MassTRAQ: a Fully Automated Tool for iTRAQ-labeled Protein Quantification

Wen-Ting Lin[1], Kuen-Pin Wu[1], Wei-Neng Hung[1], Yi-Hwa Yian[1], Yet-Ran Chen[2], Yu-Ju Chen[2], Ting-Yi Sung[1], Wen-Lian Hsu[1]
[1]Institute of Information Science and [2]Institute of Chemistry, Academia Sinica
hsu@iis.sinica.edu.tw

Abstract

Mass spectrometry has emerged to be one of the most powerful high throughput techniques for protein analysis, and iTRAQ labeling technique is the one of the newest techniques for protein identification and quantification in this field. Here, we display a new protein abundance analysis software for iTRAQ technique called MassTRAQ that can analyze up to four sets of samples at the same time. We have analyzed two samples with expression level 1:1, and have the result of 1:1.007 with standard deviation 0.004. Thus, MassTRAQ provides a good interpretation of the protein analysis.

1. Introduction

In the post-genomic era, technology development for the systematic, qualitative and quantitative protein analysis has become an important issue for global analysis of biological systems. The multidimensional chromatography based analysis enhances the detection efficiency of low-abundance proteins and is more suitable for the global analysis of complex biological systems. Recently, multiplexed set of isobaric reagents, iTRAQ, were developed for protein quantification[1]. This four-fold multiplexing reagents target at the N termini and the lysine residues in a digest mixture, and allow quantification of the relative protein expression up to four different experimental conditions.

2. The iTRAQ-based Protocol

The iTRAQ-based protocol contains four steps, as shown in Figure 1. First, we prepare proteins that are to be analyzed by the experiment. The iTRAQ labeling technique can analyze up to four sets of protein at one time. Second, each protein needs to be digested by an enzyme into a group of polypeptides. Then, each set of the polypeptides are labeled with iTRAQ reagent individually. The iTRAQ is a set of four isobaric reagents, all with a mass around 145 Daltons. Each reagent contains three groups: reporter, balance, and reactive groups. The reporter groups of the four iTRAQ reagents have molecular weights of 114, 115, 116, and 117 Daltons, respectively. The balance groups ensure that an iTRAQ-labeled peptide has the same mass, no matter labeled with which one of the four reagents. The reactive groups are attached to the N-terminal and the lysine residues of sample proteins. Each iTRAQ-labeled peptide is analyzed by tandem mass spectrometry (MS/MS) and obtains an MS/MS spectrum. The reporter groups of the iTRAQ reagents will split from the peptide and form small fragments with mass/charges (m/z) of 114, 115, 116, and 117, respectively, as shown in Figure 2. Intensity of each of these peaks represents quantity of small reporter group fragment and thus represents the quantity of a peptide sample. Other peaks in the spectrum graph are used to identify peptide sequences and therefore protein sequences.

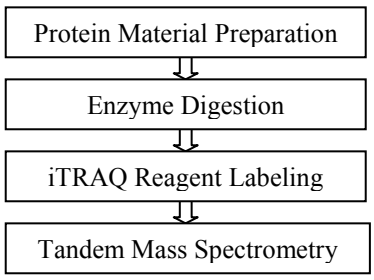

Figure 1. The iTRAQ-based protocol

3. Data Analysis and Results

The notion of the iTRAQ protocol is quite simple. Quantity of each peptide in each sample is the peak intensity of its corresponding reporter fragment in an

MS/MS spectrum. For example, as shown in Figure 2, , the intensity of peptide A (114-tagged), B (115-tagged), C (116-tagged), and D (117-tagged) is 5, 6, 3, and 5, respectively. The ratios of these pairs are 1.2(B/A), 0.6(C/A), and 1(D/A) accordingly, and these peptide ratios are theoretically the same as their corresponding protein ratios. However, tandem mass spectrometry usually generates thousands of MS/MS spectra, it is a tedious and time-consuming work for biologists to manually examine these spectra one by one. Thus the development of an automated analysis tool is essential.

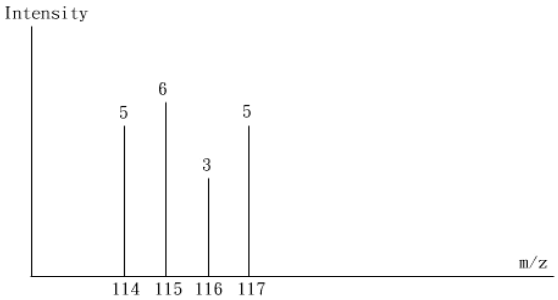

Figure 2. The four peaks of reporter groups. Peak intensity of 114, 115, 116, and 117 denote the abundance of peptide A, B, C, and D, respectively. The ratios of B/A, C/A, and D/A are 1.2, 0.6, and 1, respectively.

An automated tool should have the following two main functions:

1. Identify peptide sequences and therefore protein sequences.
2. Determine peptide ratios and therefore protein ratios according to the peak intensity of reporter groups in MS/MS spectra.

However, wet-lab results complicate the analysis problem. We inevitably face the following two problems. First, it is possible that two proteins in a sample contain the same peptide. In this case, the abundance of the mixed iTRAQ-labeled peptide is the sum of that of its parent proteins; the peptide cannot "represent" one protein. Consequently, any peptide ratio associated with such peptides cannot be used to determine any protein ratios. Second, the isotope purity and experimental configuration usually cause system errors. In other words, there may be a uniform bias to all peptide ratios and protein ratios. Thus, an automated tool should be equipped with two more functions:

3. Remove peptides that belong to more than one protein. That is, only *unique peptides* that belong to only one protein are used.
4. Perform normalization to remove system errors.

Our MassTRAQ achieves functions 1 and 3 by performing MASCOT database search. The output of the database search contains peptide sequence and unique peptide information. Ideally, peptide ratios should equal to the ratio of their parent protein if the peptides are all unique peptides. However, variations occur frequently among peptide ratios in real cases. To achieve function 2, in MassTRAQ, a protein ratio is expressed as a weighted sum of all corresponding peptide ratios, where the weight is a function of peak intensity of reporter groups. Finally, to achieve function 4, we multiply all peptide ratios by a normalizing factor to make the majority of them are 1, since we believe that in nature most protein expressions in two different cell state are unchanged.

To evaluate MassTRAQ, we synthesized two samples with expression levels 1:1, and the result of MassTRAQ is 1:1.007 with standard deviation 0.004.

4. Conclusion

Our MassTRAQ software used for iTRAq technique has four advantages: (1) MassTRAQ keeps only the unique peptides for protein quantification and thus provides more accurate results of protein quantity. (2) This software supplies adjustment of the isotope purity and prevents variation of the results. (3) MassTRAQ supports the function of normalization of the mass spectra data. It is important to distinguish proteins with significant changes from non-changed proteins. Data without normalization will produce a great number of proteins with small changes and thus results in large false-positive errors. Here, MassTRAQ equipped with normalization capability reduces the variation of experimental operation to obtain results of only few proteins with significant changes and most proteins with non-changed abundance. (4) It operates in a fully automatic fashion with good precision.

5. References

[1] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ., *Mol Cell Proteomics.*, 2004 3 1154.