Charge state-selective separation of peptides by reversible modification of amino groups and strong cation-exchange chromatography: Evaluation in proteomic studies using peptide-ce...

ARTICLE *in* JOURNAL OF PROTEOMICS · MAY 2011

Impact Factor: 3.89 · DOI: 10.1016/j.jprot.2011.04.029 · Source: PubMed

12 AUTHORS, INCLUDING:

Lazaro Betancourt
Center for Genetic Engineering and Biotech...
55 PUBLICATIONS **774** CITATIONS
SEE PROFILE

Yasset Perez-Riverol
EMBL-EBI
38 PUBLICATIONS **979** CITATIONS
SEE PROFILE

Gabriel Padron
University of Barcelona
103 PUBLICATIONS **1,350** CITATIONS
SEE PROFILE

Vladimir Besada
Center for Genetic Engineering and Biotech...
102 PUBLICATIONS **1,048** CITATIONS
SEE PROFILE

JOURNAL OF PROTEOMICS 74 (2011) 2210–2213

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Technical note

Charge state-selective separation of peptides by reversible modification of amino groups and strong cation-exchange chromatography: Evaluation in proteomic studies using peptide-centric database searches

Lázaro H. Betancourt[a,*], *Aniel Sánchez*[a], *Yasset Pérez*[a], *Jorge Fernandez de Cossio*[a], *Jeovanis Gil*[a], *Patricia Toledo*[a], *Seiji Iguchi*[b], *Saburo Aimoto*[b], *Luis J. González*[a], *Gabriel Padrón*[a], *Toshifumi Takao*[b], *Vladimir Besada*[a]

[a]*Center for Genetic Engineering and Biotechnology, Ave. 31 e/158 y 190, Cubanacán, P.O. Box 6162, 10600 Habana, Cuba*
[b]*Institute for Protein Research, Osaka University, Yamadaoka 3-2, Suita, Osaka 565-0871, Japan*

ARTICLE INFO

Article history:
Received 16 February 2011
Accepted 29 April 2011
Available online 9 May 2011

Keywords:
Proteomics
Selective-peptide-isolation
Cation-exchange
Peptide-centric-database

ABSTRACT

Here we describe an integrated approach for the selective separation of peptides from complex mixtures using strong cation-exchange chromatography. The procedure exploits the charge differences produced by reversible modification of primary amino groups in peptides, enabling their separation into three major fractions: 1) neutral peptides 2) peptides with one positive charge and 3) peptides with 2 or more positive charges. The procedure demonstrated an excellent selectivity which allowed restricted MS/MS ion searches with peptide-centric databases.

Comprehensive proteome analysis requires the identification and quantification of individual species in samples containing thousands of proteins with abundances spanning a range of several orders of magnitude. The currency of proteome analysis by mass spectrometry is the peptides generated from the proteolysis of a sample, which results in a significant increase of complexity. This problem is usually tackled through the use of multidimensional liquid chromatography techniques. Alternative strategies, such as the reduction of this complexity by obtaining representative subsets of target peptides, selectively isolated on the basis of their low-abundance amino acid content have also been extensively tested [1–7]. Selective isolation methods, however, usually fail to produce a peptide subset that is fully representative of the protein population. Furthermore, non-target peptides are frequently discarded from the analysis, losing complementary and potentially valuable information.

Our group has reported several selective isolation methods based on chemical modification of peptide primary amino groups [8–12]. In this work, an integrating approach of these methods into a new separation scheme was developed. The general workflow is showed in Fig. 1. The mixture of proteins is digested with trypsin. Then, α- and ε-amino groups of the peptides are chemically modified through a reversible reaction using 2-(Methylsulfonyl)ethyl succinimidyl carbonate (NHS-Msc) [13]. A reversible blocking group was chosen to

* *Corresponding author*. Tel.: +53 7 271 6022.
E-mail address: lazaro.betancourt@cigb.edu.cu (L.H. Betancourt).

1874-3919/$ – see front matter
doi:10.1016/j.jprot.2011.04.029

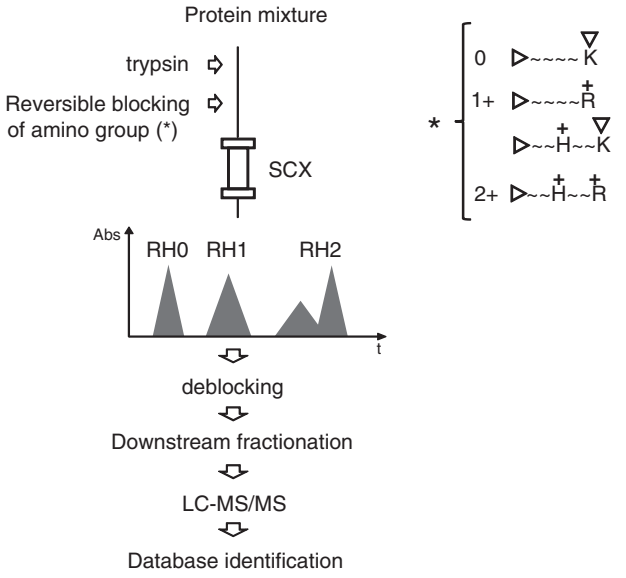

Fig. 1 – Scheme of the procedure for the selective separation and identification of RH0, RH1 and RH2 tryptic peptides from a complex mixture of proteins. The open triangle represents the reversible blocking group.

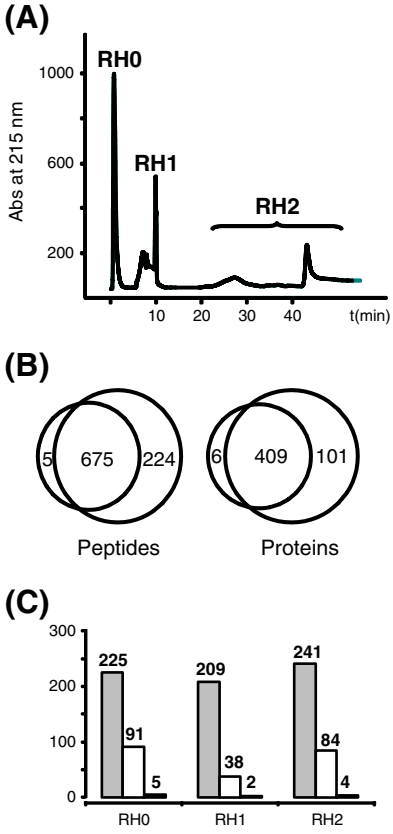

Fig. 2 – Analysis of a soluble protein extract from HuH7 cells. (A) SCX profile of RH0, RH1 and RH2 peptides. (B) Vent diagram of total number of peptides and proteins identified by protein sequence databases (smaller circles) and peptide-centric database (larger circles) searches. (C) Classification and distribution of peptides identified with peptide-centric databases. Gray bars (true positives): same assignment by each search type. White bars (false negatives): same assignment by each search type, but scored below the threshold in protein sequence database search for an FDR of 1%. Black bars (different assignment): represent cases where the assignment was different between the two search strategies.

allow the analysis of unmodified peptides, which improve mass detection and facilitate protein identification with automatic MS/MS database search engines. The reaction eliminates the positive charge of α- and ε-amino groups under acidic conditions, allowing the selective separation with a strong cation exchanger (SCX) of the mixture of derivatized peptides into three groups: 1) RH0 [R+H=0], that is, peptides devoid of histidine and arginine; 2) RH1 [R+H=1] representing those peptides containing either a histidine or an arginine; and 3) RH2 [R+H≥2] which stands for peptides with more than one histidine and/or arginine residue. The RH0 peptides are collected in the flow-through from the SCX step, the RH1 peptides are eluted in a single step of increased salt concentration, and the RH2 peptides are eluted with a steep gradient. Each peptide pool is then subjected to a procedure that regenerates their free amino groups and, if necessary, further fractionated before LC-MS/MS analysis.

The ability of the proposed method to exclusively separate peptides according to their charge state was initially evaluated on recombinant streptokinase (SKr) as a model protein and on an artificial mixture of four proteins (SKr, p64K, cytochrome c and apotransferrin). All the mass spectrometry-detected peptides were successfully sorted into their corresponding SCX fraction (data not shown) indicating the highly selective of this method.

A soluble protein extract from human liver carcinoma Huh7 cell line was also studied, using 100 μg of protein (see Supplementary data: Experimental section). Fig. 2A shows the SCX profile obtained. The three peptide groups RH0, RH1 and RH2 still display a high degree of complexity requiring additional fractionation. We chose the high pH RP approach since it is robust and user-friendly system and has shown a high degree of orthogonality with low pH RP separations. Each SCX-enriched peptide pool was separated by RP-HPLC at basic pH into 10 fractions, using an ACN/H_2O/NH_3 solvent system that also works as desalting step, prior LC-MS/MS.

A total of 680 peptide assignments (FDR=1%) were made corresponding to 415 protein identifications. Approximately 22% (92) of the proteins were identified exclusively in the RH0 fraction, 20% (83) in the RH1, 29% (121) in the RH2, and the rest 29% (119) were found in at least two of the three SCX fractions. The numbers of specific vs. unspecific peptides identified in each SCX fraction behave as follows: for RH0, 225 vs. 2; for RH1, 210 vs. 2; and for RH2, 241 vs. 5; which resulted in a selectivity of at least 98% in every case. This outcome allowed exploratory searches against peptide-centric databases.

When peptide-centric databases were used, 69% of the proteins were identified in only one of the SCX fractions (distributed as 23% (116) in RH0, 16% (83) in RH1 and 30% (151) in RH2). The remaining 31% (139) was comprised by species

detected in two or three of the fractions. Protein sequence database searches were clearly outperformed by peptide-centric databases searches, which yielded increases of 41%, 17% and 33% for the number of identified peptides and increases of 36%, 14% and 30% for the number of identified proteins in the RH0, RH1 and RH2 fractions, respectively. In addition, the use of peptide-centric database searches increased the number of proteins identified by two or more peptides: overall, 95 of the proteins previously identified with protein sequence database searches gained a total of 115 additional peptide identifications. An examination of the peptide assignments carried out with peptide-centric database searches reveals a match with the assignments performed with protein sequence database searches in 675 of the cases, with 224 new exclusive assignments and only 5 missed identifications in the former (Fig. 2B). This translates to 409 common proteins as well as 101 new identifications and only 6 missing identifications for the peptide-centric searches. Combining the results of both searches, a total of 516 proteins were identified from 904 different peptides. (For details about identifications see Supplementary data: Results from protein (IPI) and peptide (PEP) databases).

Fig. 2C shows the distribution and classification of peptides identified by peptide-centric databases according to their previous assignment by protein sequence database searches. As expected, the increase of assignments mostly originates from identification of lower scoring peptides (white bars). The improvement in identifications agrees with previous reports [14–16]. The search space is smaller in a peptide-centric database compared to a protein database; consequently better FDRs can be reached with similar peptide scores. This allowed lower scoring peptides to be accepted at the same FDR selected for protein database search.

In general, the results from the SCX fractions were highly complementary, as 70% of the proteins were identified in a single fraction for both database search experiments. Consequently, the overlap of identified proteins between the different fractions was small; for instance, less than 7% of all proteins were identified in all three fractions. This implies that all SCX fractions should always be analyzed.

The number of reports describing approaches similar to the one presented here is not frequent. Martens et al., for instance, used the COFRADIC technology for selection of methionyl, cysteinyl and amino-terminal peptides to characterize the proteome of human platelets [17]. In that study the sample had to be processed independently 3 times in order to target each peptide subgroup, a time-consuming work which also discarded a large portion of the sample and potentially valuable information.

More recently, Taouatas et al. separated Lys-N generated peptides by SCX into four categories: acetylated N-terminal peptides; singly phosphorylated peptides containing a single basic (Lys) residue; peptides containing a single basic (Lys) residue; and peptides containing more than one basic residue [18]. This work highlighted the potentialities of SCX for charge-selective separation of peptides in complex mixtures.

Several studies have addressed the analysis of rejected peptide identifications by "digging" into statistically non-significant hits [19–21]. In this sense, our attention was drawn in particular to the work of Yen et al., who attempted to increase sensitivity in shotgun proteomics by using peptide-centric database searches, applying peptide elution rules in SCX, and excluding unlikely missed tryptic cleavages [20]. However, the overlap between SCX peptide fractions in their methodology could have made unfeasible to take full advantage of the proposed procedure, which therefore yields only modest improvements in the number of identified peptide and proteins.

The present study suggested that complex mixtures of tryptic peptides, such as those generated in proteomic experiments, could be fractionated by SCX in a single workflow into subsequent, well-defined and highly enriched peptide subsets, and without discarding any peptide fraction. The procedure sacrifices the resolution that SCX typically displays with unmodified tryptic peptides, but this is compensated by RP separation at basic pH of each SCX peptide pool. Furthermore, the truly control of charge based separation and gain in selectivity, results in useful information to validate identifications and to effectively use peptide database searches.

The noticeable small numbers of peptide and protein identified was attributed to the low performance of the mass spectrometer used. It is expected that the use of state-of-art instrument will yield higher number of peptide and protein identifications, which in turn will confirm the high selectivity achievable with the present approach.

Supplementary materials related to this article can be found online at doi:10.1016/j.jprot.2011.04.029.

Acknowledgments

The authors wish to thank INSPUR (China) for the kind donation of the TS10000 computer cluster used in a large part of the bioinformatics study. We are also grateful to Mr. Alejandro M. Martin (CIGB, Cuba) for English correction of the manuscript.

REFERENCES

[1] Leitner A, Lindner W. Chemistry meets proteomics: the use of chemical tagging reactions for MS-based proteomics. Proteomics 2006;6:5418–34.
[2] Olsen JV, Andersen JR, Nielsen PA, Nielsen ML, Figeys D, Mann M, et al. HysTag — a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. Mol Cell Proteomics 2004;3:82–92.
[3] Wang H, Qian WJ, Chin MH, Petyuk VA, Barry RC, Liu T, et al. Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. J Proteome Res 2006;5:361–9.
[4] Fang X, Zhang WWJ. Affinity separation and enrichment methods in proteomic analysis. Proteomics 2008;71:284–303.
[5] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999;17:994–9.
[6] Gevaert K, Van Damme J, Goethals M, Hoorelbeke B, Demol H, Martens L, et al. Isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 *Escherichia coli* proteins. Mol Cell Proteomics 2002;1:896–903.
[7] Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas G, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol 2003;21:566–9.

[8] Betancourt L, Gil J, Besada V, González LJ, Fernández-de-Cossio J, García L, et al. SCAPE: a new tool for the Selective CApture of PEptides in protein identification. J Proteome Res 2005;4:491–6.
[9] Sánchez A, González LJ, Ramos Y, Betancourt L, Gil J, Besada V, et al. Selective isolation of lysine-free tryptic peptides delimited by arginine residues: a new tool for proteome analysis. J Proteome Res 2006;5:1204–13.
[10] Sánchez A, González-López LJ, Betancourt L, Gil J, Besada V, Fernández-de-Cossío J, et al. Selective isolation of multiple positively charged peptides for 2-DE-free quantitative proteomics. Proteomics 2006;16:4444–55.
[11] Gil J, Betancourt LH, Sardiñas G, Yero D, Niebla O, Delgado M, et al. Proteomic study via a non-gel based approach of meningococcal outer membrane vesicle vaccine obtained from strain CU385: a road map for discovering new antigens. Hum Vaccin 2009;5:347–56.
[12] Sanchez A, Sun W, Wang L, Ma J, Betancourt L, Gil J, et al. Letter: selective isolation-detection of two different positively charged peptides groups by strong cation exchange chromatography and matrix-assisted laser desorption/ionization mass spectrometry: application to proteomics studies. Eur J Mass Spectrom 2010;16:693–702.
[13] Brik A, Keinan E, Dawson PE. Protein synthesis by solid-phase chemical ligation using a safety catch linker. J Org Chem 2000;65:3829–35.
[14] Resing KA, Ahn NG. Proteomics strategies for protein identification. FEBS Lett 2005;579:885–9.
[15] Rudnick PA, Wang Y, Evans E, Lee ChS, Balgley BM. Large scale analysis of MASCOT results using a mass accuracy-based threshold (MATH) effectively improves data interpretation. J Proteome Res 2005;4:1353–60.
[16] Martens L, Van Damme P, Van Damme J, Staes A, Timmerman E, Ghesquière B, et al. The human platelet proteome mapped by peptide-centric 207 proteomics: a functional protein profile. Proteomics 2005;5:3193-204.
[17] Taouatas N, Maarten Altelaar AF, Drugan MM, Helbig AO, Shabaz M, Heck AJR. Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. Mol Cell Proteomics 2009;8:190–200.
[18] Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, et al. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. Anal Chem 2004;76:3556–68.
[19] Wielsch N, Thomas H, Surendranath V, Waridel P, Frank A, Pevzner P, et al. Rapid validation of protein identifications with the borderline statistical confidence via de novo sequencing and MS BLAST searches. J Proteome Res 2006;5: 2448–56.
[20] Tharakan R, Edwards N, Graham DRM. Data maximization by multipass analysis of protein mass spectra. Proteomics 2010;10:1160–71.
[21] Yen CY, Steve R, Mendoza AM, Meyer-Arendt K, Sun S, Cios KJ, et al. Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra. Anal Chem 2006;78:1071–84.