

An Automated High Throughput Proteolysis and Desalting Platform for Quantitative Proteomic Analysis

Albert-Baskar Arul¹, Na-Young Han¹, and Hookeun Lee^{1,2,3*}

¹Lee Gil Ya Cancer and Diabetes Institute, Gachon University, Incheon, Korea

²College of Pharmacy, Gachon University, Incheon 406-840, Korea

³Gachon Medical Research Institute, Gil Medical Center, Incheon 405-760, Korea

Received May 29, 2013; Revised June 12, 2013; Accepted June 17, 2013

First published on the web June 30, 2013; DOI: 10.5478/MSL.2013.4.2.25

Abstract: Proteomics for biomarker validation needs high throughput instrumentation to analyze huge set of clinical samples for quantitative and reproducible analysis at a minimum time without manual experimental errors. Sample preparation, a vital step in proteomics plays a major role in identification and quantification of proteins from biological samples. Tryptic digestion a major check point in sample preparation for mass spectrometry based proteomics needs to be more accurate with rapid processing time. The present study focuses on establishing a high throughput automated online system for proteolytic digestion and desalting of proteins from biological samples quantitatively and qualitatively in a reproducible manner. The present study compares online protein digestion and desalting of BSA with conventional off-line (in-solution) method and validated for real time sample for reproducibility. Proteins were identified using SEQUEST data base search engine and the data were quantified using IDEALQ software. The present study shows that the online system capable of handling high throughput samples in 96 well formats carries out protein digestion and peptide desalting efficiently in a reproducible and quantitative manner. Label free quantification showed clear increase of peptide quantities with increase in concentration with much linearity compared to off line method. Hence we would like to suggest that inclusion of this online system in proteomic pipeline will be effective in quantification of proteins in comparative proteomics where the quantification is really very crucial.

Keywords: Proteomics, Peptide linearity, IDEALQ, Online digestion, Comparative proteomics, Peptide quantification

Introduction

Mass spectrometry (MS)-based quantitative proteomics has been applied to solve a wide variety of biological problems, and several MS-based workflows have been developed for protein and peptide quantitation¹. In mass spectrometric quantitation methods it is usually assumed that the measured signal has a linear dependence on the amount of material in the sample for the entire range of amounts being studied. A prerequisite for accurate quantitation is that unwanted experimental variations in sample extraction, preparation, and analysis be minimized, and it is therefore critical that each step in the workflow is optimized for reproducibility.² Presently proteomics utilizes highly sensitive instruments for analysis of the proteome. The proteins are identified by either matching the masses of several of the observed peptides to those predicted for a protein in a database (peptide mapping), or by obtaining sequence information from a single peptide by tandem mass spectrometry (MS/MS).^{3,4}

Sample preparation is a vital step in proteomics which can greatly affect the proteome data at the end of an

experiment after investing days of work with most sophisticated instruments available nowadays. Developments over the recent years have provided us with more sophisticated and sensitive instruments to quantify highly complex biological samples. Label free proteomics usually involves three steps including sample preparation (protein extraction and digestion), peptide separation and data analysis.⁵ Label free quantification has a wide range of advantages over other methods used in proteomics. In spite of these sensitive instruments and good software, quantification may cause unreliable results if the sample preparation step is not efficient and reproducible.

In the present study we intend to introduce a new platform for the efficient and reproducible analysis of the complex biological samples using label free quantification. The proteolysis cartridges embedded with trypsin cleaves the protein efficiently in reproducible manner. We wish to introduce the online system in the present proteomics pipeline for identification of proteins in high throughput complex biological sample.

Experimental

Chemicals and reagents

Bovine serum albumin (BSA) was purchased from Sigma,

*Reprint requests to Prof. Hookeun Lee
E-mail: hkleee@gachon.ac.kr

Human lung cancer cell line (A549) was purchased from American Type Culture Collection (ATCC). Tris (2-carboxy-ethyl) phosphine (TCEP) was obtained from Thermo Fisher Scientific (Rochford, IL, USA). Iodoacetamide (IAA) was acquired from SigmaAldrich (St. Louis, MO, USA). Spin column was purchased from Havard Apparatus (Massachusetts, Boston, USA). Sequencing grade modified trypsin was obtained from Promega (Madison, WI, USA). Acetic acid, acetonitrile and high-performance liquid chromatography (HPLC) grade water from J.T. Baker (Phillipsburg, NJ, USA) were used for the study. Protease inhibitor cocktail and phosphatase inhibitor cocktail tablets were obtained from Roche (Pleasanton, CA, USA).

Sample preparation

Two different experiments were carried out using eight different concentrations of albumin using online and off-line (in-solution) digestion protocol. BSA was dissolved in milliQ water and the protein was denatured by adding 8 M Urea in 0.1 M Tris-HCl at pH 8.5. A549 cells were used for real time validation of the method.

Protein digestion and desalting

Protein was reduced with 5 mM TCEP for 30 min at 37 °C and was alkylated by blocking the cysteine residues using 15 mM IAA at 25 °C for 1 h in dark. The pH was adjusted to 8 using 1 M Tris and the Urea concentration was reduced to below 2 M using 10 mM Tris. The reduced and alkylated proteins were digested using sequence graded trypsin at 1:50 ratio for protein: trypsin at 37 °C overnight. The activity of trypsin was stopped by adding formic acid (FA) and the pH was reduced to 2-3 before desalting. The digested peptides were desalted using sep-pak columns and the peptides were eluted with 80% acetonitrile (ACN) in 0.1% FA in water. For online digestion 100 µl of the reduced and alkylated protein solution was loaded to a proteolysis cartridge automated to handle trypsinization at a flow rate of 2 µl/min. The desalting step was carried out using reversed-phase (RP) cartridge and the tryptic peptides were eluted with 80% ACN in 0.1% FA in water. Proteolysis cartridge is immobilized with trypsin using 5 µL resin of size 45-53 µm on synthetic polymer coupled with trypsin by epoxide to digest a minimum of 4 µg BSA per minute at a flow rate of 2 µl/min. On the other hand RP clean up cartridges are packed in 5 µl of underivatized polystyrene-divinylbenzene beads of pore size 100 Å and particle size range of 15-20 µm capable of binding 450 µg human insulin at a flow rate of 2 µl/min. Purified peptides were collected in 96 Greiner 650201, U-bottom plates and were used for RP clean-up. Tryptic digestion and RP clean-up were automated using Assay Map bravo platform from Agilent technologies with VWorks Automation Control 11.1 software.

LC-MS/MS analysis

The digested peptides were dried using speedvac and suspended in 0.1% trifluoroacetic acid (TFA) for LC-MS/MS analysis. Peptides were separated using a nano pump connected to high pressure liquid chromatography system in the presence of HPLC chip (large capacity chip, 150 mm, 300 Å, C18 chip, w/160 nl trap column). The peptides were loaded to the trapping column at a flow rate of 0.3 µl/min with HPLC grade water in 0.1% FA as mobile phase A and 90% ACN, 0.1% FA in HPLC grade water as Mobile phase B. The peptides were separated using 45 min gradient by mobile phase solvent A (0.1% FA in water) and mobile phase B solvent (90% ACN and 0.1% FA in water). Peptide separation was carried out using the gradient elution: 0-45, 3-45% B; 45.10-55, 45-90% B; and return to original conditions. The samples were run in triplicates with blank in between samples to clean the column. Mass spectrometry analysis was performed using 6520 Accurate-Mass Quadrupole Time-of-Flight (Q-TOF) LC-MS (Agilent Technologies, Wilmington, DE, USA) with HPLC-chip cube source.

Database search

The raw files generated by the Q-TOF were processed and analyzed using Sorcerer using the Uniprot human protein fasta database (2012, July version), including commonly observed contaminants.⁶

Label-free quantification

The IDEAL-Q software was utilized to process the LC-MS/MS data. The extracted ion chromatogram (XIC) was used to determine the peptide abundance in an LC-MS run.⁷

Results and Discussion

Linearity of peptide quantity from proteolysis

Successful analysis of proteome from biological samples needs some of the major factors like linearity, reproducible, efficient and specificity in digestion of the complex protein mixture. Since a small modification in peptide cleavage can naturally alter the peptide picture resulting in increased peptides complicating the arrangement of identified peptide during label free quantification.⁸ Hence refining the sample preparation to obtain reproducibly quantifiable peptides will be a great breakthrough in comparative proteomics. Our preliminary studies with different concentrations of BSA using proteolysis and RP cartridges showed reproducible results (data not shown). Hence we were interested in comparing the elution profile and the peptide picture of BSA using off-line method. The online system from Agilent technologies contains 9 desks combined with an automated robot to pick up cartridges, aspirate and dispense liquids at controlled conditions. The major advantage of online method over off line is the time consumption for

tryptic digestion; online method can be completed within 100 minutes (depending on the sample volume and flow rate) whereas the conventional offline method needs 16 h. Reducing the exposure time of trypsin for digestion can dramatically decrease the products of tryptic autolysis.⁹ Proteolytic digestion and RP clean up were carried out separately using automated protocol which consisted of priming and equilibrating cartridges, loading samples, washing unbound material, eluting the digested peptide with 80%ACN in 0.1% FA in water, followed by regeneration of the column and re-equilibrating using proteolysis and RP cartridges respectively. Tryptic digestion was carried out in a stepwise manner using the online system with cartridges embedded with trypsin (1) priming using 50 mM ammonium bicarbonate buffer (ABC) [100 µL at flow rate of 300 µL/min], (2) equilibration using 50 mM ABC [50 µL at flow rate of 10 µL/min], (3) digestion of sample using tryptic cartridge [200 µL at flow rate of 2 µL/min], (4) regeneration using 1 mM HCl [50 µL at flow rate of 25 µL/min]. On the other hand RP clean-up utilizes the following steps for desalting the tryptic peptides (1) priming using 80% ACN in 0.1% FA in water [100 µL at flow rate of 300 µL/min], (2) equilibration using 0.1% FA [50 µL at flow rate of 25 µL/min], (3) load sample for desalting using RP cartridge [200 µL at flow rate of 2 µL/min], (4) cartridge cup washing using 0.1% FA [50 µL], (5) internal cartridge wash using 0.1% FA [50 µL at flow rate of 25 µL/min], (6) stringent syringe was using 80%ACN in 0.1% FA in water [50 µL], (7) elution of desalted peptides using 80%ACN in 0.1% FA in water [50 µL at flow rate of 5 µL/min], (8) re-equilibrium using 0.1% FA [50 µL at flow rate of 25 µL/min].

The presence of tryptic cleavage sites (Lys or Arg) at the C-terminus is often used to validate efficiency of trypsin-ization; conversely, a substantial amount of peptides with missed cleavage sites are commonly observed in samples. The completeness of trypsinization can be detected by the presence of either none or

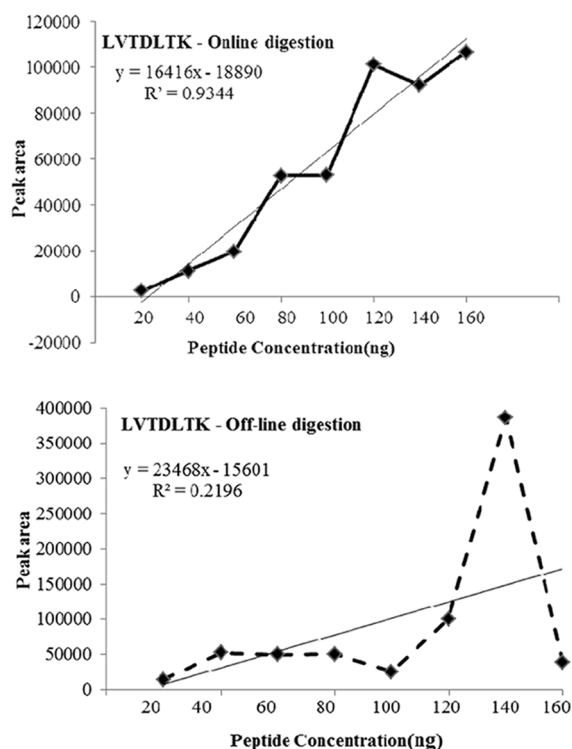
one lys residue in the observed peptides. Presence of two or more Lys residues within a peptide reflects incomplete digestion. Thus, the multiple patterns of the peptides produced provide a means to assess trypsin efficiency using any digestion conditions. The amino acid sequence coverage is also used as a measure for both the completeness of the protein digestion and the detection efficiency of the various tryptic peptides, and is common way in proteomics to define the digestion rate.¹⁰ To determine the linearity of the identified peptides and reproducibility eight different concentrations of BSA were used and the eluted peptides from online and off line methods were analyzed for MS/MS with increasing concentration to determine the peptide abundance versus sample ratio. Equal concentrations of peptides (20, 40, 60, 80, 100, 120, 140, and 160 µg) were loaded for LC-MS/MS analysis to compare the linearity and reproducibility of eluted peptides for online and off line methods. Table 1 shows the identified, tryptic, partially tryptic and non tryptic peptides with missed cleavage and the protein sequence coverage for BSA processed using online and offline method. The results of the present study shows that the offline method shows increased percentage of peptides with zero missed cleavage (fully tryptic) whereas online method shows decreased percentage of the same. Also the protein coverage was found to be very good in the offline method compared to online. Hence we wished to determine the ability of online system for linearity and reproducibility of the observed peptides. To determine the linearity of the methods regression coefficient (R^2) of identified peptides were taken as a measure. Figure 1 shows the R^2 of similar peptides identified from online and off-line digestion. Proteolytic digestion and desalting using online method showed highest linearity (high R^2 values) compared to the off-line method which showed lowest linearity.

Label free quantification uses two different criteria for measuring comparative expression of proteins in biological sample for analyzing biomarker candidate in disease and treatment. Hence given with more sophisticated analytical instrument the quality of the data remains on the perfection of

Table 1. Total number of identified peptides and proteins in bovine serum albumin (BSA) digested using Online and off line digestion.

Peptide Quantity injected (ng)	Online Digestion								Off line digestion							
	TP	FTP	PTP	NT	Missed Cleavage			Protein Coverage (%)	TP	FTP	PTP	NT	Missed Cleavage			Protein Coverage (%)
					0 (%)	1	2						0	1	2	
20	27	8	17	2	18 (67)	8	1	20.5	60	43	17	0	46 (77)	11	3	63.3
40	43	33	10	0	32 (74)	8	3	65.7	58	48	9	1	52 (90)	6	0	73.4
60	35	26	9	0	25 (71)	8	2	67.1	75	55	20	0	65 (87)	9	1	76.1
80	33	23	10	0	28 (85)	5	0	56.5	65	45	20	0	51 (78)	12	2	72.0
100	29	22	7	0	23 (79)	5	1	42.8	55	47	8	0	45 (82)	9	1	68.9
120	34	24	10	0	31 (91)	3	0	55.2	67	50	17	0	59 (88)	7	1	72.7
140	24	18	4	2	21 (88)	2	1	42.2	91	64	26	1	82 (90)	7	2	72.5
160	21	13	6	2	17 (81)	4	0	35.1	58	41	16	1	52 (90)	5	1	67.1

TP- Total peptides; FTP- Fully tryptic peptides; PTP- Partially tryptic peptides; NT- Non-tryptic peptides; () percentage of peptide compared to the corresponding total peptides identified



Number	PEPTIDE SEQUENCE	R ² of Online digestion	R ² of offline digestion
1	LVTDLTK	0.9344	0.2196
2	HPEYAVSVLLR	0.8937	0.5771
3	HADICILPDIEK	0.8731	0.7593
4	MPCTFDYI.SLILNR	0.8627	0.2538
5	SQYLQCCPFDEHVK	0.858	0.064
6	HLVDEPQNLIK	0.8152	0.2089
7	HPYFYAPELLYYANK	0.7892	0.3743
8	KQTALVELLK	0.7773	0.2756
9	SLHTLFGDELCK	0.723	0.105
10	YNGVFQECQAEDK	0.6728	0.0107
11	LCVLHEKTPVSEK	0.6534	0.0274
12	LFTFHADICILPDIEK	0.5369	0.324
13	ETYGDMADCCKEQPER	0.4904	0.1069
14	EYEATLECCAK	0.481	0.0009
15	YICDNQDTISSK	0.3989	0.058
16	DLGEEHFK	0.362	0.4968
17	ETYGDMADCCKEK	0.3613	0.0691
18	MPCTEDYLSLILNR	0.359	0.3218
19	ECCHGDLECADDR	0.3574	0.5894
20	MPCTEDYLSLILNR	0.3414	0.1117
21	CCAADDKEACFAVEGPK	0.3088	0.0567
22	TCVADESHAGCEK	0.2564	0.0532
23	ETYGDMADCCKEK	0.2541	0.1858
24	QNCDQFEK	0.1795	0.1333
25	ECCHGDLECADRADLAK	0.1687	0.0201
26	CCAADDKEACFAVEGPK	0.1432	0.006
27	LFTFHADICILPDIEK	0.0421	0.2592
28	DTHKSEIAHR	0.0406	0.4331
29	ETYGDMADCCKEK	0.0106	0.5161
30	CCIESLVNR	0.0087	0.0184
31	LKPDPNTLCDEFK	0.0035	0.4319
32	HPYFYAPELLYYANK	0.0001	0.2335

Figure 1. Comparison of R² value of peptide peak area for BSA with peptide concentration for online and off-line method.

protein digestion. Peptide peak intensity or spectral count is measured for individual LC-MS/MS runs and changes in

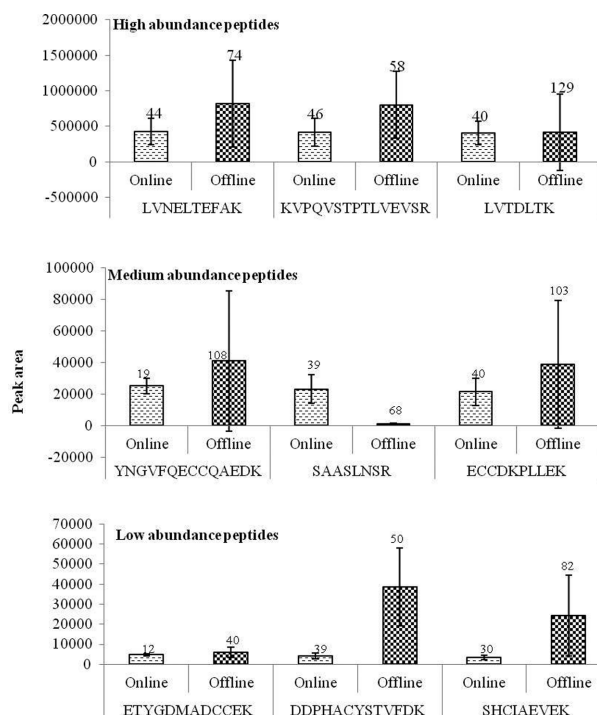


Figure 2. Distribution of peak area of common peptides according to their abundance identified in BSA.

protein abundance are calculated via a direct comparison between different analyses.¹¹ We observed a clear increase in peptide quantity with increase in initial sample concentration indicating peptide peak intensity is directly proportional to protein quantity in sample processed using online system. Figure 2 represents the average peak intensities of common high, medium and low abundance peptides with their standard deviation and error percentage. Albumin processed with online system shows consistence results with very small standard deviation and standard error percentage. Since label free quantification utilizes data from replicate run a greater difference in peak intensity between replicates will give an undesired result when in comparative proteomics. Albumin subjected to online proteolytic digestion and desalting resulted in quantitative peptides directly proportionate to the protein level in sample with a very little deviation between analyses.

Reproducibility

Real time sample validation

The present method was validated using reduced and alkylated proteins from A549 cells using online method with the same experimental conditions. The reproducibility ability of online system was determined by comparing equal loading (8 replicates) of protein and peptides for sample preparation and mass spectrum analysis respectively. The elution profile of the replicates exhibits very similar features and we were able to identify 61±6 proteins per replicate and a whole of 187 proteins were identified using sorcerer data base search tool

Table 2. Distribution of peak area of identified peptides in real time sample (A549)

	Peptide Sequence	Average peak area	SD	SE percentage
High Abundance	DASSVKR	1115483	71144	6.3
	NKPGVYTK	655868	67967	10.3
	KILASVQHMK	81224	9034	11.12
Medium Abundance	VILSIGGVVLVNLAGESEK	59182	6753	11.4
	DMEDTDDDDDDDD	58721	4205	7.1
	KVASALVPGRR	53982	9177	17.0
Low Abundance	YALYDATYETK	14541	673	4.6
	ASLENSLR	11250	772	6.8
	QITVNDLPVGR	10880	446	4.0

(data not shown). The reproducibility of the method was determined based on the standard deviations and standard error percentage of the average peak area intensity from eight replicate samples using ideal-Q. Table 2 shows the common high, medium and low abundance peptides in the samples. The standard error percentage is very low in peptides eluted from online method compared to the conventional in-solution method which can be beneficial in quantitative identification of protein. This shows the online digestion is reproducible.

Summary

The result of the present study shows online digestion produces quantifiable peptides which will be a huge advantage for label free quantification. Identification of increased number of peptides and proteins which cannot be accurately quantified will not help in identifying potential biomarkers for disease conditions. Hence there is a need for more precise methodologies to refine the sample before taken forward to analytical instruments and software tools. The present online method will be a most reliable in terms of comparative proteomics based approach using high throughput data with less time.

Acknowledgements

Funds provided by the Ministry of Health and Welfare (A111218-1212-1070100), Fountain Research Section of National Research Foundation (No. 20120006314) through Green Technology of the Ministry of Education, Science and Technology, Korea and the National Research Foundation (No. 2012036675) through Ministry of Education, Science

and Technology, Korea, and (TGF-b Biology and Therapeutics (No. R322012000102150) of the World Class University.

References

- Zhang, G.; Ueberheide, B. M.; Waldemarson, S.; Myung, S.; Molloy, K.; Eriksson, J.; Chait, B. T.; Neubert, T. A.; Fenyo, D. *Methods Mol. Biol.* **2010**, *673*, 211.
- Zhang, G.; Fenyo, D.; Neubert, T. A. *J. Proteome Res.* **2009**, *8*, 1285.
- Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. USA.* **1993**, *90*, 5011.
- Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem.* **1993**, *65*, 877.
- Sitek, B.; Waldera-Lupa, D. M.; Poschmann, G.; Meyer, H. E.; Stühler, K. *Methods Mol. Biol.* **2012**, *893*, 241.
- Eng, J. K.; Fischer, B.; Grossmann, J.; Maccoss, M. J. *J. Proteome Res.* **2008**, *7*, 4598.
- Tsou, C. C.; Tsai, C. F.; Tsui, Y. H.; Sudhir, P. R.; Wang, Y. T.; Chen, Y. J.; Chen, J. Y.; Sung, T. Y.; Hsu, W. L. *Mol. Cell Proteomics* **2010**, *9*, 131.
- Burkhart, J. M.; Schumbrutzki, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R. P. *J. Proteomics* **2012**, *75*, 1454.
- Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, Y.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H. *J. Proteome Res.* **2010**, *9*, 5422.
- Xu, F.; Wang, W. H.; Tan, Y. J.; Bruening, M. L. *Anal. Chem.* **2010**, *82*, 10045.
- Wenhong, Z.; Jeffrey, W. S.; Chun-Ming, H. *J. Biomed. Biotechnol.* **2010**, *18*, 6.