i - \overline{\text{Pred}})^2}} \quad (1)$$

$$\text{COS} = \frac{\text{Real} \cdot \text{Pred}}{\|\text{Real}\| \|\text{Pred}\|} = \frac{\sum_{i=1}^n \text{Real}_i \times \text{Pred}_i}{\sqrt{\sum_{i=1}^n (\text{Real}_i)^2} \sqrt{\sum_{i=1}^n (\text{Pred}_i)^2}} \quad (2)$$

$$\text{SPC} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

Here, n represents the number of predicted ions, i.e., the length of the vector. Real_i and Pred_i denote the actual and predicted intensities of the i th ion, respectively. $\overline{\text{Real}}$ and $\overline{\text{Pred}}$ represent the mean intensities of the actual and predicted ion intensity vectors, respectively. d_i is the rank difference between Real_i and Pred_i in their respective sequences. These metrics serve as the standard measures utilized in this article for analyzing the upper limits.

PCC measures the linear correlation between two variables, giving a value between -1 and 1 . Its advantage is in detecting linear relationships, but it might not capture non-linear relationships well. COS measures the cosine of the angle between two vectors, useful in high-dimensional spaces. It's beneficial for comparing orientation but not magnitude, making it suitable for text analysis but less effective when magnitude is important. SPC assesses how well the relationship between two variables can be described using a monotonic function. It's advantageous for non-linear relationships and is not influenced by outliers, unlike PCC. However, it might not be as sensitive as PCC in detecting linear relationships. Moreover, SPC imposes a stricter ranking on longer vectors, resulting in lower scores compared to PCC and COS.

2.4. Precision Analysis Method

In this study, three intensity vector approaches were employed for analyzing the

upper limit of accuracy. The first method focused solely on b and y backbone ions, which are the most abundant ion types in spectra and are fundamental for all theoretical tandem mass spectrometry prediction methods. Although Zhou *et al.* [8] analyzed the accuracy upper limit of b and y ions, they only considered the PCC metric and suggested there was room for improvement in accuracy. The second approach analyzed using the aforementioned 18 ion types, indicating that b and y ions alone are insufficient for representing the entire spectrum, as they account for only half of the total spectral intensity. The final method adopted the preprocessing technique used by PredFull, scaling the m/z ratio by a factor of 10, rounding it, and representing the entire spectrum in a 20,000 dimensional vector. This approach, not relying on specific ion types, incorporates all ion types into the vector, with each index representing m/z and scalar values indicating intensity at those positions.

For the first two methods, ions specified for each spectrum are extracted and represented in an intensity vector in the order of b, y, or a, b, y ions, with ion intensities defaulting to 0.0 for absent peaks. The third approach directly utilizes the vector processed by PredFull's preprocessing method as the intensity vector. Typically, a peptide corresponds to multiple experimental spectra. Assuming a peptide fragments into K spectra, the total number of repeated spectrum pairs is calculated as shown in formula (4). Similarity analysis of these spectrum pairs' vectors is then performed to determine the theoretical upper limit of prediction accuracy for repeated spectra.

$$S = \sum_{i=1}^{K-1} i \quad (4)$$

3. Analysis and Results

3.1. Upper Limit of Ionisation of the Backbone

The study employed two approaches for analyzing the upper limit of backbone ions. The first focused on the accuracy limit for b and y backbone ions, and the second involved the 18 types of backbone ions previously annotated. Experiments were conducted using the PXD004732 dataset to ensure consistency across other parameters, with nearly 1.5 million spectrum pairs collected from data at NCE 25 for analysis. The evaluation utilized multiple metrics, including PCC, COS, and SPC, to assess performance.

As shown in **Figure 2**, box plot a represents the results of the experimental analysis using only the intensity of b and y ions, and the median values of the three metrics are 0.996, 0.997, and 0.964, respectively. whereas box plot b represents the results of the experimental analysis of the 18 ions, and the median values of the three metrics in the results are 0.996, 0.996, and 0.893, respectively. the analysis reveals that the distribution of box plots of the PCC and COS is more centralised, which indicates that the similarity value is higher under this indicator, while in SPC the distribution is discrete and the accuracy of the 18 ions decreases compared to the b and y ions. As shown in **Figure 3**, by analysing the proportion of each indicator above the critical value, it can be seen that for

PCC and COS there is almost no great difference, in these two indicators more than 99% of the spectrogram pairs can reach a similarity of more than 0.900, with the increase in the number of ions the vector length increases the SPC indicators have a greater impact, but in general, almost all the spectrograms of the pairs of similarity is more than 0.70.

This study conducted a comparative analysis between two classical theoretical mass spectrometry prediction methods, pDeep and Prosit, which are both limited to predicting b and y ions. Predictions were made using both methods on the same dataset, with relevant metrics presented in **Table 3**. The analysis revealed that Prosit generally outperformed pDeep, although Prosit was slightly inferior to pDeep in terms of the SPC metric. Despite achieving high prediction accuracy, both methods still have considerable room for improvement compared to the accuracy upper limit analysis presented in this paper, especially in metrics above 0.90.

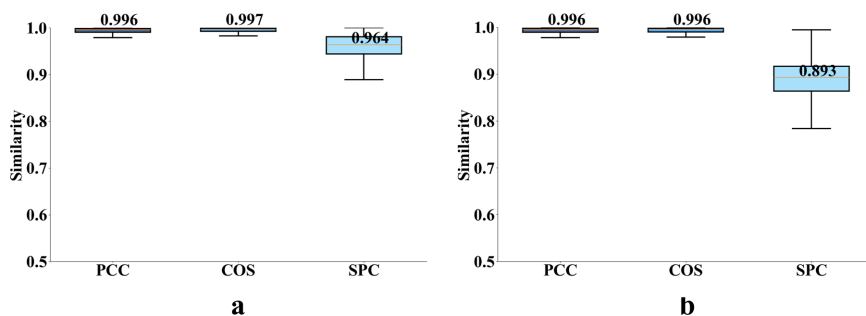


Figure 2. Distribution of similarity indicators.

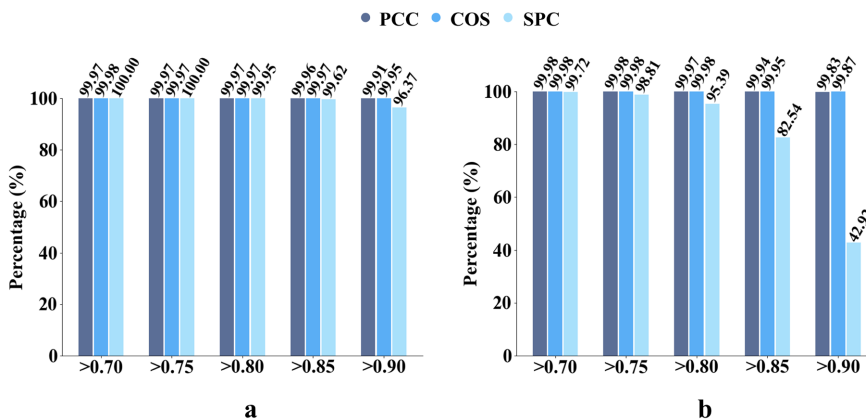


Figure 3. Percentage of each similarity indicator.

Table 3. Upper limit of prediction accuracy for different methods.

Methods	PCC median	COS median	SPC median	PCC > 0.90	COS > 0.90	SPC > 0.90
pDeep	0.984	0.987	0.933	82.20	88.02	68.94
Prosit	0.989	0.991	0.928	96.63	97.54	67.27
Our Method	0.996	0.997	0.964	99.91	99.95	96.37

3.2. Full Spectrum Upper Limit Analysis

Spectra encompass not only backbone ions but also numerous non-backbone ions, precursor ions, and noise ions. Noise ions generate peaks that significantly interfere with spectral analysis. For the PredFull prediction method, the presence of numerous noise ions can disrupt results, reducing prediction accuracy. In an experiment analyzing the upper limit of full spectrum accuracy using the same dataset as the backbone ion accuracy limit experiment, 1.2 million spectrum pairs were examined. The medians of the three metrics were 0.930, 0.930, and 0.656, respectively (as shown in **Figure 4**). The decrease in accuracy compared to the backbone ion accuracy limit is primarily due to the increased number of ions, including more noise ions, which significantly impacts accuracy. An analysis of the critical values of these metrics showed a consistent trend, with a 0% proportion of SPC > 0.90, indicating that the SPC metric, which has stricter rank correlation for longer vectors, loses accuracy with an increased number of ions.

This paper compares the PredFull method as shown in **Table 4** below. In each median index, PredFull is about 0.1 - 0.2 lower, and at the same time, it performs poorly in each index of >0.90, which is only about 1/6 of the upper limit value. This shows that the PredFull method for predicting the full spectrum still has a very large room for improvement in accuracy, and there is still room for further exploration for full spectrum prediction.

3.3. NCE Analysis

NCE is a crucial parameter in mass spectrometry that directly influences the degree of peptide fragmentation. Typically, as NCE increases, peptides fragment more thoroughly, resulting in spectra with higher peak signal intensities. This is because higher NCE levels cause more peptide bonds to break, generating more fragment ions. Conversely, at lower NCE levels, insufficient energy may not induce peptide bond breakage, leading to lower coverage of fragment ions in the spectra. Therefore, to assess the impact of NCE on spectral prediction, this study conducted an upper limit accuracy analysis experiment on repeated spectra under different NCE settings.

The experiments utilized the PXD004732 dataset, which fragments peptides at six NCEs: 20, 23, 25, 28, 30, and 35. Repeated spectrum pairs were identified between two NCE settings, resulting in fifteen different NCE combinations, with an average of 300,000 spectrum pairs per combination. The analysis focused on b and y ions, and the results are displayed in **Figure 5** through three heatmaps representing the median distributions for PCC, COS, and SPC, with the lower diagonal indicating the upper limit of accuracy at identical NCEs. The findings reveal a direct correlation between the proximity of NCE values and spectral similarity: closer NCE values yield higher similarity. Specifically, a 2 unit difference in NCE maintains similarity; however, a 5 unit difference reduces median similarity by approximately 0.05, and a 10 unit difference decreases it by about 0.25, underscoring the significant impact of NCE variation, especially larger differences, on peptide fragmentation.

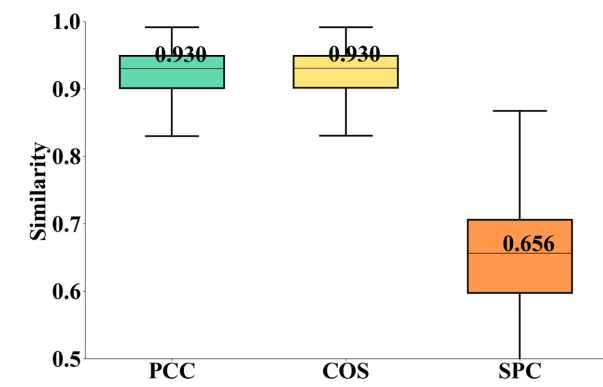


Figure 4. Distribution of similarity indicators.

Table 4. Comparison of PredFull prediction accuracy upper limits.

Methods	PCC median	COS median	SPC median	PCC > 0.90	COS > 0.90	SPC > 0.90
PredFull	0.812	0.813	0.462	12.81	12.97	0.00
this paper	0.930	0.930	0.656	75.54	75.91	0.00

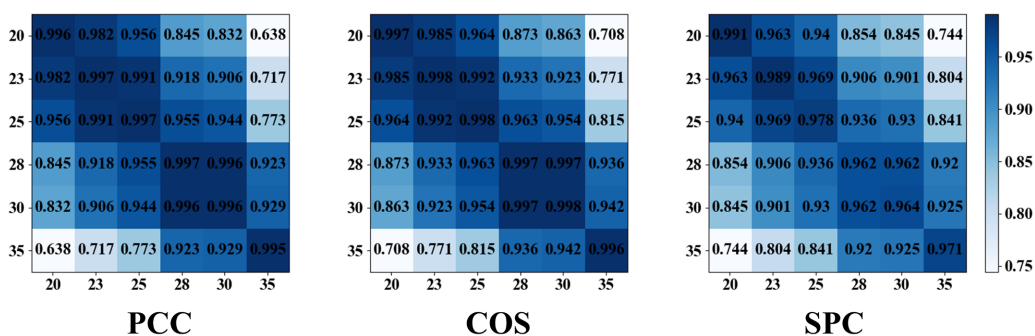


Figure 5. Differences in b/y ion similarity at different energies.

3.4. Instrument Type Analysis

While existing methods incorporate instrument type, a deeper analysis of its specific effects is necessary. The PredFull method, when training the HCD model, did not differentiate between instrument types, observing that spectra from different instruments were highly similar. This study investigates the impact of instrument type on peptide fragmentation by analyzing repeated spectra of the same peptide across different instruments. To minimize confounding variables, datasets with the same NCE and closely related species, including Lumos, QE, QE-HF, and Elite, were collected for analysis at NCEs 25 and 27. Similarity was calculated using intensity vectors composed of 18 ion types across different instruments at the same NCE.

The designed comparative experiments included: 1) a comparison between the PXD001250 dataset generated by QE-HF and the PXD000561 dataset generated by Elite (NCE = 27); 2) a comparison between the PXD001250 dataset generated by QE and the PXD004732 dataset generated by Lumos (NCE = 25); and 3 & 4) comparisons within the same instrument type to demonstrate the internal

Table 5. PCC index of repeated spectra with different instruments.

Methods	PCC median	PCC mean	PCC > 0.75	PCC > 0.80	PCC > 0.85	PCC > 0.90
Experiment 1	0.919	0.873	87.95	83.13	75.90	61.45
Experiment 2	0.908	0.881	89.96	83.17	71.66	54.16
Experiment 3	0.984	0.957	96.22	95.44	94.25	91.84
Experiment 4	0.988	0.982	100.00	99.74	99.48	98.70

consistency of QE and Lumos instruments. Results indicated significant differences in PCC values across different instruments (as seen in **Table 5**), with higher similarity observed within the same instrument type, suggesting that the fragmentation patterns of the same peptide vary across different instruments.

4. Discussion

The existing theoretical tandem mass spectrometry spectrum prediction methods have achieved a certain level of accuracy, and the prediction of b and y backbone ions is approaching the theoretical limit. However, for other ion types, not only a lack of prediction methods, but there is also a significant gap between the prediction accuracy and the similarity limit analyzed in this study. Especially concerning the problem of full spectrum prediction, there is still considerable room for improvement in prediction accuracy. Additionally, regarding the NCE and instrument type, our analysis reveals that NCE has a significant impact on experimental spectra, with increasing NCE differences leading to a gradual decrease in similarity. While the impact of instrument type on peptide fragmentation may not be as significant as NCE, it is still noteworthy, especially as prediction accuracy approaches the theoretical limit. Therefore, in the process of developing new theoretical tandem mass spectrometry prediction methods in the future, it is essential not only to further enhance prediction accuracy but also to comprehensively consider various characteristics (NCE and Instruments), thereby providing more accurate prediction methods for the analysis of mass spectrometry data.

Conflicts of Interest

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

References

- [1] Sun, S.W., Yang, F., Yang, Q., *et al.* (2012) MS-Simulator: Predicting Y-Ion Intensities for Peptides with Two Charges Based on the Intensity Ratio of Neighboring Ions. *Journal of Proteome Research*, **11**, 4509-4516. <https://doi.org/10.1021/pr300235v>
- [2] Wang, Y.J., Yang, F., Wu, P., *et al.* (2015) OpenMS-Simulator: An Open-Source Software for Theoretical Tandem Mass Spectrum Prediction. *BMC Bioinformatics*, **16**, Article No. 110. <https://doi.org/10.1186/s12859-015-0540-1>

- <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0540-1>
- [3] Zhang, Z.Q. (2004) Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Analytical Chemistry*, **76**, 3908-3922. <https://doi.org/10.1021/ac049951b>
- [4] Zhang, Z.Q. (2005) Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides with Three or More Charges. *Analytical Chemistry*, **77**, 6364-6373. <https://doi.org/10.1021/ac050857k>
- [5] Arnold, R., *et al.* (2006) A Machine Learning Approach to Predicting Peptide Fragmentation Spectra. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*, **11**, 219-230.
- [6] Li, S.J., *et al.* (2011) On the Accuracy and Limits of Peptide Fragmentation Spectrum Prediction. *Analytical Chemistry*, **83**, 790-796. <https://doi.org/10.1021/ac102272r>
- [7] Dong, N.-P., Liang, Y.-Z., Xu, Q.-S., *et al.* (2014) Prediction of Peptide Fragment Ion Mass Spectra by Data Mining Techniques. *Analytical Chemistry*, **86**, 7446-7454. <https://doi.org/10.1021/ac501094m>
- [8] Zhou, X.-X., Zeng, W.-F., Chi, H., *et al.* (2017) pDeep: Predicting MS/MS Spectra of Peptides with Deep Learning. *Analytical Chemistry*, **89**, 12690-12697.
- [9] Zeng, W.-F., Zhou, X.-X., Zhou, W.-J., *et al.* (2019) MS/MS Spectrum Prediction for Modified Peptides Using pDeep2 Trained by Transfer Learning. *Analytical Chemistry*, **91**, 9724-9732. <https://doi.org/10.1021/acs.analchem.9b01262>
- [10] Tarn, C. and Zeng, W.F. (2021) pDeep3: Towards More Accurate Spectrum Prediction with Fast Few-Shot Learning. *Analytical Chemistry*, **93**, 5815-5822. <https://doi.org/10.1021/acs.analchem.0c05427>
- [11] Liu, K.Y., *et al.* (2020) Full-Spectrum Prediction of Peptides Tandem Mass Spectra using Deep Neural Network. *Analytical Chemistry*, **92**, 4275-4283. <https://doi.org/10.1021/acs.analchem.9b04867>
- [12] Gessulat, S., Schmidt, T., Zolg, D.P., *et al.* (2019) Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning. *Nature Methods*, **16**, 509-518. <https://doi.org/10.1038/s41592-019-0426-7>
<https://www.nature.com/articles/s41592-019-0426-7>
- [13] Ekvall, M., *et al.* (2022) Prosit Transformer: A Transformer for Prediction of MS2 Spectrum Intensities. *Journal of Proteome Research*, **21**, 1359-1364. <https://doi.org/10.1021/acs.jproteome.1c00870>
- [14] Zolg, D.P., Wilhelm, M., Schnatbaum, K., *et al.* (2017) Building ProteomeTools Based on a Complete Synthetic Human Proteome. *Nature Methods*, **14**, 259-262. <https://doi.org/10.1038/nmeth.4153>
<https://www.nature.com/articles/nmeth.4153>
- [15] Wilhelm, M., Zolg, D.P., Graber, M., *et al.* (2021) Deep Learning Boosts Sensitivity of Mass Spectrometry-Based Immunopeptidomics. *Nature Communications*, **12**, Article No. 3346. <https://doi.org/10.1038/s41467-021-24263-w>
<https://www.nature.com/articles/s41467-021-23713-9>
- [16] Kulak, N.A., *et al.* (2014) Minimal, Encapsulated Proteomic-Sample Processing Applied to Copy-Number Estimation in Eukaryotic Cells. *Nature Methods*, **11**, 319-324. <https://doi.org/10.1038/nmeth.2834>
<https://www.nature.com/articles/nmeth.2834>
- [17] Sharma, K., Schmitt, S., Bergner, C.G., *et al.* (2015) Cell Type- and Brain Region-Resolved Mouse Brain Proteome. *Nature Neuroscience*, **18**, 1819-1831.

- <https://doi.org/10.1038/nn.4160>
<https://www.nature.com/articles/nn.4160>
- [18] Pinto, M.P., Manda, S.S., Kim, M.-S., *et al.* (2014) Functional Annotation of Proteome Encoded by Human Chromosome 22. *Journal of Proteome Research*, **13**, 2749-2760. <https://doi.org/10.1021/pr401169d>
<https://pubs.acs.org/doi/10.1021/pr401169d>
- [19] Kim, M.S., Pinto, S.M., Getnet, D., *et al.* (2014) A Draft Map of the Human Proteome. *Nature*, **509**, 575-581. <https://doi.org/10.1038/nature13302>
<https://www.nature.com/articles/nature13302>
- [20] Olsen, J.V., Macek, B., Lange, O., *et al.* (2007) Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nature Methods*, **4**, 209-712. <https://doi.org/10.1038/nmeth1060>
<https://www.nature.com/articles/nmeth1060>
- [21] Yuan, Z.F., Liu, C., Wang, H.-P., *et al.* (2012) pParse: A Method for Accurate Determination of Monoisotopic Peaks in High-Resolution Mass Spectra. *Proteomics*, **12**, 226-235. <https://doi.org/10.1002/pmic.201100081>
<https://pubmed.ncbi.nlm.nih.gov/22106041/>
- [22] Chi, H., Liu, C., Yang, H., *et al.* (2018) Comprehensive Identification of Peptides in Tandem Mass Spectra Using an Efficient Open Search Engine. *Nature Biotechnology*, **36**, 1059-1061. <https://doi.org/10.1038/nbt.4236>
<https://www.nature.com/articles/nbt.4236>