

## The Multi-Q Web Server for Multiplexed Protein Quantitation

Chuan-Yih Yu, Yin-Hao Tsui, Yi-Hwa Yian,

Ting-Yi Sung\* and Wen-Lian Hsu\*

Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C.

\*To whom all correspondence should be addressed. Email: [tsung@iis.sinica.edu.tw](mailto:tsung@iis.sinica.edu.tw),  
[hsu@iis.sinica.edu.tw](mailto:hsu@iis.sinica.edu.tw)

### Abstract

Multi-Q Web Server provides an automated data analysis tool for multiplexed protein quantitation based on the iTRAQ labeling method. Multi-Q is designed as a generic platform that can accommodate various input data formats from search engines and mass spectrometer manufacturers. In comparison with its previous stand-alone version, this new web server version provides many enhanced features and flexible options for quantitation. The work flow of the web server is represented by a quantitation wizard so that the tool is easy to use. It also provides friendly user interface in configuration settings. The web server default output a standard report for quantitation results. In addition, it allows users to customize their outputs report and information of user's interest can be easily highlighted. The output provides visualization of mass spectral data so that users can conveniently validate the results. Multi-Q Web Server is a fully automated and easy to use quantitation tool suitable for large-scale multiplexed protein quantitation.

<http://ms.iis.sinica.edu.tw/Multi-Q-Web>

### 1. Introduction

The iTRAQ (1) labeling method combined with shotgun proteomic techniques represents a new dimension in multiplexed protein quantitation for relative protein expression of different cell states. The quantitation strategy for iTRAQ-labeling proteomic experiments is based on a set of four isobaric reagents, each of which is comprises of three groups: reporter, balance, and reactive groups. After cell lysis, reduction, alkylation, and protein digestion, the four states samples are separately labeled on N-terminals and lysine residues by the reactive group of iTRAQ. The labeled peptides are then combined and analyzed by liquid chromatography (LC) and tandem mass spectrometry (MS/MS).

Fully automated liquid chromatography-tandem mass spectrometry (LC-MS/MS) usually generates tens of thousands of MS/MS spectra (the size of raw data usually amounts to several gigabytes). Many thousands of peptides can be collectively analyzed by multiple LC-MS/MS runs in each proteomic experiment such that

hundreds, or even thousands, of proteins can be identified. Moreover, the high-throughput spectral data always contains much noise. Thus data analysis for proteomic quantitation becomes intractable and there is a pressing need for automated quantitation tool.

Though i-TRACKER (2) and ProQUANT (Applied Biosystems, Foster City, CA) have been available for iTRAQ-labeling quantitation, they have some limitations. The former is limited to peptide level quantitation and no protein level analysis is provided. The latter is limited to instruments developed by Applied Biosystems. Other commonly used quantitation tools such as RelEx (3) and MSQuant (4) are used for stable isotope labeling technique, a different labeling technique other than iTRAQ for protein quantitation. However, quantitation of RelEx and MSQuant is based on MS spectra and designed for two-plex quantitation, whereas iTRAQ-labeling quantitation is based on MS/MS spectra and designed for four-plex quantitation.

Therefore, we present an automatic quantitation tool for iTRAQ-labeling quantitation, called Multi-Q, which is flexible and accepts various types of protein search results and spectral data. To ensure more accurate quantitation results and to reduce manual validation efforts, in the design of Multi-Q, we paid particular attention to the following issues. First, iTRAQ's relatively large sample complexity inevitably includes identical peptides produced by two homologous proteins. These degenerate (non-unique) peptides not only result in ambiguous protein identification, but also introduce protein quantitation errors. Multi-Q quantitation is based on the nondegenerate peptides. Second, instrument's dynamic range (5) of detection will very likely affect quantitation results. Thus, Multi-Q is designed to allow users to input the threshold for dynamic range filtering.

Multi-Q is previously available as a stand-alone program executable on Windows platform (6). Currently, we develop a web server version of Multi-Q. Users can conveniently download the Multi-Q Web Server and run Multi-Q as an internal server. In this version, we further provide several enhanced functions and friendly visualization interface as described in Usage.

## **2. Method**

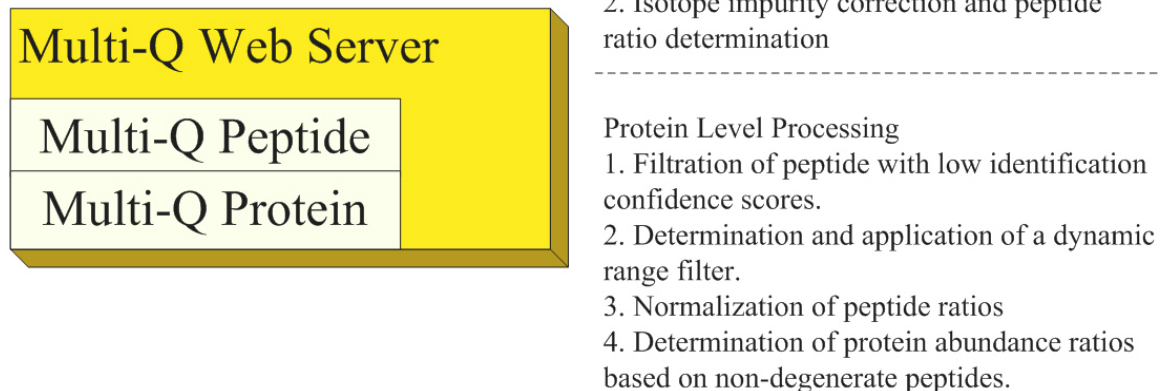


Figure 1: Multi-Q web server work flow.

The data analysis of Multi-Q follows the following procedure. On the basis of the identification results and the MS/MS spectra, signature ions of the iTRAQ-labeled peptides are selected, processed, and quantified. Then, for protein ratio determination, peptides with low identification confidence are removed. We filter out peptide data outside the instrument's dynamic range after assessment of the experimental data prior to peptide ratio normalization. Then, the normalized ratios of quantified, nondegenerate iTRAQ peptides are weighted according to their peak intensities to calculate the average protein ratios. The final protein quantitation results are exported to an html page. The above procedure is Multi-Q's default procedure of calculating protein ratios. Multi-Q allows users to modify or omit some steps of the default procedure (will be shown in Section 3.2).

In what follows, we describe each step with more details.

### Step 1. Prepare input data

The major MS manufacturers often use different data storage formats such as WIFF (Applied Biosystems), RAW (Thermal Finnigan and Waters), and BAF (Bruker Daltonics) for mass spectra acquisition. In order to accept different spectral data formats, we use the standard mzXML (7) file as the input spectral data format. For RAW and BAF spectral files, converters have been available to convert them into mzXML files. For WIFF spectral files, we provide a tool called mzFast for data conversion from WIFF to mzXML. The mzFast program generates a "reduced" mzXML file that spectral **peaks** with  $m/z$  over 120 are removed to reduce the cost of data storage and the processing time.

Multi-Q accepts search results from MASCOT (8) (CSV format) and SEQUEST (9) (XML format obtained from executing PeptideProphet (10) and ProteinProphet

(11)). Multi-Q internally generates two data structures: one is peptide list containing peptides' identification confidence scores and degeneracy information, and the other is protein list containing the probability of correct identification of all proteins. Using these two data structures, Multi-Q can generate a protein summary list with annotated information. Particularly, in MASCOT CSV files, peptides are indexed by query number, which has its corresponding spectrum information. Since quantitation is based on the scan number, we have developed a wiff2scan program to map the query number to the scan number. MASCOT users need to run wiff2scan first to generate the mapping file (.table file). Data flow diagram is shown in Figure 2.

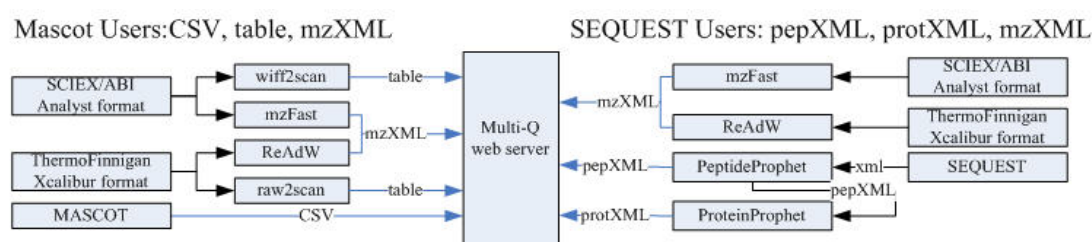


Figure 2: Data flow diagram.

## Step 2. Peptide level processing

After the data has been input, Multi-Q selects iTRAQ labeled peptides with confident MS/MS identification, detects signature ions, and performs automated quantitation of peptide abundance.

### Step 2.1 Signature ion detection and background subtraction

Peptide ratio determination is performed by identifying the signature peaks in an MS/MS spectrum with  $m/z$  of 114 to 117 and comparing their peak intensities. Multi-Q provides an option for users to determine whether or not to perform smoothing and background subtraction. To differentiate true signature peaks from noise, we smooth mass spectra first by the 3-point moving average method and select peaks using smoothed spectra. The mass tolerance of signature peak detection is determined internally in Multi-Q.

After peak selection, Multi-Q performs background subtraction. The spectrum baseline is defined as the mean of all the valleys in the smoothed curve. Note that the valleys are determined by calculating the first and second order derivatives of the curve. The peak intensity is then calculated by subtracting the baseline from the original data.

### Step 2.2 Isotope impurity correction and peptide ratio determination

The iTRAQ quantitation strategy is based on the quartet samples labeled with isotopically distinct tags; and the relative peak intensities of the resultant isotope clusters are determined to represent changes in peptide abundance. However, each batch of iTRAQ reagents contains trace levels of isotopic impurities that must be corrected to avoid variations in the true peak intensity. Since the isotopic distributions of the 114, 115, 116, and 117 signature peaks interfere with each other, over-representation of heavy signature ions will occur. Using the impurity information in the “Certificate of Analysis” provided by the iTRAQ reagent manufacturer, the interference of an isotopic reagent with its two predecessors and two successors can be corrected by solving a simultaneous equation system.

Subsequently, Multi-Q calculates the peptide ratios according to the peak intensities of the signature peaks. The ratio calculation is weighted average by default, and Multi-Q allows users to use unweighted average, i.e., each ion contributing equally.

### **Step 3. Protein level processing**

After determining the peptide ratios, Multi-Q reconstructs the proteins’ abundance ratios from the ratios of their iTRAQ labeled peptides. Multi-Q uses only confident peptides to determine protein ratios, and the threshold of peptide confidence score can be input by users. Multi-Q provides an option for users to input the dynamic range of their instrument. (Please see Section 3.2 for determining the dynamic range.) For protein quantitation, peptides ratios outside the dynamic range are removed prior to peptide ratio normalization and protein ratio determination.

#### **Step 3.1 Normalization of peptide ratios**

Multi-Q performs a normalization procedure on the ratios within the dynamic range with the Gaussian distribution (12). The normalization factor, the reciprocal of the mean of the fitted Gaussian distribution, can be used to correct the systematic bias (13).

#### **Step 3.2 Determination of protein abundance ratios based on nondegenerate Peptides**

Users can decide whether or not to use only nondegenerate peptides to calculate protein ratios. Protein ratio is calculated according to their corresponding peptide ratios. The default calculation is weighted average of peptide ratios, and the weight is determined by peptide abundance. Users are allowed to use unweighted average.

#### **Step 3.3 Output the results**

Initially, Multi-Q allows users to determine the denominators and numerators of their peptide and protein ratios. Then Multi-Q reports ratios accordingly.

Multi-Q reports some statistics of experiment results and a protein summary list which includes quantitation results and peptides contained in each protein. In addition, Multi-Q provides a new interface that provides visualization of results and information. Users can compose their own report by choosing fields to be included in protein summary and peptide summary. After specifying output composition, a new window will pop up to show quantitation results. Users can save the result as web page and use search function provided by the browser to quickly find the protein or peptide of their interest. Users can also view and save spectral information of a peptide.

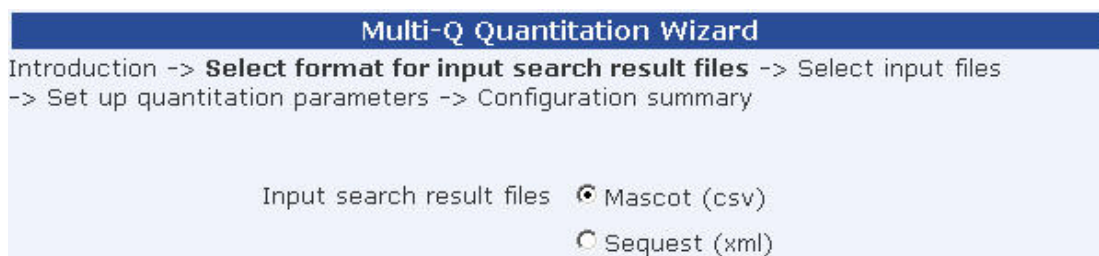
### 3. USAGE

Users can download the Multi-Q web server at

<http://ms.iis.sinica.edu.tw/Multi-Q-Web> and follow the installation guide.

#### 3.1 Input phase

Multi-Q web server sets the *Experiment\_Data* folder as the root folder. We recommend users create a folder under the root to store all input files (including search result, spectral files, and .table file) of one experiment. Figures 3 and 4 show the input phase; here, we only show input for MASCOT.

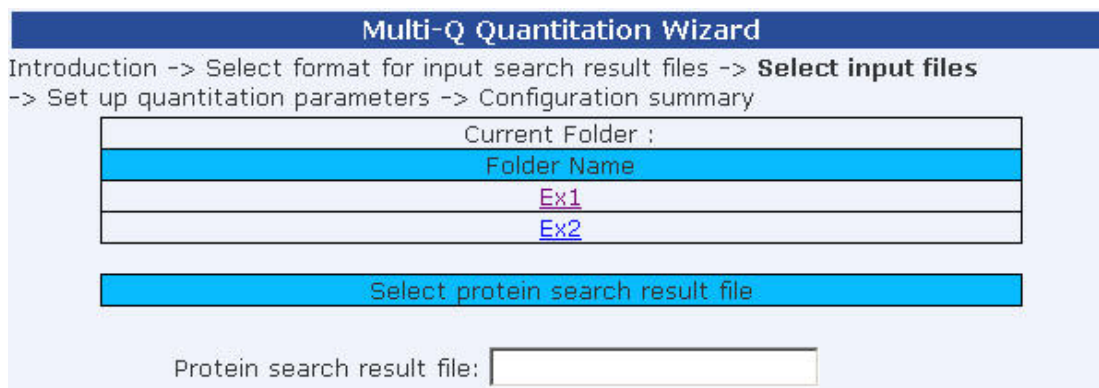


Multi-Q Quantitation Wizard

Introduction -> **Select format for input search result files** -> Select input files  
-> Set up quantitation parameters -> Configuration summary

Input search result files  Mascot (csv)  
 Sequest (xml)

Figure 3: Select the search result type.



Multi-Q Quantitation Wizard

Introduction -> Select format for input search result files -> **Select input files**  
-> Set up quantitation parameters -> Configuration summary

Current Folder :  
Folder Name  
Ex1  
Ex2

Select protein search result file

Protein search result file:

Figure 4: Input phase for MASCOT user.

### 3.2 Set up parameters for quantitation

In this stage, users can input their isotope impurity correction table and dynamic range, define their own ratio calculation, and choose peptides for protein quantitation as shown in Figure 5.

**Multi-Q Quantitation Wizard**  
Introduction -> Select format for input search result files -> Select input files  
-> **Set up quantitation parameters** -> Configuration summary

Sample/standard data type     Sample     Standard  
Sample:normalization function will be performed  
Standard:normalization function will not be performed

Reagent	% of -2	% of -1	% of +1	% of +2
iTRAQ Reagent 114	0.0	1.0	5.9	0.2
iTRAQ Reagent 115	0.0	2.0	5.6	0.1
iTRAQ Reagent 116	0.0	3.0	4.5	0.1
iTRAQ Reagent 117	0.1	4.0	3.5	0.1

**Configuration**

Perform curve smoothing and background subtraction  
 Use confident peptides score > 0.0 for protein quantitation  
 Use only nondegenerate peptides for protein quantitation  
 Weighted calculation to determine protein ratios  
 Dynamic Range

Low Intensity 0.0  
High Intensity 0.0

Ratio 1(R1)    Ratio 2(R2)    Ratio 3(R3)  
Numerator    m/z 115    m/z 116    m/z 117  
Denominator    m/z 114    m/z 114    m/z 114

Next

Figure 5: Set up quantitation parameters.

We recommend first-time users to first run Multi-Q on standard data to determine the dynamic range of their instrument. Users need to choose “Standard” data type, not the Sample data type, and leave the “Dynamic Range” box unchecked; and Multi-Q will output the dynamic range (see Figure 6) in addition to quantitation results. Then the reported (or modified) dynamic range can be input to subsequent quantitation of other experiments, and the “Sample” data type should be selected.

### 3.3 Output

Multi-Q provides copious information for proteins and peptides as shown below.

Terms in this report are explained in Multi-Q User Guide available at

<http://ms.iis.sinica.edu.tw/Multi-Q-Web>.

### Output composition

[Save CSV format report](#)

Statistics of experiment results

Ratio 1(R1)	Ratio 2(R2)	Ratio 3(R3)
m/z 115	m/z 116	m/z 117
m/z 114	m/z 114	m/z 114

	No. of proteins with ratio 0	No. of proteins with rational-value ratios	Normalization Factor
R1	36 (36/115)	79 (79/115)	-0.001
R2	30 (30/115)	85 (85/115)	1.185
R3	32 (32/115)	83 (83/115)	0.013

	No. of proteins with std not available	No. of proteins with std 0	No. of proteins with rational-value std
Std 1	34 (34/115)	36 (36/115)	45 (45/115)
Std 2	27 (27/115)	30 (30/115)	58 (58/115)
Std 3	31 (31/115)	32 (32/115)	52 (52/115)

Protein List : 115

Acc#	Score	Mass	R1	R2	R3	Std 1	Std 2	Std 3	Desc.
<a href="#">IIM1_HUMAN</a>	158	97108	0.03	1.29	0.02	N/A	0.05	0.03	(Q14974) Importin beta-1 subunit (Karyopherin beta-1 subunit) (Nuclear factor P97) (Importin 90)
<a href="#">K22O_HUMAN</a>	49	65830	0.01	1.24	0.02	0.00	0.08	0.01	(Q01546) Keratin, type II cytoskeletal 20A (Cytokeratin 20A) (CK 20A)

Figure 6: Quantitation results for standard data quantitation.

In addition to the default output report (Figure 6), Multi-Q provides flexibility for users to customize their output reports by clicking “Output composition” on top of the web page shown in Figure 6. Users can highlight fields of interest and conveniently browse peptides of very low or high ratios among the large number of fields in the output. Figure 7 shows an example of customized report.



Multi-Q Quantitation Result	
Date	2007/4/12
Time	下午 06:03
Search Name	F004675.csv
Config	
Smoothing and background subtraction	Not Applied
Data type	Sample
Peptide Score >	0
Nondegenerate Peptide	Use nondegenerate peptide
Weighted calculation	Applied
Dynamic Range	
	Low Intensity 0
	High Intensity 0

**Isotope Impurity Correction Table**

Reagent	% of -2	% of -1	% of +1	% of +2
iTRAQ Reagent 114	-	-	0.059	0.002
iTRAQ Reagent 115	-	0.02	0.056	0.001
iTRAQ Reagent 116	0	0.03	0.045	-
iTRAQ Reagent 117	0.001	0.04	-	-

**Ratio**

	Ratio 1(R1)	Ratio 2(R2)	Ratio 3(R3)
Numerator	m/z 115	m/z 116	m/z 117
Denominator	m/z 114	m/z 114	m/z 114

\*NonD\* stands for is Nondegenerate peptide

AscNo	Score	Mass	Ratio1	Ratio2	Ratio3	STD1	STD2	STD3	Description								
DMB1_HUMAN	158	97108	0.03	1.29	0.02	N/A	0.05	0.03	(Q14974) Importin beta-1 subunit (Karyopherin beta-1 subunit) (Nuclear factor P97) (Importin 90)								
spec.	score	sequence	z	m/z 114	m/z 115	m/z 116	m/z 117	Ratio 1	Ratio 2	Ratio 3	Mod	Is Label	> Score	NonD	R1	R2	R3
50	5.7	VAAGLQIK	2	326.0486	0	406.1349	8.726358	0	1.25	0.03	iTRAQ (Nterm)	Yes	Yes	Yes	No	Yes	Yes
122	27.72	VAALQNLVK	2	498.6845	15.7753	659.8676	6.290183	0.03	1.32	0.01	iTRAQ (Nterm)	Yes	Yes	Yes	Yes	Yes	Yes
550	45.4	LAATNALLNSLEFTK	2	115.9179	4.104576	168.5619	0	0.04	1.45	0		No	Yes	No	No	No	No

Figure 7: An example of customized report.

Moreover, Multi-Q provides a visualization interface to show the four MS/MS spectra (raw data, smoothed data, with background, with isotope impurity correction) by clicking the spectrum number in the peptide summary table. This feature allows users to conveniently compare the quantitation results with the MS/MS spectra. Figure 8 shows an MS/MS spectrum with isotope impurity correction.

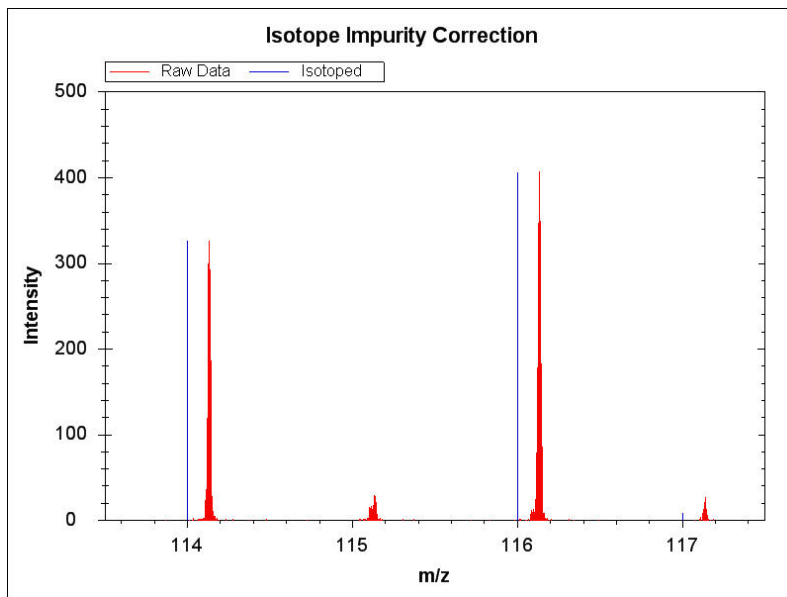


Figure 8: An example of MS/MS spectra with isotope impurity correction

#### 4. Conclusion

The Multi-Q Web Server can help users to achieve accurate, robust, high-throughput computation for multiplexed protein quantitation. In comparison with the stand-alone version of Multi-Q, this new web server provides many enhanced features and flexibility. First, it can determine the dynamic range for the first-time users of Multi-Q given standard protein mixture data is input. Then the reported dynamic range can be used for subsequent quantitation. Second, it allows users to choose their parameters for quantitation, including confidence score threshold, whether or not to use nondegenerate peptides for quantitation, and defining ratios, in order to fit users' preference. Third, the web server version further strengthens the result output so that users can easily find useful information. Multi-Q Web server provides a friendly interface for users to customize their own output report, in addition to the default report. In the customized report, users can highlight information of their interest, e.g., peptides or proteins of very low or high ratios. Since the output is shown in web page, users can use the search function of their browser to find targets of their interest. Multi-Q Web Server can show MS/MS spectra in the output for users to conveniently validate the results. In summary, Multi-Q Web Server is a fully automated and easy to use quantitation tool suitable for large-scale multiplexed protein quantitation.

#### Acknowledgements

We thank Dr. Yu-Ju Chen and Dr. Y. R. Chen for useful discussion and assistance in developing Multi-Q Web Server.

#### References

1. Ross, P. L., Huang, Y. N., Marchese, J. N., 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. and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics*, 3, 1154-1169.
2. Shadforth, I. P., Dunkley, T. P., Lilley, K. S. and Bessant, C. (2005) i-Tracker: for quantitative proteomics using iTRAQ. *BMC Genomics*, 6, 145.
3. MacCoss, M. J., Wu, C. C., Liu, H., Sadygov, R. and Yates, J. R., III. (2003) A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal. Chem.*, 75, 6912-6921.
4. Schulze, W. X. and Mann, M. (2004) A novel proteomic screen for

- peptideprotein interactions. *J. Biol. Chem.*, 279, 10756-10764.
5. Skoog, D. A. and Leary, J. J. (1992) Principle of Instrumental Analysis, 4th ed. *Saunders College Publishing: New York*.
  6. Lin, W. T., Hung, W. N., Yian, Y. H., Wu, K. P., Han, C. L., Chen, Y. R., Chen, Y. J., Sung, T. Y. and Hsu, W. L. (2006) Multi-Q: A Fully Automated Tool for Multiplexed Protein Quantitation. *J. Proteome Res.*, 5, 2328-2338.
  7. Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W. and Aebersold, R. (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.*, 22, 1459-1466.
  8. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20, 3551-3567.
  9. Eng, J. K., McCormack, A. L. and Yates, J. R., III. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in protein database. *J. Am. Soc. Mass Spectrom.*, 5, 976-989.
  10. Keller, A., Nesvizhskii, A. I., Kolker, E. and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.*, 74, 5383-5392.
  11. Nesvizhskii, A. I., Keller, A., Kolker, E. and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.*, 75, 4646-4658.
  12. Golub, G. H. and Van Loan, C. F. (1996) Matrix Computations, 3rd ed. *The Johns Hopkins University Press: Baltimore, MD*.
  13. Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W., Webb-Robertson, B. M., Smith, R. D. and Lipton, M. S. (2006) Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.*, 5, 277-286.