Preprocessing of Tandem Mass Spectrometric Data Based on Decision Tree Classification

Jing-Fen Zhang[1,2]*, Si-Min He[1], Jin-Jin Cai[1], Xing-Jun Cao[3], Rui-Xiang Sun[1], Yan Fu[1], Rong Zeng[3], and Wen Gao[1,2]

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China; [2] Graduate School of Chinese Academy of Sciences, Beijing 100080, China; [3] Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

In this study, we present a preprocessing method for quadrupole time-of-flight (Q-TOF) tandem mass spectra to increase the accuracy of database searching for peptide (protein) identification. Based on the natural isotopic information inherent in tandem mass spectra, we construct a decision tree after feature selection to classify the noise and ion peaks in tandem spectra. Furthermore, we recognize overlapping peaks to find the monoisotopic masses of ions for the following identification process. The experimental results show that this preprocessing method increases the search speed and the reliability of peptide identification.

Key words: peptide identification, Q-TOF tandem mass spectra, feature selection, decision tree

Introduction

Mass spectrometric analysis and database searching have been used as well-known approaches for peptide and protein identification (*1*). During the experiment, the peptides separated from liquid chromatographers are fragmented and ionized by collision-induced dissociation (CID) and the ions are measured by mass spectrometer for mass/charge ratios (m/z). Consequently, the peptides are identified (or sequenced) by these m/z values of ions in tandem spectra with a sequence database searching.

Due to the variety of the fragment ions under CID and the existence of a large amount of spectral noise, it is difficult to determine the sequence of a peptide from its tandem spectrum. Generally, a quadrupole time-of-flight (Q-TOF) spectrum of a peptide has 500 to 8,000 or even more peaks (*2*), but only 1%–5% of these peaks are real ones that correspond to the important and known fragment ions and are useful for peptide identification. To increase the accuracy of peptide identification and decrease the computation complexity, the preprocessing of tandem mass spectra is introduced before database searching in order to select the peaks corresponding to fragment ions and minimize the number of selected peaks.

To date, several methods have been proposed for the preprocessing of tandem data, including threshold filtering, denoise transforming, and deisotoping. The threshold filtering method is the most straightforward approach. As peaks with very small abundance values are unlikely to be real ones, this method selects the peaks above a given threshold or chooses a specific number of the most intensive peaks in the specified m/z intervals (*3*–*7*). As we know, abundance is not the fundamental attribute of real peaks. Many important b-type ions have very low abundance. In addition, for various spectra, the quality, namely the intensity baseline of noise, is totally different. Therefore, using thresholds to remove the noise is not perfect. In the denoising mechanism, some well-known procedures such as wavelet transformation have been used to denoise the raw tandem mass spectra (*6*). However, the parameters such as the wavelet base functions, the order, and the level of decomposition would impact the potential spectrum distortion by this procedure. In deisotoping, the isotopes are removed so that every fragment ion is represented only by one peak and the complexity of spectra is greatly reduced (*6*, *7*). Since peak overlappings, that is, two or more different ions have confused isotope masses, are observed frequently in spectra, deisotoping without identifying whether a peak corresponds to the monoisotope of one ion or the isotope of another ion leads to the loss of some overlapped but important fragment ions.

To address the above issues, we present a new pre-

*** Corresponding author.**
E-mail: jfzhang@jdl.ac.cn

processing method for Q-TOF tandem mass spectra based on decision tree classification. Firstly, instead of threshold filtering and denoise transforming, we use a Gaussian mixture model (GMM) to estimate the baseline of noise and treat the baseline just as one feature to distinguish noise and real peaks. Secondly, a key concept of isotope pattern vector (IPV) is introduced to characterize the isotope cluster of a fragment ion. The complex overlapping of isotope peaks are considered before deisotoping. Then we investigate the difference among noise, single fragment ions, and overlapping ions based on features such as the baseline of noise and IPV. Finally, a decision tree is constructed to classify the peaks, and the monoisotopic masses of all potential ions are calculated.

We applied our preprocessing method on four different datasets and conducted extensive experiments to evaluate the specificity and sensitivity of classification. We also evaluated the effect of the preprocessing on the speed and accuracy of the Mascot (*4*) and pFind (*8*) searches. The experimental results show that this data preprocessing approach can increase the search speed and the reliability of peptide identification.

Methods

Gaussian mixture model

Factors including the signal to noise ratio of precursor and the imperfect laboratorial environment such as temperature shifts in the laboratory may all impact the quality of spectrum. Therefore, the intensity distribution of noise is different for various spectra. For example, Figures 1 and 2 show the spectra of peptides CCAADDKEACFAVEGPK and YLGYLEQLLR, respectively. It can be observed that the intensity baseline of noise peaks in Figure 1 is much higher than that in Figure 2.

The peaks corresponding to noise are randomly produced by mass spectrometry during CID. Therefore, the variable of the intensity of noise obeys a normal distribution approximatively and a GMM can be established, in which the Gaussian curve represents the distribution of the intensity of noise. Intuitively, the centroid of the Gaussian curve corresponding to noise is treated as the baseline. Practically, the mean and standard deviations are used to characterize the baseline of noise, denoted as $I_{baseline} = (I_{mean}, I_{deviation})$, and the value of $I_{baseline}$ is obtained by the Expectation-Maximization (EM) algorithm to estimate the parameters of GMM. It is noted that we use the relative intensities instead of the absolute values of the intensities of peaks in spectra. The highest value in intensity is 100%. Using the MATLAB toolbox, the calculated results of (I_{mean}, $I_{deviation}$) for the data in Figures 1 and 2 are (2.290144%, 0.350236%) and (1.012099%, 0.076899%), respectively, which are consistent with the observation of the noise in the two spectra.

Isotope pattern vector

Isotopes are elements that contain the same number of protons and electrons but differ in the number of neutrons in nucleus. The elements of H, C, N, O, and S have stable isotope distributions in nature (*9*). Most proteins are composed of the above five elements and thereby have relatively stable isotope patterns. We use IPV to digitally describe the profile of the isotopes of an ion. Suppose that the monoisotopic mass of a fragment ion P (with molecular formula $C_{n1}H_{n2}N_{n3}O_{n4}S_{n5}$) is M, and its first four isotopes (with one, two, three, and four extra neutrons, respectively) are P_1, P_2, P_3, and P_4, we can define the IPV of P as:

$$IPV = (M, T_1, T_2, T_3, T_4, \Delta m_1, \Delta m_2, \Delta m_3, \Delta m_4)$$

where T_k is the relative abundance of P_k with respect to P, and Δm_k is the mass difference between P_k and P, for k=1∼4, respectively.

Theoretical IPV

Since the five elements of H, C, N, O, and S have stable isotope distributions, the theoretical IPV (tIPV) of a fragment ion is definite and can be deduced from its elemental components, that is, from its molecular formula. We assume that each extra neutron of an atom in the peptide appears independently. Then the tIPV for the given formula $C_{n1}H_{n2}N_{n3}O_{n4}S_{n5}$ can be deduced from the probability of the isotopes of each element. For example, we show the deduction of M, T_1, T_2, Δm_1, and Δm_2 as follows:

$$M = (12.0000, 1.0078, 14.0030, 15.9949, 31.9721) \times (n_1, n_2, n_3, n_4, n_5)^{\mathrm{T}}$$

$$T_1 = n_1 q_C + n_2 q_H + n_3 q_N + n_4 q_{O_1} + n_5 q_{S1}$$

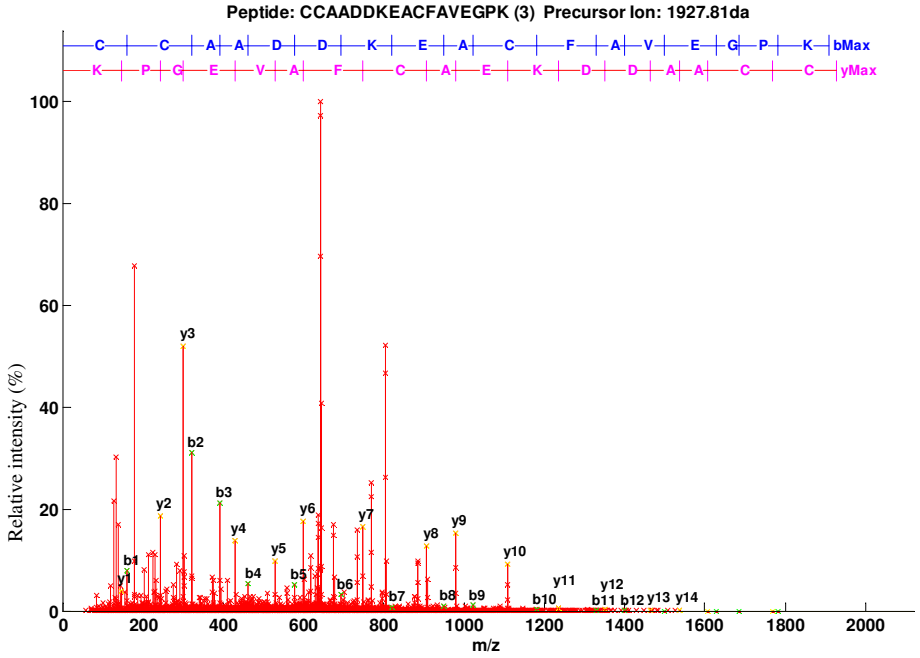

Fig. 1 The tandem mass spectrum of peptide CCAADDKEACFAVEGPK in which the precursor holds 3 charges.

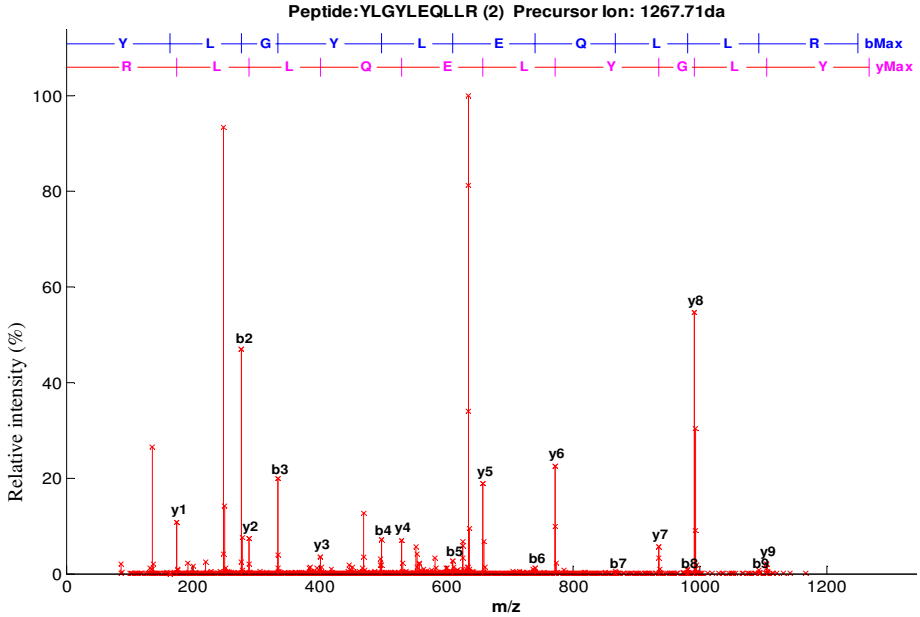

Fig. 2 The tandem mass spectrum of peptide YLGYLEQLLR in which the precursor holds 2 charges.

$$
\begin{aligned}
T_2 &= n_4q_{O_2} + n_5q_{S_2} + \frac{1}{2}T_1^2 - \frac{1}{2}(n_1q_C^2 + n_2q_H^2 + n_3q_N^2 + n_4q_{O_1}^2 + n_5q_{S_1}^2) \\
\Delta m_1 &= (n_1q_C\Delta C + n_2q_H\Delta H + n_3q_N\Delta N + n_4q_{O_1}\Delta O_1 + n_5q_{S_1}\Delta S_1)/T_1 \\
\Delta m_2 &= [n_4q_{O_2}\Delta O_2 + n_5q_{S_2}\Delta S_2 + n_1(n_1-1)q_C^2\Delta C + n_2(n_2-1)q_H^2\Delta H + n_3(n_3-1)q_N^2\Delta N \\
&\quad + n_4(n_4-1)q_{O_1}^2\Delta O_1 + n_5(n_5-1)q_{S_1}^2\Delta S_1 + n_1n_2q_Cq_H(\Delta C+\Delta H) + n_1n_3q_Cq_N(\Delta C+\Delta N) \\
&\quad + n_1n_4q_Cq_{O_1}(\Delta C+\Delta O_1) + n_1n_5q_Cq_{S_1}(\Delta C+\Delta S_1) + n_2n_3q_Hq_N(\Delta H+\Delta N) \\
&\quad + n_2n_4q_Hq_{O_1}(\Delta H+\Delta O_1) + n_2n_5q_Hq_{S_1}(\Delta H+\Delta S_1) + n_3n_4q_Nq_{O_1}(\Delta N+\Delta O_1) \\
&\quad + n_3n_5q_Nq_{S_1}(\Delta N+\Delta S_1) + n_4n_5q_{O_1}q_{S_1}(\Delta O_1+\Delta S_1)]/T_2
\end{aligned}
$$

where q_C, q_H, and q_N are the relative abundance of ^{13}C to ^{12}C, D to H, and ^{14}N to ^{15}N; ΔC, ΔH, and ΔN are the mass differences between ^{13}C and ^{12}C, D and H, and ^{14}N and ^{15}N, respectively; q_{O_1}, q_{O_2} (q_{S_1}, q_{S_2}) are the ratios of ^{17}O to ^{16}O, ^{18}O to ^{16}O (^{33}S to ^{32}S, ^{34}S to ^{32}S), respectively; ΔO_1, ΔO_2 (ΔS_1, ΔS_2) are the mass differences between ^{17}O and ^{16}O, ^{18}O and ^{16}O (^{33}S and ^{32}S, ^{34}S and ^{32}S), respectively.

Experimental IPV

We can calculate the experimental IPV (eIPV) of a fragment ion P if the isotope peaks of the ion are measured by mass spectrometer. We characterize an ion peak in mass spectrum in terms of (m/z, *intensity*), where m/z is the value of the mass to charge ratio and *intensity* is the relative height of the peak. Considering a group of isotope peaks (p_0, p_1, p_2, p_3, p_4) corresponding to an ion, the interval of the corresponding m/z values among p_0, p_1, p_2, p_3, and p_4 is around 1 Da when the ion holds a single charge, while the interval is around 0.5 Da when the ion holds double charges. In general, the interval is $1/z$ Da when the ion holds z charges. Contrariwise, the charge of an ion can be deduced by the m/z interval of the isotope peaks.

To calculate the eIPV for P, we find the corresponding isotope cluster of peaks (p_0, p_1, p_2, p_3, p_4) in tandem spectrum with the (m/z, *intensity*) pair (Mz_k, I_k), k=0∼4, and calculate the number of charge z from the interval between Mz_k. After normalizing z=1, the (m/z, *intensity*) pairs are converted to (M_k, I_k), where $M_k = Mz_k \times z - (z-1) \times 1.0078$, k=0∼4. Then the eIPV can be obtained by:

$$\begin{aligned} eIPV &= (M_0, I_1/I_0, I_2/I_0, I_3/I_0, I_4/I_0, M_1 - M_0, \\ &\quad M_2 - M_0, M_3 - M_0, M_4 - M_0) \end{aligned}$$

Feature selection and decision tree classification

The next step is to investigate the difference between noise and fragment ions based on some proposed features, and construct a decision tree to classify the peaks based on the values of the features. Firstly, since the peaks higher than the baseline of noise are more likely to be real peaks, it is necessary to find the baseline of noise $I_{baseline}$ = (I_{mean}, $I_{deviation}$) of each spectrum. Secondly, each fragment ion has theoretical isotopes while noise does not have. Therefore, noise and real peaks can be distinguished based on the concept of IPV. Considering the measure error of mass spectrometer, the isotope peaks of a fragment ion should be observed and the experimental isotope pattern should match its theoretical isotope pattern. Thirdly, there are complex overlapping ions with different charge states and noise data, hence it is very important to recognize the charge state of fragment ions and the case of overlapping to determine all the monoisotopic masses of ions. Therefore, we select some features such as the charge state, the mass corresponding to the peak, the intensity distance between the peak and the baseline of noise, and the distance between eIPV and tIPV. Finally, we investigate the difference between noise and fragment ions, learn the rules from some training samples, and construct a decision tree to classify the peaks into three classes: Class 1: noise; Class 2: real peaks corresponding to single ions; Class 3: real peaks corresponding to overlapping ions.

As described above, the interval of the m/z value of the isotope peaks is around $1/z$ Da if the ion holds z charges. For a given peak p_0, we scan the spectrum and find out the overall groups of potential isotope peaks in tandem spectrum by supposing three different charge states for z=1, 2, or 3, and within a tolerance of $0.05/z$ Da for the interval. For the above isotope cluster of peaks (p_0, p_1, p_2, p_3, p_4) with the (m/z, *intensity*) pair (Mz_k, I_k), k=0∼4, it is noted that if there is no peak at the k^{th} isotopic interval within the given tolerance, then we set the virtual peaks ($p_k, p_{k+1}, \ldots, p_4$) by setting the intensity I_j as zero, $j = k$∼4. Therefore, we can always obtain at least three groups of potential isotope peaks for p_0. Then it will be judged accordingly that which group corresponds to the fragment ion.

On the other hand, although the formula of a fragment ion is unknown during the preprocessing, the tIPV of an ion can be estimated by the expected (or mean) isotope pattern of an average peptide of the given mass (*10*). The average peptide is a peptide with an amino acid composition corresponding to the statistical distribution of amino acids in the non-redundant database and the expected $tIPV = (M_0, T_1, T_2, T_3, T_4, \Delta m_1, \Delta m_2, \Delta m_3, \Delta m_4)$ can be obtained. Therefore, we calculate the value of the features for each potential group of isotope peaks and obtain:

$V = (M_0, z, I_0 - I_{mean} - 3 \times I_{deviation}, I_0 - I_{mean} + 3 \times I_{deviation}, I_1/I_0 - T_1, I_2/I_0 - T_2, I_3/I_0 - T_3, I_4/I_0 - T_4, M_1 - M_0 - \Delta m_1, M_2 - M_0 - \Delta m_2, M_3 - M_0 - \Delta m_3, M_4 - M_0 - \Delta m_4)$

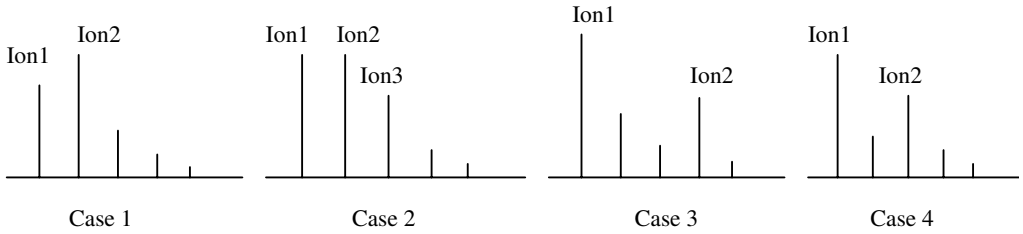

Fig. 3 Four profiles of the overlapping cases in which Ion 1, Ion 2, and Ion 3 represent the monoisotopes of each ion involved in overlapping.

We select some peaks as training samples to observe the difference between the value corresponding to noise and that to real peaks. Specifically, we judge whether a peak is noise or it corresponds to an ion or it involves overlapped ions when the peptide sequence corresponding to the spectrum is known. There are four kinds of overlappings considered as follows: Case 1: two ions with 1 Da difference in mass; Case 2: three consecutive ions with 1 Da difference in mass; Case 3: two ions with 3 Da difference in mass; Case 4: two ions with 2 Da difference in mass. The four profiles of the overlapping cases are shown in Figure 3. Then we select three classes of peaks corresponding to noise, single ions, and overlapped ions, respectively. Finally, the decision tree to classify these peaks is constructed by using the WEKA C4.5 toolbox.

According to the rules of the decision tree, all of the peaks in spectra can be classified by the calculated values of V for its potential isotope peak groups. It is noted that each peak will be classified into one and only one class. Specifically, a given peak p_0 is judged as noise if all of the values of V corresponding to the overall groups of potential isotope peaks are classified into Class 1. If it is classified into Class 2, then the monoisotopic mass $M = Mz \times z - (z-1) \times 1.0078$ is selected to present a potential fragment ion. Furthermore, if peak p_0 is classified into Class 3, then two or three monoisotopic masses will be obtained according to the overlapping cases. Finally, some masses corresponding to the peaks that have been classified into Classes 2 and 3 are selected prior to database searching.

Application

We applied our preprocessing method on four different datasets of Q-TOF mass spectra, including 54 spectra from tryptic digestion peptides (*11*), 20 spectra of Glu-Fibrino peptide B, 9 spectra of the mixture of standard peptides measured during different time, and 7 spectra of the tryptic peptides of bovine serum albumin protein (the Research Centre for Proteome Analysis, Key Laboratory of Proteomics, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences), which were denoted as PepLutefisk, PepGFB, PepMix, and PepBSA, repectively.

For performance metrics, we gave some definitions as follows. Firstly, a peak is called a real peak if its corresponding mass matches with a known theoretical ion; otherwise, it is called an invalid peak. In this paper, the known theoretical ions include the predominant *a*-, *b*-, and *y*-type of ions (*12*, *13*), immonium ions (*14*, *15*), and other less important ions such as *c*-, *x*-, and *z*-type of ions (*12*, *13*), internal fragment ions formed by a combination of *a*- and *y*-type cleavages (*14*, *15*), and ions with lost ammonia and water (*16*). It is noted that there are some peaks that really correspond to fragment ions but the corresponding masses cannot match with any known theoretical ions because the knowledge of collision rules in CID is not complete at present. Consequently, the invalid peaks include noise peaks and some peaks corresponding to fragment ions but its ion type is unknown to human beings. Secondly, it is called a true positive (TP) if a real peak is classified correctly; otherwise it is called a false negative (FN). Similarly, it is called a true negative (TN) if an invalid peak is classified correctly; otherwise it is called a false positive (FP). Finally, sensitivity and specifity are used to measure the performance of classification. Here, sensitivity is defined as TP/(TP+FN) and specifity is defined as TN/(TN+FP).

In our experiment, 900 cases were selected as training samples and 429,156 cases were selected as testing samples. The experimental results are summarized in Table 1. From the table, it can be observed that the ratios of peak selection in the four datasets are all lower than 5%. The low selecting ratios can improve the speed of database searching greatly since the less the number of selected peaks, the simpler the

Table 1 Classification Performance of the Preprocessing

Data	No. of spectra	No. of total peaks/ No. of selected peaks	Ratio of peak selection	Sensitivity	Specifity
PepLutefisk	54	89,256/3,721	4.168%	97.94%	99.06%
PepGFB	20	180,088/2,408	1.337%	97.77%	99.66%
PepMix	9	51,836/1,799	3.471%	93.68%	97.99%
PepBSA	7	18,720/789	4.215%	94.50%	97.76%

Table 2 Detailed Performance on Sensitivity of the Preprocessing

Data	No. of selected peaks	No. of real peaks in spectra*	No. of TP	No. of FN *a*-, *b*-, *y*-type/other type
PepLutefisk	3,721	2909	2,849	11/49
PepGFB	2,408	1796	1,756	1/39
PepMix	1,799	775	726	9/40
PepBSA	789	379	358	3/18

*Peaks whose corresponding masses match with the known type of theoretical ions.

computing of the subsequent identification process.

As we know, it is the real peaks that make certain the identification of peptides. The more selected real peaks, the higher the accuracy of identification. Therefore, the sensitivity of classification is very important for the identification. The detailed results on sensitivity are depicted in Table 2, where two kinds of FN samples are given in the last column: one is the peaks corresponding to the predominant *a*-, *b*-, and *y*-type of ions, and the other is the peaks corresponding to other less important types of ions. From the data, it can be observed that the former FN is much less than the later FN, which means that the lost but important information in classification is few. Compared with sensitivity, the specifity of preprocessing is less important for two reasons: Firstly, the number of invalid peaks is related to the purity of testing samples and the knowledge of collision rules in CID while the knowledge of collision rules is not sufficient and needs improvement, hence the computing of specifity is not absolutely objective; Secondly, most peaks are invalid, thus a small number of classification error has little effect on the value of specifity.

We also evaluated the effect of the preprocessing on the speed and accuracy of the Mascot (*4*) and pFind (*8*) searches. On one hand, the experimental tests were performed with pFind. The results showed that under the same parameters of searching, the accuracy of identification was increased a little while the speed of searching was improved up to 5–10 times. On the other hand, all the experiments were performed by submitting the data to Mascot through the Internet. Therefore, only the accuracy level of searching was compared since the testing of speed was not applicable. We submitted two kinds of data to Mascot: the original spectrum data and the spectrum data after our preprocessing. Comparing with the search results, we can see that: (1) If the peptide can be identified by the original data, that is, the expected peptide sequence is listed at the first position by the Mascot search, it can also be identified by the data after our preprocessing, which means that the process does not destroy the data. (2) Compared with the search scores including "Score" and "Expectation value" in Mascot search results, there were 70% data (spectra) in which the scores for the data after our preprocessing were much better than those for the original data. (3) For some spectra, such as the spectrum of peptide QNCDQFEK (in which the amino acid C is carbamidomethylated) and the spectrum of peptide DDPHACYSTVFDK, the query for the original data gave the expected sequence after the fifth position, while the query for the processed data gave the correct answer at the first position. Therefore, the search after our preprocessing is more reliable. In the future research, we will focus on improving the sensitivity and specifity of the preprocessing.

Conclusion

In this study, we present a new preprocessing method for Q-TOF tandem mass spectra to increase the accuracy of database searching for peptide (protein) identification. Instead of threshold filtering and denoise

transforming, we use a GMM to estimate the baseline of noise and treat the baseline just as one feature to distinguish noise and real peaks. In addition, based on the natural isotopic information inherent in tandem mass spectra, we construct a decision tree after feature selection to classify the noise and ion peaks and recognize overlapping peaks. The experimental results show that this preprocessing increases the search speed largely and improves the reliability of peptide identification.

Acknowledgements

This work was supported by the National Basic Research Program (973 Program) of China (No. 2002CB713807) and the National Key Technologies R&D Program of China (No. 2004BA711A21). The authors thank Dr. R. S. Johnson for kindly providing the Q-TOF data, and also thank Bin-Peng Ma and Xiao-Biao Wang from the Institute of Computing Technology for insightful discussions.

References

1. Aebersold, R. and Mann, M. 2003. Mass spectrometry-based proteomics. *Nature* 422: 198-207.
2. Cotter, R.J. 1997. *Time-of-Flight Mass Spectrometry.* American Chemical Society, Washington DC, USA.
3. Eng, J.K., *et al.* 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5: 976-989.
4. Perkins, D.N., *et al.* 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551-3567.
5. Baginsky, S., *et al.* 2002. AuDeNS: a tool for automatic *de novo* peptide sequencing. *Technical Report*, No. 383, ETH Zurich.
6. Rejtar, T., *et al.* 2004. Increased identification of peptides by enhanced data processing of high-resolution MALDI TOF/TOF mass spectra prior to database searching. *Anal. Chem.* 76: 6017-6028.
7. Gentzel, M., *et al.* 2003. Preprocessing of tandem mass spectrometric data to support automatic protein identification. *Proteomics* 3: 1597-1610.
8. Fu, Y., *et al.* 2004. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics* 20: 1948-1954.
9. Hoefs, J. 1997. *Stable Isotope Geochemistry* (fourth editon). Springer-Verlag, Berlin, Germany.
10. Zhang, J., *et al.* 2005. Predicting molecular formulas of fragment ions with isotope patterns in tandem mass spectra. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 2: 217-230.
11. Taylor, J.A. and Johnson, R.S. 2001. Implementation and uses of automated *de novo* peptide sequencing by tandem mass spectrometry. *Anal. Chem.* 73: 2594-2604.
12. Roepstorff, P. and Fohlman, J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 11: 601.
13. Johnson, R.S., *et al.* 1987. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* 59: 2621-2625.
14. Falick, A.M., *et al.* 1993. Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* 4: 882-893.
15. Papayannopoulos, I.A. 1995. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* 14: 49-73.
16. Rouse, J.C., *et al.* 1995. A comparison of the peptide fragmentation obtained from a reflector matrix-assisted laser desorption-ionization time-of-flight and a tandem four sector mass spectrometer. *J. Am. Soc. Mass Spectrom.* 6: 822-835.