MSPoisDM: A Novel Peptide Identification Algorithm Optimized for Tandem Mass Spectra

Simin Zhu, Chuanjie Yang, Wenya Wu

China Fire and Rescue Institute, Beijing 102202, China

Abstract. Tandem mass spectrometry (MS/MS) plays an extremely important role in proteomics research. Thousands of spectra can be generated in modern experiments, how to interpret the LC-MS/MS is a challenging problem in tandem mass spectra analysis. Our peptide identification algorithm, MSPoisDM, is integrated the intensity information which produced by target-decoy statistics, although intensity information often undervalued. Furthermore, in order to combine the intensity information for better, we propose a novel concept scoring model which based on Poisson distribution. Compared with commonly used commercial software Mascot and Sequest at 1% FDR, the results show MSPoisDM is robust and versatile for various datasets which obtained from different instruments. We expect our algorithm MSPoisDM will be broadly applied in the proteomics studies.

Keywords: Index-Terms Tandem Mass Spectrum, Protein Identification Algorithms, Weight of Pair Amino Acid Fracture.

1. Introduction

In biological sample analysis, mass spectrometry (MS)-based proteomics has evolved into an indispensable approach [7, 11]. In proteomics experiment, proteins can be cleaved into peptides by enzyme-selected, then separated and enter MS for subsequently analyzing [7, 12]. Thousands of fragmentation spectra generated in modern proteomics experiment, how to infer peptide sequence is a challenging and propose peptide identification algorithms are necessity [1, 12].

Algorithms model greatly affect the efficient and accuracy of searching spectra [10]. Scoring function is the core of protein identification algorithms, integrated the current algorithms types, they could be divided into four categories as follows [14, 16, 19].

(1) Correlation matching model: by mathematical simulation of protein digestion and mass spectrometry detection process, the theory of enzyme peptides transform into the corresponding predicted spectrum, then the degree of correlation between predicted spectrum and experimental spectrum needed to be evaluated by mathematical approach, obtained the appropriate search results. Representative algorithms such as Sequest [6]and pFind [12].

(2) Probabilistic matching model: utilizing the statistical probability which obtained by census the frequency of a certain m/z value in a certain error range in protein database to indicate the credibility of matching [15, 17], then constructed a reasonable algorithm model, searching the correct peptides. Representative algorithms such as Mascot, OMSSA [2], X!Tandem [4], Andromeda and ProVerB [18].

(3) Random matching model: using the information of proteins mass distribution and theory of enzyme peptides mass distribution, the corresponding spectrum was divided into several intervals, by calculating random probability of each section matching to the selected peptides, then built the identification algorithms model. Representative algorithms such as SCOPE [3] and Probity [8].

(4) Empirical weight matching model: by assigned different empirical weights to key ions, consecutive occurrence, intensities, pair-wise amino acid patterns and ect. Representative algorithms such as MassWiz [19] and SQID [13].

Integrated the characterizes information to scoring algorithms was inevitable, which could improve the confidence of searching results and efficiency. m/z value often as the main characterize information to be assigned the mainstream search engines, contained Mascot, Seuqest, X!Tandem and OMSSA. Peak intensity characterizer often undervalued because of its unreliability, in order to integrate the intensity information, favourable peaks selecting manner was imperative, SQID and ProVerB utilized diverse manner to select efficiency peaks, and studied the fragmentation intensity patterns, built the protein identification algorithms, achieved excellent results. Dispec proposed a novel concept characterize information based on peptide

matching discriminability (PMD) [1], which abundant reflected the properties of experimental spectrum. Hence, appended more luxuriant characterize information would improve the efficiency and reliable of identification results [6].

In this paper, we proposed a novel peptide identification algorithm, named MSPoisDM, which integrated a brand-new concept characterize information Peak Intensity Identification-ability (PII). PII measured the degree of real matching, meantime, we built a novel scoring model for adding PII information. To validate the reliability and accuracy of MSPoisDM, we utilized diverse datasets from various mass spectrometer platforms to test, compared with Mascot and Sequest at 1% FDR level [9], MSPoisDM showed more robust and higher identification.

2. Methods

2.1 MSPoisDM Identification Algorithm

MSPoisDM, which in virtue of Poisson distribution to construct a novel scoring model and consider to add the PII information. We adopted Matlab (version: 8.1.0.604. (R2013a)) as the programming language. How to via the training experimental spectra to obtain PPI characterize information was crucial for our identification algorithm. Here, we through the following aspects for introducing the algorithm designing process.

(1) Isotopes discarding: plentiful isotopes exist in nature, so experimental spectra did. The existence of isotope peaks led to more random matching instead of real matching, the key of discarding isotopes was that correctly judged isotope peaks or not. The specific rules as following: if the two peaks closer than $1\pm 0.25\ Da$ were considered as isotope peaks and the lower intensity peak needed to be discarded. This treatment could reduce random matches and enhance the accuracy.

(2) Peaks selecting: different peptide identification algorithms had diverse manner to select efficient peaks. Sequest and SQID selected the strongest 200 and 80 peaks in each experimental spectra respectively; OMSSA divided the spectra into several bins and then selected the top 5 peaks in each bin. MS Amanda selected the m most intense peaks in each $100\ Da$. In this article, we adopted dynamic approach which had been reported by ProVerB to select peaks, MSPoisDM selected the top 6 peaks in $100\ Da$ window.

(3) Extracting PPI characterize information: integrated abundant characterizes information to enhance the accuracy of scoring algorithms were necessary. PPI was a measure of real matching or not, and the specific extraction process included three aspects:

(a) Training dataset: utilized the rational datasets for training was extremely important. The training spectra of MSPoisDM was extracted from the identified spectra which Mascot, Sequest and ProVerB all identified and be controlled by 1% FDR level. Hence, we considered the training spectra were collect identified.

(b) Statistical method: different statistical method generated various results, in order to obtain PII characterize information, statistical process comprised three aspects: firstly, confirm key ion type, here we only defined b, $b-NH_3$, $b-H_2O$, y, $y-NH_3$, $y-H_2O$ as key ion type; secondly, divided the peak intensities which had been normalized into 12 intervals, the details showed in table 1. Meantime, the method of normalization just like the following formula:

$$I_{Normalization} = I_{Original} \Big/ I_{Max} \quad (1)$$

Where I_{Max} was the mean of the top three peaks, enhance the reliability of the statistical method. Third, searching the training data set based on forward and reverse reference sequence respectively, recorded the matching results.

(c) Quantitative mathematical: quantified the statistical results involved on the above was crucial for MSPoisDM, we adopted the following formula to quantify, which not only retained variation, but also made the quantitative results get better smoothness. Specific process as follows:

$$PPI(i,j) = \frac{N(F,i,j) \Big/ N(R,i,j)}{\sum_{i=1}^{6}\sum_{j=1}^{12} N(F,i,j) \Big/ N(R,i,j)} \quad (2)$$

Where i denoted the sum of key ion types; j denoted the sum of the number of intensity interval; F denoted the forward reference sequence; R denoted the reverse reference sequence; $N(F,i,j)$ was the number of fragment ion matches which ion type was i and intensity value located at j interval based on forward sequence; $N(R,i,j)$ was the number of fragment ion matches which ion type was i and intensity value located at j interval based on reverse sequence; $PPI(i,j)$ reflected the degree of real matching which ion type was i and intensity value located at j interval. Table 1 was the calculated PPI value.

(4) False Discovery Rate (FDR) calculated: no search engine could ensure all identified results were correct, the peptide spectrum matches (PSMs) were exported to calculate the FDR threshold. We used our in-house Matlab code to extract Mascot and MSPoisDM output result files which peptide length ≥ 6, Sequest results were extracted from output files which PSMs with the highest rank and $\Delta Cn \geq 0.1$, meantimem, the peptide length ≥ 6. And the FDR was calculated by Kall's menthod, respectively. The specific formula as follows:

$$FDR = \frac{no.\ of\ decoy\ PSMs\ above\ threshold\ value}{no.\ of\ target\ PSMs\ above\ threshold\ value} \quad (3)$$

(5) Scoring algorithm: scoring function is the heart of the peptide identification algorithms. In this article, firstly, we considered three-dimensional characterizes into MSPoisDM, contained fragment matches, consecutive fragment matches and b/y-ion matches [1]; secondly, constructed the scoring function based on Poisson distribution, respectively; finally, integrated PPI

characterize information into scoring model. Specific details as follows:

(a) Fragment matches: proposed a universal scoring function for various strategies is hard. We solved the problem by utilizing Poisson distribution to build appropriate function. the formula of Poisson distribution as below:

$$P_0\{X = K_0\} = \lambda_0^{K_0} \cdot e^{-\lambda_0} \Big/ K_0! \tag{4}$$

Where K_0 reflected the number of fragment matches; P_0 reflected the probability of K_0 matches, which embodied the confidence of fragment matches; λ_0 reflected the theoretical mean of fragment matches, and the value of λ_0 could be calculated from the following formula.

$$\lambda_0 = 0.06 N_0 \tag{5}$$

Where 0.06 reflected the random match probability, because of we selected the top 6 peaks from each 100 Da in the experimental spectra; N_0 reflected the number of theoretical matches. Meantime, when $\varepsilon_0 = [\lambda_0]$, the probability of Poisson distribution arrived maximum.

$$P_{0\_\max}\{X = \varepsilon_0\} = \lambda_0^{\varepsilon_0} \cdot e^{-\lambda_0} \Big/ \varepsilon_0! \tag{6}$$

The preliminary score of fragment matches calculated from the following formula.

$$S_{0\_Pre} = \mathrm{sgn}(K_0 - \varepsilon_0) \cdot \lg \frac{P_{0\_\max}}{P_0} \tag{7}$$

Table 1. PPI Value, c omposed by various key ion type and intensity interval

	[0,0.05)	[0.05,0.1)	[0.1,0.2)	[0.2,0.3)	[0.3,0.4)	[0.4,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1)	[1,+∞)
b	0.90	1.08	1.44	2.04	2.59	3.11	3.53	3.86	4.09	4.18	4.34	4.34
b-NH_3	0.74	1.00	1.29	1.56	1.77	1.91	1.90	1.98	2.01	2.00	1.97	1.72
b-H_2O	1.93	2.41	2.94	3.46	3.81	3.97	3.96	3.99	4.03	4.03	3.85	2.76
y	4.04	5.05	6.77	10.00	13.96	18.38	21.13	23.90	27.09	30.21	29.25	32.39
y-NH_3	3.53	3.37	3.06	2.60	2.22	1.97	1.76	1.55	1.43	1.33	1.29	1.12
y-H_2O	1.11	0.98	0.74	0.53	0.42	0.34	0.31	0.28	0.26	0.25	0.24	0.24

Where S_{0_Pre} is the preliminary score of fragment matches, the function showed the more fragment matched, the higher score obtained. In order to integrate the PPI information, we needed to re-scored the fragment matches.

$$S_{0\_Final} = \begin{cases} S_{0\_Pre} \times e^{\sum_{i=1}^{K_0} PPI_i}, & S_{0\_Pre} \geq 0 \\ S_{0\_Pre} \times e^{-\sum_{i=1}^{K_0} PPI_i}, & S_{0\_Pre} < 0 \end{cases} \tag{8}$$

Where S_{0_Final} was the final score of fragment matches, $e^{\sum_{i=1}^{K_0} PPI_i}$ and $e^{-\sum_{i=1}^{K_0} PPI_i}$ reflected the confidence of the matching efficient.

(b) Consecutive fragment matches: consecutive fragment matches characterize information was hard to integrate into scoring algorithms, we via to take fragment matches and consecutive fragment matches as two independent information and scored separately, which could improve the efficiency of MSPoisDM. If termed two fragment matches as consecutive matches, they must satisfy two conditions: belonged to the same ion-type and the differ just equal the mass of a residue. And the scoring process as follows:

$$P_1\{X = K_1\} = \lambda_1^{K_1} \cdot e^{-\lambda_1} \Big/ K_1! \tag{9}$$

Where K_1 denoted the number of consecutive fragment matches; P_1 reflected the probability of K_1 consecutive fragment matches; λ_1 reflected the theoretical mean of consecutive fragment matches and calculated from the following method:

$$\lambda_1 = 0.09083 \cdot K_0 \Big/ N_0 \cdot N_1 \tag{10}$$

Where $0.09083 \cdot K_0 \Big/ N_0$ reflected the random consecutive match probability, which reported by ProVerB; N_1 reflected the number of theoretical consecutive matches. When $\varepsilon_1 = [\lambda_1]$, the probability arrived at maximum.

$$P_{1\_\max}\{X = \varepsilon_1\} = \lambda_1^{\varepsilon_1} \cdot e^{-\lambda_1} \Big/ \varepsilon_1! \tag{11}$$

Like the scoring strategy of fragment matches, here, we also needed to calculate the preliminary of the consecutive matches. The specific process as follows:

$$S_{1\_Pre} = \mathrm{sgn}(K_1 - \varepsilon_1) \cdot \lg \frac{P_{1\_\max}}{P_1} \tag{12}$$

Where S_{1_Pre} is the preliminary score of the consecutive fragment matches. Then the final scoring strategy which integrated PPI information as following formula:

$$S_{1\_Final} = \begin{cases} S_{1\_Pre} \times e^{\sum_{j=1}^{K_1} II_j}, & S_{1\_Pre} \geq 0 \\ S_{1\_Pre} \times e^{-\sum_{j=1}^{K_1} II_j}, & S_{1\_Pre} < 0 \end{cases} \tag{13}$$

If two matches were a couple of consecutive match, and corresponding PPI value were PPI_s and PPI_t respectively, and the value of II_j could be calculated by following formula:

$$II_j = PPI_s + PPI_t \tag{14}$$

(c) b/y-ion matches: b/y-ion were the mainstream ion type under CID environment. Evaluated the efficiency of b/y-ion matches could improve the robust and accuracy of peptide identification algorithm. Hence, we considered the b/y-ion matches as the independent information into scoring model.

$$P_2\{X = K_2\} = \lambda_2^{K_2} \cdot e^{-\lambda_2} / K_2! \quad (15)$$

Where K_2 reflected the number of b/y-ion matches; P_2 reflected the probability of K_2 b/y-ion matches; λ_2 reflected the theoretical mean of b/y-ion matches and calculated from the following method:

$$\lambda_2 = 0.02 \cdot N_2 \quad (16)$$

Where 0.02 was the b/y-ion random probability, N_2 was the number of the theoretical b/y-ion matches. When $\varepsilon_2 = [\lambda_2]$, the probability obtained the maximum value.

$$P_{2\_\max}\{X = \varepsilon_2\} = \lambda_2^{\varepsilon_2} \cdot e^{-\lambda_2} / \varepsilon_2! \quad (17)$$

Similarly, the preliminary score of b/y-ion matches could be obtained by the following formula:

$$S_{2\_Pre} = \mathrm{sgn}(K_2 - \varepsilon_2) \cdot \lg \frac{P_{2\_\max}}{P_2} \quad (18)$$

Meantime, the final score of b/y-ion matches as follows:

$$S_{2\_Final} = \begin{cases} S_{2\_Pre} \times e^{\sum_{k=1}^{K_2} PPI_{k,b/y}}, & S_{2\_Pre} \geq 0 \\ S_{2\_Pre} \times e^{-\sum_{k=1}^{K_2} PPI_{k,b/y}}, & S_{2\_Pre} < 0 \end{cases} \quad (19)$$

Where S_{2_Final} was the score of b/y-ion mates.

(d) The score of candidate peptide: the score of candidate peptide could be calculated by the following formula:

$$S = S_{0\_Final} + S_{1\_Final} + S_{2\_Final} \quad (20)$$

Where S was the score of candidate peptide, it measured the similarly degree between experimental spectra and theoretical spectra. The highest score of candidate peptides was treated as the final searching result.

2.2 MS/MS Datasets and Search Engine

We utilized various data sets which based on different instrument platforms. Standard mixtures of 18 proteins obtained from four types of MS instruments, included Thermo Finnigan LTQ-FT, Thermo Finnigan LCQ DECA, Thermo Finnigan LTQ and Micromass/Waters QTOF Ultima. In order to narrate convenience, we abbreviated the names of mentioned above as FT, LCQ, LTQ and QTOF, and public download website is https://regis-web.systemsbiology.net//PublicData sets/. The public data sets of E.coli downloaded from http://macrottelab.org/MSdata/Data_03/, which contained three sub-datasets, named E.coli1, E.coli2 and E.coli3. S.pneumoniae D39 dataset based on LTQ-Orbitrap was obtained from http://bioinformatics.jnu.edu.cn /software/proverb/, not only served as the training data set for extracting PPI characterize information, but also utilized as test data set; the data set of yeast was obtained from ophthalmology central of Sun Yat-Sen university, the data set was generated by HCD.

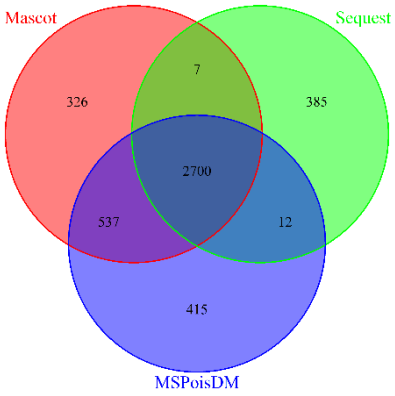
Mascot and Sequest were widely used in proteomics research, which were adopted to compare with MSPoisDM. The version of Mascot and Sequest were 2.3 and 28.13 respectively. When used Mascot engine to search, the dta format files transferred into Mascot generic format (Mgf) files by merge.pl which was download from Mascot official web. Dta format files as input for Sequest and MSPoisDM. In addition, searching criteria were applied for Mascot, Sequest and MSPoisDM, specific contained cysteine (+57.021464 Da, Carbamidomethylation) defined as fixed modification, methionine (+15.994915 Da, oxidation) defined as variable modification and full tryptic specificity. Other parameters were set in table 2.

Table 2. Parameters of precursor and fragment ion

Instruments	Mascot and MSPoisDM		Sequest	
	PIT	FIT	PIT	FIT
LCQ_Deca	3.0 Da	0.5 Da	3.0 Da	1.0 Da
LTQ	3.0 Da	0.5Da	3.0 Da	1.0 Da
LTQ-FT	10 ppm	0.5 Da	10 ppm	1.0 Da
QTOF	0.2 Da	0.5 Da	0.2 Da	1.0 Da
LTQ-Orbitrap	10 ppm	0.5 Da	10 ppm	1.0 Da

3. Results

MSPoisDM was compared with Mascot and Sequest after FDR calculation, the data set of S.pneumoniae D39 showed MSPoisDM & Mascot had higher overlap than MSPoisDM & Sequest. The overlap of MSPoisDM &Mascot was 88.3%, but MSPoisDM & Sequest only was 74.0%. Figure 1 showed the overlap between the two from Mascot, Sequest and MSPoisDM.



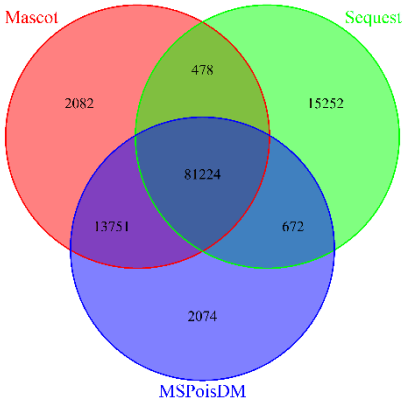


Fig. 1 Overlpa between the two from Mascot, Sequest and MSPoisDM based on S.pneumoniae D39 data set; (Left) the number of identified peptides, (Right) the number of identified sepctra

The data sets of standard mixtures of 18 proteins, which instruments contained FT, QTOF, LCQ and LTQ. MSPoisDM identified peptides was the most of the algorithms which mentioned above from the instruments except LTQ, showed its robust and steady. Meantime, MSPoisDM identified more spectra than Sequest from any MS instrument. Figure 2 and Figure 3 were the identified peptides and spectra from the four instruments mentioned above respectively.

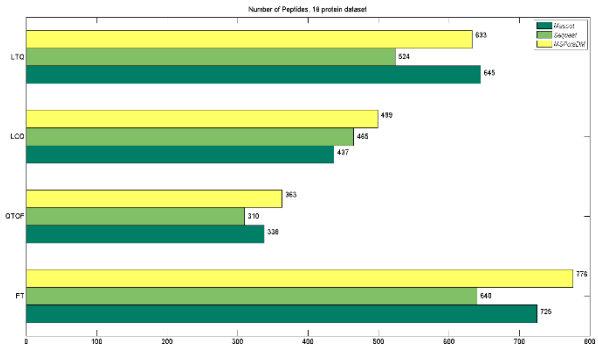


Fig. 2 Identified peptides from 18 protein data sets of Mascot, Sequest and MSPoisDM

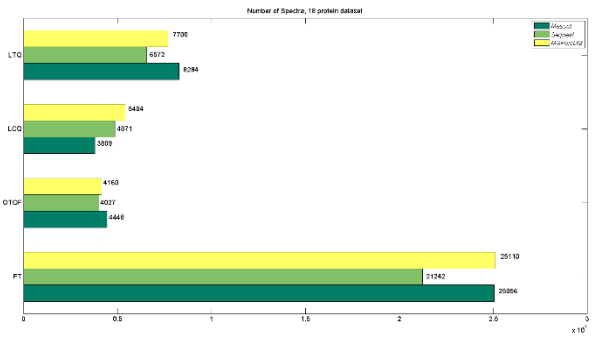


Fig. 3 Identified spectra from 18 protein data sets of Mascot, Sequest and MSPoisDM

The data sets of E.coli, which contained three subsets E.coli1, E.coli2 and E.coli3, the identified peptides and spectra from E.coli data sets were the most of all the three search engines, showed MSPoisDM had high identified and superiority. Specify details revealed in Figure 4 and Figure 5.

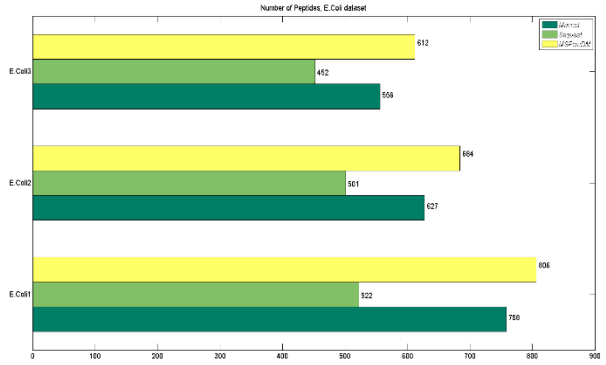


Fig. 4 Identified peptides from E.coli data sets of Mascot, Sequest and MSPoisDM

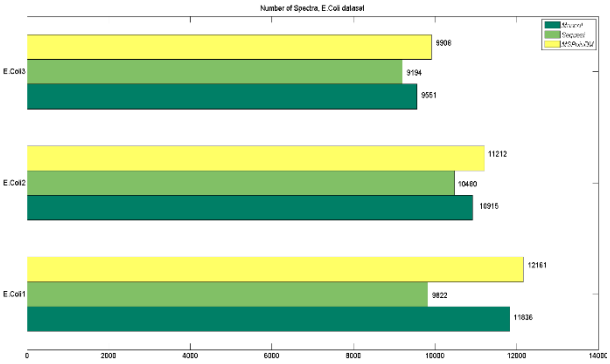


Fig. 5 Identified spectra from E.coli data sets of Mascot, Sequest and MSPoisDM.

For verifying the accuracy of MSPoisDM, we should calculate the overlap between the two search engines of all. The peptides which identified at least by two search engines were defined as high confidence peptides. According to the Figure 6 and Table 3 showed MSPoisDM had more high confidence peptides than others.

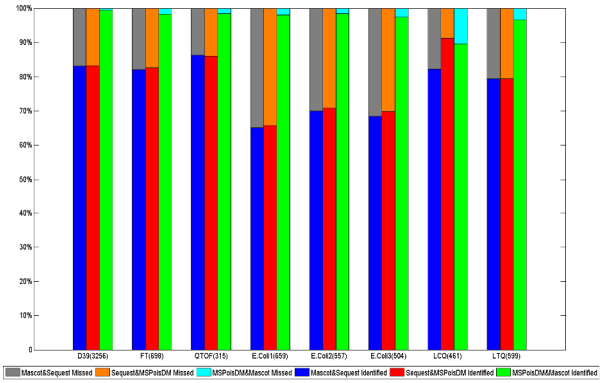


Fig. 6 High confidence peptides of Mascot, Sequest and MSPoisDM

In order to recount convenience, in Table 3 and Table 4, Mascot, Sequest, MSPoisDM, high confidence peptides and high confidence spectra abbreviated M, S, MP, H_P and H_S(Xiao, et al., 2013).

Table 3. High confidence peptides of Mascot, Sequest and MSPoisDM from varies data sets

	M & S	S & MP	M & MP	M & S & MP	H_P
D39	2707	2712	3237	2700	3256
	83.13%	83.29%	99.42%	82.92%	
FT	573	577	686	569	698
	82.09%	82.66%	98.28%	81.52%	
QTOF	272	271	310	269	315
	86.35%	86.03%	98.41%	85.40%	
*E.coli*1	429	434	646	425	659
	65.10%	65.86%	98.03%	64.49%	
*E.coli*2	390	395	548	388	557
	70.02%	70.92%	98.38%	69.66%	
*E.coli*3	345	352	491	342	504
	68.45%	69.84%	97.42%	67.86%	
LCQ	379	421	413	376	461
	82.21%	91.32%	89.59%	81.56%	
LTQ	476	476	579	466	599
	79.47%	79.47%	96.66%	77.80%	

Table 4. High confidence spectra of Mascot, Sequest and MSPoisDM from varies data sets

	M & S	S & MP	M & MP	M & S & MP	H_S
D39	81702	81896	94975	81224	96125
	85.00%	85.20%	98.80%	84.50%	
FT	18080	18296	23810	17986	24214
	74.67%	75.56%	98.33%	74.28%	
QTOF	3212	3162	3906	3093	4094
	78.46%	77.23%	95.41%	75.55%	
*E.coli*1	7499	7418	10854	7345	11081
	67.67%	66.94%	97.95%	66.28%	
*E.coli*2	7769	7688	10103	7600	10360
	74.99%	74.21%	97.52%	73.36%	
*E.coli*3	6845	6776	8865	6681	9124
	75.02%	74.27%	97.16%	73.22%	
LCQ	3274	4324	3556	3149	4856
	67.42%	89.04%	73.23%	64.85%	
LTQ	5805	5325	6626	4933	7890
	73.57%	67.49%	83.98%	62.52%	

Figure 7 showed FDR from the range of 0 ~ 5%, MSPoisDM identified peptides was the most in 0.005 ~ 0.45 of all the mentioned peptide identification algorithms, showing its reliable and significance.

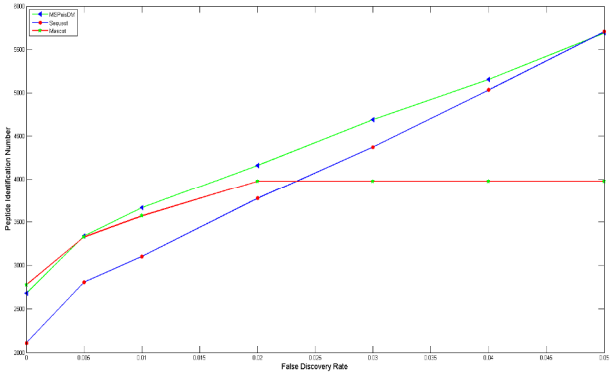


Fig. 7 FDR from the range of 0 ~5%, the identified peptides change trend of Mascot, Sequest and MSPoisDM

For verifying the generality of MSPoisDM, we utilized the data set of yeast which was generated by HCD. The searching results showed MSPoisDM identified more peptides than Mascot. Figure 8 and Figure 9 revealed the identified peptides and spectra of Mascot and MSPoisDM respectively.

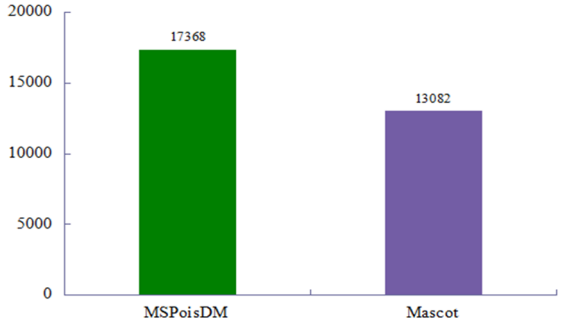


Fig. 8 Identified peptides from yeast data set of Mascot and MSPoisDM

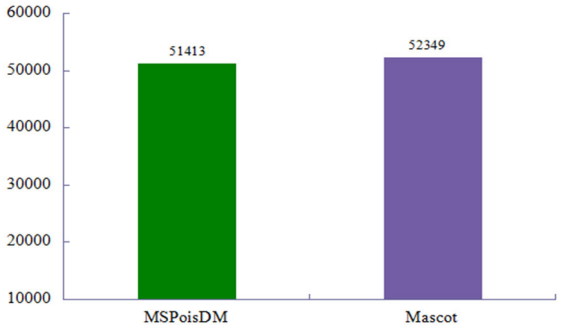


Fig. 9. Identified spectra from yeast data set of Mascot and MSPoisDM

4. Discussion

MSPoisDM proposed a novel peptide identification algorithm optimized for tandem mass spectra, and integrated the PPI characterize information, according to the diversity data sets from different MS instruments, showed MSPoisDM robust, accuracy and steady. Meantime, for verifying the generality of MSPoisDM, we adopted the yeast data set from HCD to test [5], and MSPoisDM identified more peptides than Mascot. Hence,

MSPoisDM was a universal peptide identification algorithm for tandem mass spectra.

Acknowledgements

We are grateful to Beijing scriptless comprehensive emergency drill technology research service for the help of the algorithm design and proposals.

References

1. Xiao CL, Chen XZ, Du YL, et al. Dispec: A Novel Peptide Scoring Algorithm Based on Peptide Matching Discriminability[J]. Plos One, 2013, 8: e62724-e62724.
2. Geer LY, Markey SP, Kowalak JA, et al. Open Mass Spectrometry Search Algorithm. Journal of Proteome Research 2004, 3: 958-964.
3. Bafna V, Edwards N. SCOPE: a probabilistic model for scoring tandem mass spectra against a peptide database. Bioinformatics 2001, 17: 13-21.
4. Bjornson RD, Carriero N J, Colangelo C, et al et al. X!!Tandem, an improved method for running X!tandem in parallel on collections of commodity computers. Journal of Proteome Research 2008, 7: 293-299.
5. Chi H, Sun RX, Yang B, et al. pNovo: de novo peptide sequencing and identification using HCD spectra. Journal of Proteome Research 2014, 9: 2713-2724.
6. Chick JM, Gygi SP, Nusinow DP, et al. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. Nature Biotechnology 2015, 33: 882-882.
7. Dorfer V, Pichler P, Stranzl T, et al. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra.[J]. Journal of Proteome Research, 2014, 13: 3679-3684.
8. Eriksson J, Fenyö D. Probity: a protein identification algorithm with accurate assignment of the statistical significance of the results. Journal of Proteome Research 2004, 3: 32-36.
9. Goeminne LJ, Gevaert K, Clement L. Experimental design and data-analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob. Journal of Proteomics 2018, 171: 23-36.
10. Jian L, Niu X, Xia Z, et al. A Novel Algorithm for Validating Peptide Identification from a Shotgun Proteomics Search Engine. Journal of Proteome Research 2013, 12: 1108.
11. Käll L, Storey JD, Maccoss MJ, et al. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. Journal of Proteome Research 2008, 7: 29-34.
12. Li D, Fu Y, Sun R, et al. pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. Bioinformatics 2005, 21: 3049-3050.
13. Li W, Ji L, Goya J, et al. SQID: An intensity-incorporated protein identification algorithm for tandem mass spectrometry. Journal of Proteome Research 2011, 10: 1593-1602.
14. Lin MS, Cherny JJ, Fournier CT, et al. Assessment of MS/MS SearchAlgorithms with Parent-ProteinProfiling. Journal of Proteome Research 2014, 13: 1823-1832.
15. Perkins DN, Pappin DJC, Creasy DM, et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999, 20: 3551-3567.
16. Lelong C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. Journal of Proteomics 2010, 73: 2064-2077.
17. Townsend C, Furukawa A, Schwochert J, et al. CycLS: Accurate, whole-library sequencing of cyclic peptides using tandem mass spectrometry. Bioorganic & Medicinal Chemistry 2018, 6: 1232-1238.
18. Xiao CL, Chen XZ, Du YL, et al. Binomial probability distribution model-based protein identification algorithm for tandem mass spectrometry utilizing peak intensity information. Journal of proteome research 2013, 12: 328-335.
19. Yadav AK, Kumar D, Dash D. MassWiz: a novel scoring algorithm with target-decoy based analysis pipeline for tandem mass spectrometry. Journal of proteome research 2011, 10: 2154-2160.