Nucleic Acids Research, 2007, Vol. 35, Web Server issue ***W707–W712***
doi:10.1093/nar/gkm345

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.

Received January 31, 2007; Revised April 20, 2007; Accepted April 22, 2007

ABSTRACT

The Multi-Q web server provides an automated data analysis tool for multiplexed protein quantitation based on the iTRAQ labeling method. The web server is designed as a platform that can accommodate various input data formats from search engines and mass spectrometer manufacturers. Compared to the previous stand-alone version, the new web server version provides many enhanced features and flexible options for quantitation. The workflow of the web server is represented by a quantitation wizard so that the tool is easy to use. It also provides a friendly interface that helps users configure their parameter settings before running the program. The web server generates a standard report for quantitation results. In addition, it allows users to customize their output reports and information of interest can be easily highlighted. The output also provides visualization of mass spectral data so that users can conveniently validate the results. The Multi-Q web server is a fully automated and easy to use quantitation tool that is suitable for large-scale multiplexed protein quantitation. Users can download the Multi-Q Web Server from http://ms.iis.sinica.edu.tw/Multi-Q-Web.

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 comprised of three groups: reporter, balance and reactive groups. After cell lysis, reduction, alkylation and protein digestion, the four-state 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; hence, the amount of raw data is usually in the order of 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. However, the high-throughput spectral data always contains a great deal of noise, which makes data analysis for proteomic quantitation intractable. Thus, there is a pressing need for an automated quantitation tool.

Although i-TRACKER (2) and ProQUANT (Applied Biosystems, Foster City, CA) have been used for iTRAQ-labeling quantitation, they have some limitations. The former is limited to peptide level quantitation and no protein level analysis is provided, while 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 the stable isotope labeling technique, a different labeling technique other than iTRAQ for protein quantitation. In RelEx and MSQuant, quantitation 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 have developed 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 the amount of manual validation, we paid particular attention to the following issues when we designed Multi-Q. 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. In contrast, Multi-Q quantitation is based on the non-degenerate peptides. Second, an instrument's dynamic range (5) of detection will very likely affect the accuracy of the quantitation results. Thus,

*To whom correspondence should be addressed. Tel: 886-2-2788-3799 ext 1711; Fax: 886-2-2782-4814; Email: tsung@iis.sinica.edu.tw,
Correspondence may also be addressed to Wen-Lian Hsu. Tel: 886-2-2788-3799 #1804; Fax: 886-2-2782-4814; Email: hsu@iis.sinica.edu.tw

© 2007 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

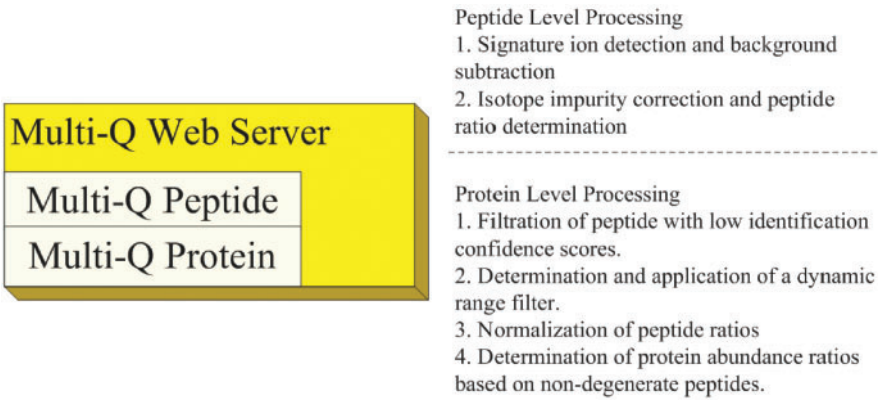

Figure 1. The Multi-Q web server workflow.

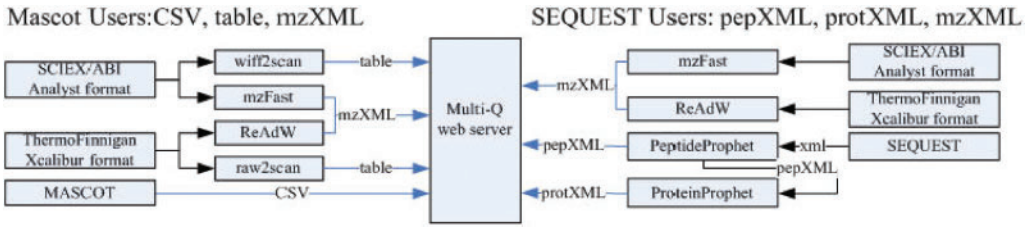

Figure 2. The Multi-Q web server data flow diagram.

Multi-Q is designed to allow users to input a threshold for dynamic range filtering.

Multi-Q is already available as a stand-alone program executable on the Windows platform (6). We have now developed a web server version of Multi-Q. Users can conveniently download the web server and run the program as an internal server. In this version, we provide several enhanced functions as well as a user-friendly visualization interface, as described in the USAGE section.

METHOD

The data analysis procedure of Multi-Q is as follows (Figure 1). Based on 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 assess the experimental data and filter out peptides outside the instrument's dynamic range prior to peptide ratio normalization. Then, the normalized ratios of quantified, non-degenerate iTRAQ peptides are weighted according to their peak intensities to calculate the average protein ratios, and the final protein quantitation results are exported to an HTML page. The above procedure is Multi-Q's default procedure for calculating protein ratios. Users are allowed to modify or omit some steps of this procedure (see Set up parameters for quantitation in USAGE).

We now describe each step in detail.

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 are 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 by removing spectral peaks with 'm/z' over 120, which reduces the cost of data storage and the processing time.

The web server accepts search results from MASCOT (8) (CSV format) and SEQUEST (9) [XML format obtained by executing PeptideProphet (10) and ProteinProphet (11)]. It internally generates two data structures: 1) a peptide list containing the peptides' identification confidence scores and degeneracy information, and 2) a protein list containing the correct identification probabilities of all proteins. Using these two data structures, Multi-Q generates a protein summary list with annotated information. Specifically, in MASCOT CSV files, peptides are indexed by the query number, which has its corresponding spectral 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). The data flow diagram of Multi-Q is shown in Figure 2.

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.

Signature ion detection and background subtraction. Peptide ratios are determined by identifying the signature peaks in an MS/MS spectrum with 'm/z' of 114 to 117 and comparing their peak intensities. To differentiate true signature peaks from noise, we smooth the mass spectra first with the 3-point moving average method and select peaks using the smoothed spectra. The mass tolerance of signature peak detection is determined internally by the Multi-Q web server.

After peak selection, the Multi-Q web server 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. However, in our experience, not performing smoothing and background subtraction only affects the quantitation results very slightly, but performing the tasks requires a great deal of time. Therefore, the Multi-Q web server provides an option so users can decide whether or not to perform smoothing and background subtraction.

Isotope impurity correction and peptide ratio determination. The iTRAQ quantitation strategy is based on the quartet samples labeled with isotopically distinct tags. The relative peak intensities of the resultant isotope clusters are identified to represent changes in peptide abundance. However, each batch of the 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, the Multi-Q web server calculates the peptide ratios according to the peak intensities of the signature peaks. The ratio calculation uses a weighted average by default, and the Multi-Q web server allows users to employ an unweighted average, i.e. each ion contributes equally.

Step 3. Protein level processing

After determining the peptide ratios, the Multi-Q web server reconstructs the proteins' abundance ratios from the ratios of their iTRAQ labeled peptides. Only confident peptides are used to determine protein ratios, and the threshold for peptide confidence scores can be input by users. The web server also provides an option for users to input the dynamic range of their instruments see Set up parameters for quantitation in USAGE. For protein quantitation, peptides ratios outside the dynamic range are removed prior to peptide ratio normalization and protein ratio determination.

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

Determination of protein abundance ratios peptides. The Multi-Q web server uses non-degenerate peptides to calculate protein rations by default. However, it allows users to decide whether or not to use non-degenerate peptides only. Protein ratios are calculated according to their corresponding peptide ratios. The default calculation uses a weighted average of peptide ratios, and the weight is determined by the peptides' abundance. Users are allowed to use an unweighted average.

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

The Multi-Q web server reports some statistics of the experiment results and a protein summary list, which includes the quantitation results and peptides in each protein. It also has a new interface that provides visualization of results and information. Users can compose their own reports by choosing the fields to be included in the protein summary and peptide summary. After the output's composition has been specified, a pop up window shows the quantitation results. Users can save the results as a web page and use search function provided by the browser to quickly find the proteins or peptides of interest. They can also view and save spectral information about a peptide.

USAGE

Users can download the Multi-Q web server from http://ms.iis.sinica.edu.tw/Multi-Q-Web and follow the installation guide to install the program.

Input phase

The Multi-Q web server sets the 'Experiment_Data' folder as the root folder. We recommend that users create a folder under the root to store all input files (including search results, spectral files, and table files) of an experiment. Figures 3 and 4 show the input phase. We only show the input for MASCOT.

Set up parameters for quantitation

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

We recommend that new users first run the Multi-Q web server on standard data to determine the dynamic range of their instruments. To do this, the 'Standard' data type, not

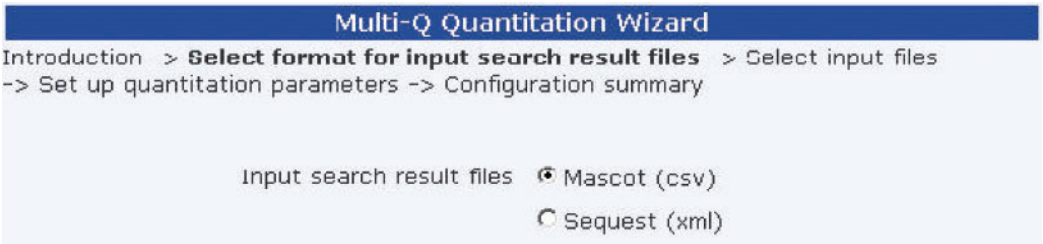

Figure 3. Select the search result type.

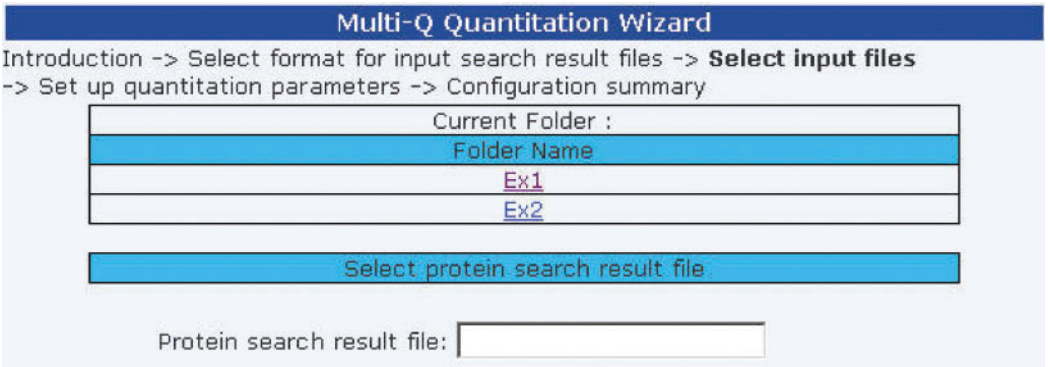

Figure 4. Input phase for MASCOT users.

the 'Sample' data type, should be selected; and the 'Dynamic Range' box should be left unchecked. The web server will output the dynamic range (Figure 6) in addition to the quantitation results. Then, the reported (or modified) dynamic range can be input for subsequent quantitation in other experiments. In this step, the 'Sample' data type should be selected.

Output

The Multi-Q web server provides copious amounts of information for proteins and peptides, as shown below. Terms used in this report are explained in the User Guide available at: http://ms.iis.sinica.edu.tw/Multi-Q-Web.

In addition to the default output report (Figure 6), the web server allows users to customize their output reports by clicking 'Output composition' at the top of the web page shown in Figure 6. Users can highlight fields of interest and conveniently browse peptides with very low or high ratios among the large number of fields in the output. Figure 7 shows an example of a customized report.

Finally, the Multi-Q web server provides a visualization interface that shows the four MS/MS spectra (raw data, smoothed data, spectrum with background, and spectrum with isotope impurity correction). By clicking the spectrum number in the peptide summary table, users can easily compare their quantitation results with the MS/MS spectra. Figure 8 shows an MS/MS spectrum with isotope impurity correction.

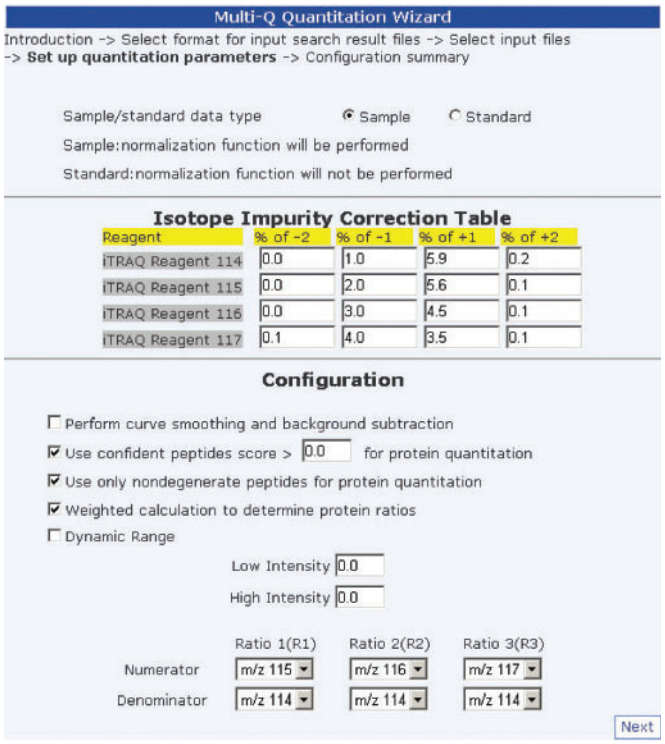

Figure 5. Set up quantitation parameters.

CONCLUSION

The Multi-Q web server can help users achieve accurate, robust, high-throughput computation for multiplexed protein quantitation. Compared to the stand-alone version of Multi-Q, this new web server provides many enhanced features as well as more flexibility. First, given the standard protein mixture data as input, it determines

Output composition

Save CSV format report

Statistics of experiment results

Gen XML

Gen CSV

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	0 (0/115)	79 (79/115)	0.000
R2	0 (0/115)	85 (85/115)	0.000
R3	0 (0/115)	83 (83/115)	0.000
	No. of proteins with std not available	No. of proteins with std 0	No. of proteins with rational-value std
Std 1	70 (70/115)	0 (0/115)	45 (45/115)
Std 2	57 (57/115)	0 (0/115)	58 (58/115)
Std 3	63 (63/115)	0 (0/115)	52 (52/115)

Evaluated Dynamic Range	
High Intensity (Avg.)	277.6717
Low Intensity (Avg.)	125.2666

Protein List : 115

ReCalc

Acc#	Score	Mass	R1	R2	R3	Std 1	Std 2	Std 3	Desc.
IMB1_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)
K22O_HUMAN	49	65830	0.01	1.24	0.02	0.00	0.08	0.01	(Q01546) Keratin, type II cytoskeletal 2 oral (Cytokeratin 2P) (K2P) (CK 2P)

Figure 6. Quantitation results for a standard data experiment.

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

Acc#/Mr	Score	Mass	Ratio1	Ratio2	Ratio3	STD1	STD2	STD3	Description
IMB1_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 a customized report.

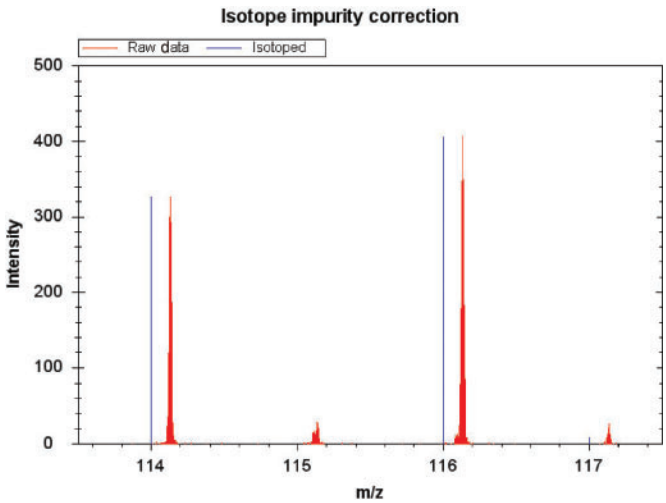

Figure 8. An example of an MS/MS spectrum with isotope impurity correction.

the dynamic range of first-time users' instruments. Then the reported dynamic range can be used for subsequent quantitation. Second, to satisfy user preferences, it allows users to choose their parameters for quantitation, such as setting the confidence score threshold, deciding whether or not to use non-degenerate peptides for quantitation, and defining ratios. Third, the web server version further strengthens the result output so that users can easily find useful information. The 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 interest, e.g. peptides or proteins with very low or very high ratios. Since the output is shown on a web page, users can use the search function of their browsers to find items of interest. The Multi-Q web server can show MS/MS spectra in the output so that users can easily validate their results. In summary, the Multi-Q web server is a fully automated and easy to use quantitation tool that is suitable for large-scale multiplexed protein quantitation.

ACKNOWLEDGEMENTS

We thank Dr Yu-Ju Chen and Dr Y. R. Chen for useful discussions and assistance in developing the Multi-Q web server. This work is partially supported by The Thematic Research Program, Academia Sinica, under grant 94B003-1. Funding to pay the Open Access publication charges for this article was provided by Academia Sinica.

Conflict of interest statement. None declared.

REFERENCES

1. Ross,P.L., Huang,Y.N., Marchese,J.N., Williamson,B., Parker,K., Hattan,S., Khainovski,N., Pillai,S., Dey,S. *et al.* (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 peptide-protein interactions. *J. Biol. Chem.*, **279**, 10756–10764.
5. Skoog,D.A. and Leary,J.J. (1992) *Principle of Instrumental Analysis, 4th edn.* 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. *et al.* (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 edn.* 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.