Evaluation of Phenylthiocarbamoyl-Derivatized Peptides by Electrospray Ionization Mass Spectrometry: Selective Isolation and Analysis of Modified Multiply Charged Peptides for Liqu…

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2010

Impact Factor: 5.64 · DOI: 10.1021/ac1012738 · Source: PubMed

Anal. Chem. **2010,** *82,* 8492–8501

Evaluation of Phenylthiocarbamoyl-Derivatized Peptides by Electrospray Ionization Mass Spectrometry: Selective Isolation and Analysis of Modified Multiply Charged Peptides for Liquid Chromatography−Tandem Mass Spectrometry Experiments

Aniel Sanchez,[†] Yasset Perez-Riverol,[‡] Luis Javier González,[†] Jesus Noda,[†] Lazaro Betancourt,[†] Yassel Ramos,[†] Jeovanis Gil,[†] Roberto Vera,[‡] Gabriel Padrón,[†] and Vladimir Besada*[,†]

Proteomics Department and Bioinformatic Department, Physical-Chemistry Division, Biomedical Research, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana, Cuba

Edman degradation in the gas phase has been observed by collision activated dissociation of *N*-terminal phenylthiocarbamoyl (PTC) protonated peptide to yield abundant complementary b_1 and y_{n-1} ion pairs. Here, we demonstrated the relation between the observed losses of aniline and/or the entire PTC derivatizing group with the availability of mobile protons using electrospray ionization mass spectrometry. In order to select the peptides with more efficient fragmentation, while simplifying the mixture of peptides, we extend the phenylisotiocyanate (PITC) derivatization of amino groups to the selective isolation of multiply charged peptides (those having the number of arginines and histidines residues higher than one) using a procedure previously developed in our group. Thus, it was possible to identify in the filtered protein database the sequence of the isolated multiply charged peptides derived from a single protein and a complex mixture of proteins extracted from *Escherichia coli* using only the molecular mass and the N-terminal amino acid information. For this purpose, we developed a novel bioinformatic tool for automatic identification of peptides from liquid chromatography−tandem mass spectrometry (LC−MS/MS) experiments, which potentially can be used in high-throughput proteomics.

Proteomics based on multidimensional liquid chromatography and tandem mass spectrometry (MS/MS) to separate and fragment peptides have allowed the highest throughput for protein identification. This is in spite of the very low percentage of MS/MS spectra from liquid chromatography−mass spectrometry (LC−MS) that can be correctly assigned.[1] However, the measured peptide mass acts as a filter that directly reduces the number of potential false positive assignments.[2,3] With good scoring, higher mass accuracy proportionately increases the certainty of identification, a concept that applies to intact peptide as well as their fragmentation products.

The peptide molecular mass by itself, even with low ppm accuracy, is normally considered nonsufficient evidence for identification of the peptide,[4] but their combination with partial sequencing can yield determinant information for identification of peptide and protein in the database.

Gaskell et al.[5] showed that the equivalent of a single Edman degradation stage can be realized for ions in the gas phase. The peptides are coupled in solution with phenylisotiocyanate (PITC) and ionized either in an electrospray ionization (ESI) or a matrix-assisted laser desorption ionization (MALDI) ion source. After activation, the phenylthiocarbamoyl (PTC)-derivatized peptide dissociates specifically to yield the b_1 fragment. In consequence, this opens the possibility to determine the peptide mass and its N-terminal residue in a single mass spectrum.

Gaskell and co-workers used this reaction and, supported by bioinformatic tools, identified yeast proteins using gel electrophoresis and MALDI-MS.[6] This procedure was also applied to an apomyoglobin tryptic digestion using a Fourier transform ion cyclotron resonance MS (ESI-FTICR).[7] Some PTC-derivatizated peptides showed loss of aniline and PITC neutral group instead

* Corresponding author. Vladimir Besada Ph.D., Head of Proteomics Department, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Ave 31, e/158 y 190, Cubanacán, Playa, Ciudad de la Habana, Cuba. E-mail: vladimir.besada@cigb.edu.cu.

[†] Proteomics Department.

[‡] Bioinformatic Department.

(1) Yen, C. Y.; Russell, S.; Mendoza, A. M.; Meyer-Arendt, K.; Sun, S.; Cios, K. J.; Ahn, N. G.; Resing, K. A. *Anal. Chem.* **2006**, *78*, 1071–1084.

(2) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513–523.

(3) Fang, R.; Elias, D. A.; Monroe, M. E.; Shen, Y.; McIntosh, M.; Wang, P.; Goddard, C. D.; Callister, S. J.; Moore, R. J.; Gorby, Y. A.; Adkins, J. N.; Fredrickson, J. K.; Lipton, M. S.; Smith, R. D. *Mol. Cell. Proteomics* **2006**, *5*, 714–725.

(4) Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18132–18138.

(5) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. *J. Mass Spectrom.* **1997**, *32*, 225–231.

(6) Brancia, F. L.; Butt, A.; Beynon, R. J.; Hubbard, S. J.; Gaskell, S. J.; Oliver, S. G. *Electrophoresis* **2001**, *22*, 552–559.

10.1021/ac1012738
Published on Web 09/20/2010

of a specific b_1 fragment, and the fragmentations were not generalized neither completely explained as the authors pointed out in the conclusions of the manuscript.[7] Finally, the results were attractive for identification of proteins in database only for low parts per million (ppm) accuracy, due fundamentally to the complexity of the proteomes. In both papers, the authors demonstrate the feasibility of the reaction and the potential application for proteome analysis.

In addition, Wohlhueter and co-workers demonstrated that other isothiocyanate analogues bearing basic moieties can derivatize peptides and significantly improve the MS sensitivity of tagged analytes, while promoting Edman fragmentation.[8] On the other hand, Yao and co-workers proposed this active chemical modification for improving the fragment ion detection in multiple reaction monitoring (MRM) studies.[9]

In this work, we studied the fragmentation pattern of PTC derivatized peptides using electrospray ionization mass spectrometry. We used a mixture of peptides with several charge states, observing a similar behavior for each charge state group explained by mobile proton theory.

In order to increase the number of PTC-derivatized peptides with highly efficient fragmentation (considering the abundant complementary b_1 and y_{n-1} pairs) while simplifying the mixture of peptides, we combined the PITC reaction with a selective isolation of multiply charged peptides as published previously[10] (those which the sum of arginines and histidines residues are higher than one, RH peptides). Initially, the *in silico* analysis of PTC-derivatized RH peptides for several proteomes showed a considerable increase in the number of peptides that could be identified based on their mass and N-terminal residue information only (unique peptides).

Finally, a single mixture of peptides produced by tryptic digestion of streptokinase and a complex mixture of peptides from *Escherichia coli* were analyzed, and protein identification was achieved using only the molecular mass, the N-terminal amino acid information, and a filtered database composed exclusively of RH peptides.

MATERIALS AND METHODS

The sequencing-grade trypsin was purchased from Promega (WI). Acetonitrile (ACN) and water were HPLC grade and were obtained from Caledon (Ontario, Canada). The peptides were synthesized by the Fmoc chemistry, and the recombinant streptokinase (rSK) were manufactured at the Center for Genetic Engineering and Biotechnology (Havana, Cuba). TFA, formic acid, and PITC were purchased from Pierce (IL). The bovine serum albumin (BSA) was obtained from Sigma (MO). *E. coli* strain W3110 cells were processed by sequential solubilization described by Molloy,[11] and the fraction soluble in Tris was analyzed with the present procedure.

Reduction and S-Alkylation. The proteins were dissolved separately in 50 μL of 200 mM HEPES buffer (pH 8.0) containing 2 M of guanidium hydrochloride and were reduced by adding DTT to a final concentration of 10 mM, flushed with nitrogen, and incubated 4 h at 37 °C. The free thiol groups were alkylated by adding acrylamide to a final concentration of 20 mM, and the reaction proceeded for an additional 60 min at room temperature.

Tryptic Digestions. The proteins, dissolved in the reduction and S-alkylation buffer, were diluted with two volumes of water and digested with trypsin using an enzyme−substrate ratio of 1:100 at 37 °C during 10 h.

PITC Derivatization. After trypsin digestion, the mixture of peptides was dried under vacuum centrifugation and then converted into a PTC derivative by dissolving the mixture in ethanol−water−triethylamine−phenyl isothiocyanate (75/10/10/5 by volume). The reaction proceeded for 20 min at 50 °C followed by vacuum centrifugation. The dried, derivatized product was first dissolved in heptane−ethyl acetate (90:10 v/v; 100 μL), and then an equal volume of water was added. This mixture was shaken vigorously and centrifuged. The upper phase was discarded, and the extraction of the aqueous phase was repeated twice before ESI MS analysis.

Selection of Multiply Charged PTC-Peptides (RH Peptides) by Strong Cation Exchange Chromatography. The PTC-derivatized peptides were desalted using ZipTipC18 (Millipore), and the multiply charged peptides (RH peptides) were selectively captured using a SCX minicolumn, NuTip (Glygen Corp). The NuTip was equilibrated with TFA (0.05%), and peptides were absorbed onto the NuTip and extensively washed with the same equilibration buffer. RH peptides were eluted with a solution containing 1 M NaCl and 20% ACN. The sample was desalted using ZipTipC18 (Millipore) previous to the analysis by nanoESI-MS.

MS Analysis. Mass spectrometric measurements were performed using a hybrid quadrupole orthogonal acceleration tandem mass spectrometer QTof-2 (Micromass, Manchester, U.K.). Capillary and cone voltages were 900 and 35 V, respectively, in the nanoESI experiment. The spectra were acquired in the m/z range from 400 to 2000 Th. Data acquisition and processing were performed using a Masslynx system (version 3.5) from Micromass (Manchester, U.K.).

MS/MS Analysis. The solution of multiply charged PTC-derivatized peptides from rSK was injected by infusion mode to the mass spectrometer QTof-2 (Micromass, Manchester, U.K.) using a metal-coated borosilicate capillary (Micromass, U.K.).

For the analysis of *E. coli* proteins the LC−MS/MS were performed on an integrated nanoLC Agilent 1100 equipped with a microautosampler. The injected peptides from 40 μL were trapped and desalted on a PS-DVB monolithic trap column (5 mm × 200 μm i.d.) from Dionex (Sunnyvale, CA) during 20 min with 0.1% of formic acid delivered by an auxiliary pump at 20 μL/min. With the valve switched, the retained peptides were back-flushed and loaded onto the capillary PS-DVB monolithic column (50 mm × 100 μm i.d.) from Dionex (Sunnyvale, CA). The separation column was previously equilibrated in solution A (formic acid 0.2% in water), and the peptides were eluted with a linear gradient increasing the concentration of solution B (acetonitrile/water 80/

(7) Van der Rest, G.; He, F.; Emmett, M. R.; Marshall, A. G.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 288–295.
(8) Wang, D.; Fang, S.; Wohlhueter, R. M. *Anal. Chem.* **2009**, *81*, 1893–900.
(9) Diego, P. A.; Bajrami, B.; Jiang, H.; Shi, Y.; Gascon, J. A.; Yao, X. *Anal. Chem.* **2010**, *82*, 23–27.
(10) Sanchez, A.; Gonzalez, L. J.; Betancourt, L.; Gil, J.; Besada, V.; Fernandez-de-Cossio, J.; Rodriguez-Ulloa, A.; Marrero, K.; Alvarez, F.; Fando, R.; Padron, G. *Proteomics* **2006**, *6*, 4444–4455.
(11) Molloy, M. P.; Herbert, B. R.; Walsh, B. J.; Tyler, M. I.; Traini, M.; Sanchez, J. C.; Hochstrasser, D. F.; Williams, K. L.; Gooley, A. A. *Electrophoresis* **1998**, *19*, 837–44.

20 v/v) at 0.75%/min. The eluates at a flow rate of 300 nL/min were delivery to the QTof-2 mass spectrometer.

For both injection modes, the precursor ions were selected to be fragmented automatically once their intensity rose above a defined threshold (4 counts s^{-1}), and each MS/MS step was completed after two scans of 2 s. The total acquisition time was 2 min. To acquire the MS/MS spectra, the first quadrupole was used to select the precursor ion within a window of 4 Th. A pressure of $\sim 3 \times 10^{-2}$ Pa collision gas (argon) was used in the hexapole collision cell to yield the fragment ions. The collisions ramps used for PTC-derivatized peptides fragmentation are shown in the Supporting Information, Supplementary Material 2.

Data acquisition and processing were performed using a Masslynx system (version 3.5) from Micromass (Manchester, U.K.). Protein identification was made using the Internet-available search engine MASCOT[12] (http://www.matrixscience.com) or an in house program written in Java. The mgf files were obtained from the raw data using the Mascot distiller program.[12]

Computer in House Programs. AAmanager is a computer program coded in Java that was developed for console use to scan the Swiss-Prot sequence database and calculates (1) the number of proteins of a given organism that might be identified by isolating selective RH peptides with molecular masses comprised between 800 and 3500 Da, (2) the total number of tryptic peptides per protein (KR peptides/protein) in the analyzed proteome, and (3) the average number of RH peptides/protein that could be isolated selectively using the method developed in our group.[10]

N terminal software was coded in Java that allows the identification of peptide sequences in the database from the information included in the mgf files. The scoring method used for each sequence is based on the experimental finding that approximately 70−90% of the total intensity of fragment ions observed in the MS/MS spectra of PTC-peptides that corresponds to the b_1 and its complementary y_{n-1} ions.[9] Generally, the most intense signal in the low-mass region could be assigned to the b_1 fragment. In consequence, with this report[9] the score was calculated as follows

$$I_n = I_i / I_{\max}$$

$$\text{score} = I_n / O_i$$

where I_i is the intensity of the possible b_1 fragment, I_n is the normalized intensity of the possible b_1 fragment, $I_{\max}$ is the intensity of the most intense ion in the low mass region (193−322 Da), O_i is the intensity order (sort descending) of I_i among all possible b_1 fragments. Peptide sequences with a score equal to 1 or equal to I_n were considered as confident identifications or probable identifications, respectively.

Algorithm to Identify Unique Peptides. A decision tree algorithm was developed in Java to compute the identification of unique peptides based on the above-described theoretical properties. Each node of the tree represents a property of the peptide (Nterm, molecular mass). The algorithm constructs a tree representing the possible combinations of the properties to identify unique peptides in the sequence database. The root node of the tree corresponds to the molecular mass of the peptide because it is the most discriminating property to reduce the number of peptides to be analyzed by the remaining nodes of the decision tree. In each node of the tree, the algorithm takes a set of input peptides and calculates the absolute values for a given property and stores only those that are unique peptides. Then, it sends the child nodes the remaining peptides as input data for further analysis in order to identify other unique peptides after considering the next property. The identification function for the unique peptides is the following logical expression:

$$(V_{n+1}\, V_{n-1}) > 4(\text{error})$$

A peptide sequence is unique if the property value (V_n) differs from its previous value V_{n-1} in 2 units of property error and in 2 units of error for the next peptide property value V_{n+1}. The algorithm provides as a final result a set of unique peptides after combining the N-terminal amino acids and the molecular mass at several parts per million of mass accuracy. In order to know if the number of unique peptides increases by using the selective isolation methods, the algorithm was run with the whole tryptic peptide database and the individual subdatabases containing the tryptic peptides were selectively isolated (multiply charged peptides).

RESULTS AND DISCUSSION

Influence of the Charge-State and the Presence of Basic Amino Acids on the Fragmentation Pattern of PTC- Peptides. Previous studies showed that several PTC-peptides ions do not yield an abundant b_1 fragment ion in the ESI-MS/MS, which seriously limits the information regarding the N-terminal amino acid because the loss of aniline and/or the entire PTC derivatizing group can compete with the Edman fragmentation pathway.[7] Nowadays, no general trend (based on nature of the N-terminal amino-acid, charge-state of the peptide, or availability of mobile protons) of these losses have been published to explain this behavior.[7] In this experiment, four synthetic peptides with sequences SSFSMLR, SSMSTLR, RSSYSTL, and QWISLGDR were derivatized with PITC and analyzed in an ESI-QTOF mass spectrometer for studying the influence of basic amino acids and the protonation extent of peptide ions in the gas phase on the fragmentation observed in MS/MS spectra, specifically the pairs b_1 and y_{n-1} fragments and the loss of aniline and/or the entire PTC derivatizing group. The fragmentation patterns of these peptides were studied using two different charge states: the singly- and the doubly charged ions. For singly charged peptides, similar spectra were observed, with predominant signals of the b_1 fragment and losses of aniline and PITC groups (see Figure 1).

According to the theory of the mobile proton for this kind of peptide, where the number of charges is equal to the number of arginine residues in the peptide sequence, the proton is located on the guanidinium group (fixed proton).[13] In consequence, the collision energy required to produce the b_1 fragment ion from the singly charged precursor must be higher than the energy needed to fragment the corresponding multiply charged forms, thus the collision energy used was higher than 35 eV for all

(12) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.

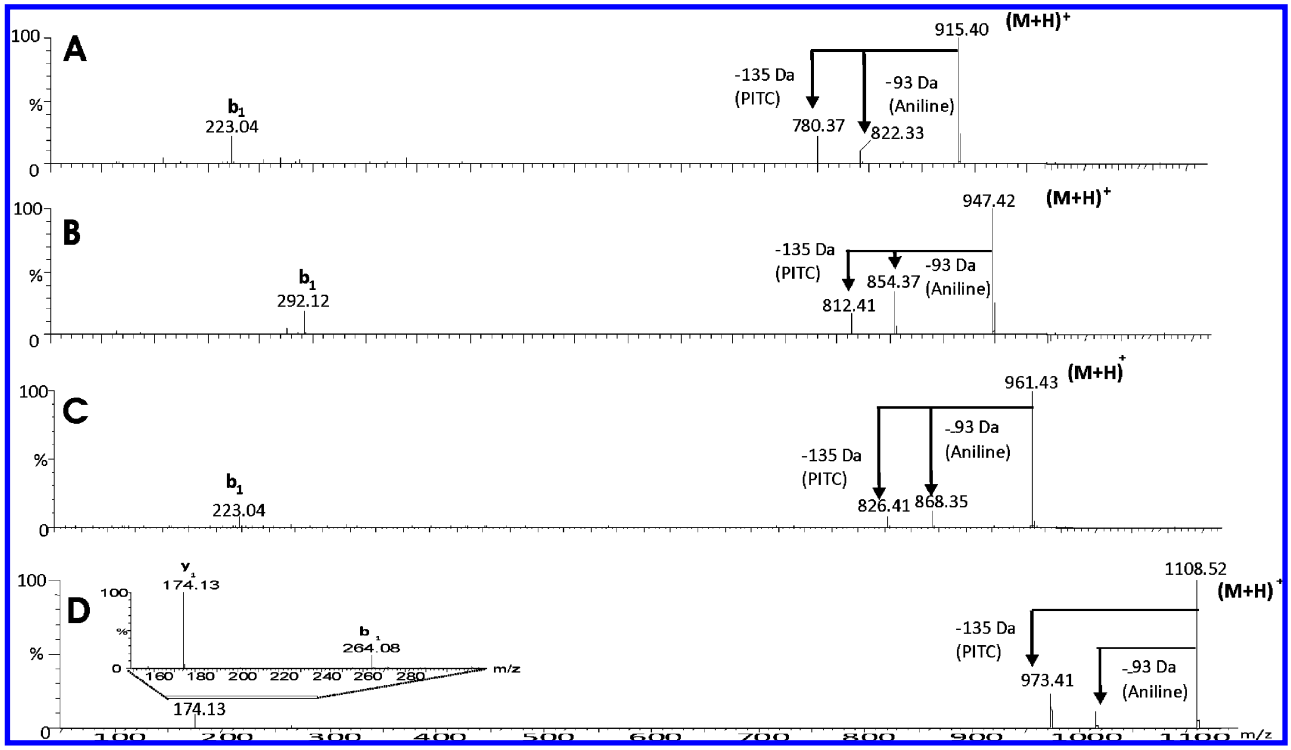

Figure 1. ESI-MS/MS spectra of single charged PTC-derivatized synthetic peptides: (A) SSFSMLR, (B) RSSYSTL, (C) SSMSTLR, and (D) QWISLGDR. The inset in part D showed the expanded low-mass region where the b_1 and y_1'' fragments are observed. All MS/MS spectra have a similar pattern, showing the loss of 93 and 135 Da beside the b_1 fragment.

analyzed peptides. In these conditions, only the signals corresponding to the loss of aniline and PTC groups appear with similar intensities to the b_1 fragment ion in the MS/MS spectra. This result suggests that the probability to obtain other fragment ions in addition to b_1 as well as the losses of aniline and the PITC group increases while increasing the collision energy; therefore, a more complex MS/MS spectrum is obtained.

However, an additional fragment was observed in the mass spectrum of peptide QWISLGDR, corresponding to the y_1 fragment (m/z = 174 Da, Figure 1D). It is well-known that the presence of aspartic acid residues in single charged peptides with a fixed proton (in arginine residues) provokes the C bias fragment promoted by the carboxylic proton.[14] In this case, the aspartic acid is adjacent to the C-terminal arginine residue. For doubly charged peptide analysis, the mass spectrum patterns were similar (see Figure 2). Here, the number of charges is higher than the number of arginine residues in each peptide sequence, therefore one proton is mobile. The energy used was lower than 20 eV, and the most abundant fragment ions were the b_1 and their complementary y_{n-1}, showing that gas-phase Edman degradation is more efficient for multiple charged peptides ions, as previously reported.[7] In this analysis, the peptide QWISLGDR was not observed as double charged ion in the experimental conditions used. Both results (for singly- and doubly charged peptides) demonstrate that it is possible to explain and predict the most intense signals in the mass spectrum for known peptide sequences, including the losses of aniline and/or the entire PTC derivatizing group according to the theory of the mobile proton. These findings are very useful to determine the PTC-peptides exhibiting more efficient fragmentation upon collision induced dissociation yielding intense b_1 and y_n-1 ions.

PTC-Derivatized Tryptic Digests of Two Model Proteins: Bovine Serum Albumin (BSA) and Recombinant Streptokinase (rSK). To increase the number of analyzed PTC-derivatized peptides and confirm the regularities observed above, two model proteins BSA and rSK were digested with trypsin and the resultant peptides modified with PITC. The reaction proceeded successfully for all the tryptic peptides of both proteins, and in some cases multiple PTC-derivatized peptides were observed due to the presence of lysine residues. Tryptic peptides (26) were subjected to MS/MS analysis under different collision energies, and their assignments are shown in Table 1. The minimum collision energy (MCE) was defined as the minimal energy necessary to obtain the b_1 fragment.

Signals of the PTC-derivatized peptides with one, two, or three positive charges were observed in the ESI MS spectrum. A more detailed analysis of the charge state of each peptide allows the classification of PTC-peptides into two main groups. The first one is composed by peptides with fixed protons (Table 1, peptides 1−9); in which the number of charges is equal to the number of arginines in their sequences. The peptides in the second group (Table 1, peptides 10−26) have more protons than arginines in each peptide sequence (mobile protons). The MS/MS spectra of peptides belonging to the first group (Table 1, peptides 1−9) showed some common characteristics, such as the intense signals due to the losses of aniline (− 93 Da) and the PITC group (− 135 Da) from the precursor ions that were obtained in all cases with an MCE greater than 30 eV. These results are in good agreement with those obtained for singly charged synthetic peptides derivatized with PITC described in the previous section. The peptides YLYEIAR (no. 3, Table 1) and NLDFR (no. 7, Table 1) showed y_3 and y_2 fragment ions, respectively, beside the b_1 fragment. These fragments could be easily explained due to the presence

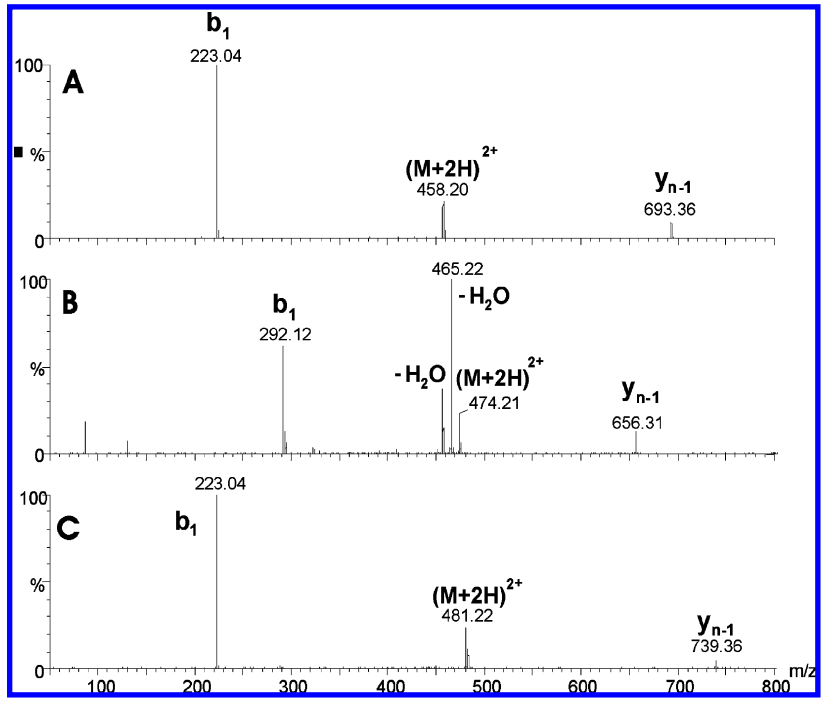

Figure 2. ESI-MS/MS spectra of doubly charged PTC-derivatized peptides: (A) SSFSMLR, (B) RSSYSTL, and (C) SSMSTLR. All spectra showed similar pattern with the b_1 and y_{n-1} as the most intense backbone fragment ions.

Table 1. Summary of the PTC-Derivatized Tryptic Peptides of Two Model Proteins Studied by ESI-MS/MS

	sequence	z^a	(R + H)^b	m/z	$b_1{}^c$	$y_{n-1}{}^d$	other fragments^e (Da)	protein^f	MCE^g (eV)
1	^{475}C*C*TESLVNR^{483} ^{h}	1	1	1301.55	nd^i	nd^i	$y_n - 93$, $y_n - 135$	BSA	45
2	^{77}VASLR^{81}	1	1	680.36	X	nd^i	$y_n - 93$, $y_n - 135$	BSA	32
3	^{137}YLYEIAR^{143}	1	1	1062.50	X	nd^i	y_1, y_3, $y_n - 93$, $y_n - 135$	BSA	45
4	^{336}RHPEYAVSVLLR^{347}	2	3	787.90	X	nd^i	$y_n - 93$, $y_n - 135$, 273	BSA	40
5	^{326}DLYDPR^{331}	1	1	913.37	nd^i	nd^i	y_2, y_1	rSK	30
6	^{396}EVYSYLR^{402}	1	1	1064.47	X	nd^i	y_1, 273	rSK	35
7	^{321}NLDFR^{325}	1	1	799.32	X	nd^i	y_1, y_2, $y_n - 93$, $y_n - 135$	rSK	35
8	^{311}SEQLLTASER^{320}	1	1	1268.57	nd^i	nd^i	y_1	rSK	35
9	^{373}RPEGENASYHLAYDKDR^{389}	2	3	1145.97	X	nd^i	$y_n - 93$, y − 135, 273	rSK	40
10	^{1}DTHK^{4}	1	1	770.28	nd^i	X	$y_n - 93$, $y_n - 135$, $y_{n-1} - 93$, $y_{n-1} - 135$	BSA	25
11	^{291}SHLK^{294}	1	1	754.30	X	X	$y_n - 93$, $y_n - 135$, [HL] = 251	rSK	28
12	^{373}RPEGENASYHLAYDKDR^{389}	3	3	764.32	X	X	$y_n - 93$, $y_n - 135$	rSK	27
13	^{373}RPEGENASYHLAYDK^{387}	2	2	1010.43	X		$y_n - 93$, $y_n - 135$	rSK	28
14	^{243}EC*C*HGDLLEBADDR^{256} ^{h}	2	1	963.89	X	X	nd^i	BSA	29
15	^{42}LVNELTEFAK^{51}	2	0	717.33	X	X	nd^i	BSA	15
16	^{459}LC*VLHEK^{461} ^{h}	2	1	591.77	X	X	nd^i	BSA	15
17	^{286}SHC*IAEVEK^{294} ^{h}	2	1	678.80	X	X	nd^i	BSA	23
18	^{52}TC*VADESHAGBEK^{64} ^{h}	2	1	881.35	X	X	nd^i	BSA	15
19	^{413}KVPQVSTPTLVEVSR^{427}	2	1	955.51	X	X	nd^i	BSA	18
20	^{5}SEIAHR^{10}	2	2	424.21	X	X	nd^i	BSA	18
21	^{336}RHPEYAVSVLLR^{347}	3	3	525.63	X	X	nd^i	BSA	18
22	^{234}TILPMDQEFTYHVK^{247}	2	1	996.42	X	X	nd^i	rSK	20
23	^{123}DGSVTLPTQPVQEFLLSGHVR^{143}	2	2	1208.14	X	X	nd^i	rSK	25
24	^{221}DSSIVTHDNDIFR^{233}	2	2	827.35	X	X	nd^i	rSK	25
25	^{211}THPGYTIYER^{220}	2	2	686.30	X	X	nd^i	rSK	20
26	^{355}VEDNHDDTNR^{364}	2	2	675.24	X	X	nd^i	rSK	20

^{a} Number of protons of the precursor selected for MS/MS analysis. ^{b} Corresponds to the number of positive charges the arginine and histidine residues bear in the α and ε-amino-blocked peptides dissolved at acidic pH. ^{c} X denotes the peptides giving the b_1 fragment. ^{d} X denotes the peptides giving the y_{n-1} fragment. ^{e} Additional fragments ions observed in the MS/MS spectra of the PTC-derivatized peptides. ^{f} Protein corresponding to the peptide analyzed. ^{g} Minimal collision energy (MCE) required for observing the b_1 fragment. ^{h} C* means propionamidated cystein residue. ^{i} nd means that the specified fragment ions were not detected in the MS/MS spectra using the minimal collision energy.

of glutamic acid or aspartic acid adjacent to the cleavage site that provokes the C bias fragment promoted for its carboxylic proton.[13] The same behavior was observed for peptide SEQLLTASER (no. 8, Table 1), corresponding to the appearance of a y_1 fragment; however, the fragment b_1 was not observed, similar to the peptide DLYDPR (no. 5, Table 1), where fragment y_2 was the most intense signal in its MS/MS spectrum. Here, the proline effect[15] is added to the role of aspartic acid. Additionally, for peptide C*C*TESLVNR (*, propionamidated cystein residue), the b_1 fragment was not observed; and for other peptides,

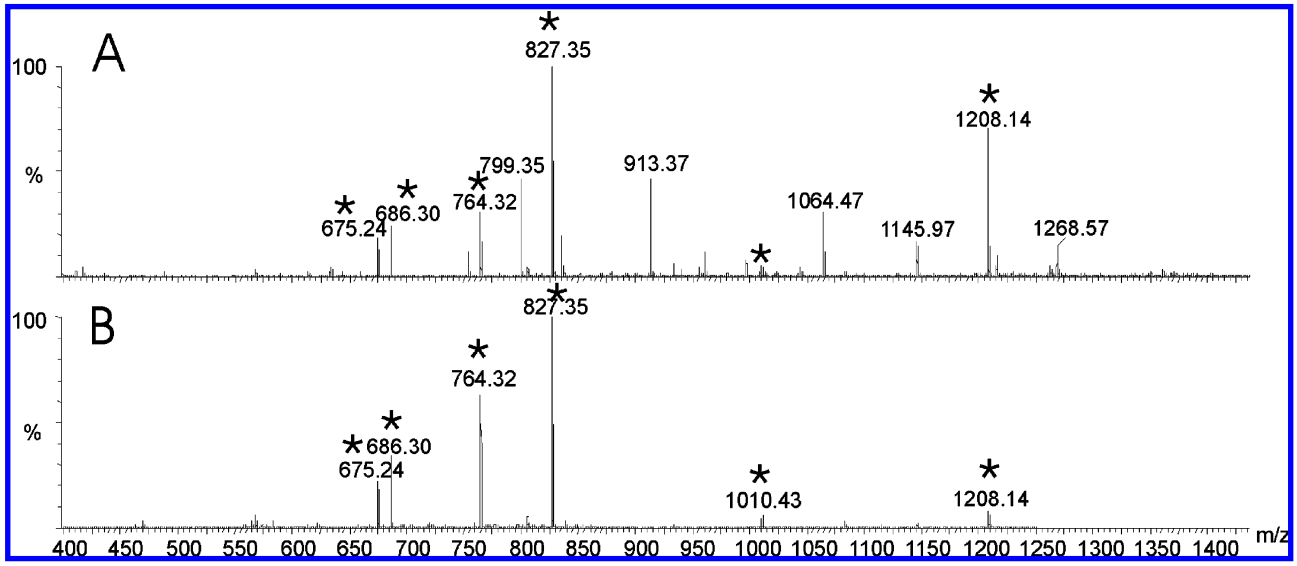

Figure 3. ESI-MS spectra of (A) PTC-peptides derived from the tryptic digestion of rSK. (B) Selective isolation of six multiply charged PTC-derivatized peptides using SCX chromatography.[10] The signals labeled with an asterisk represent the six multiply charged peptides (RH peptides).

the y_1 fragment and a peak at 273 Da were also detected, which cannot be explained from this limited set of experimental results.

For the second group, some peptides have an interesting characteristic (peptides 10, 11, 12, and 13 in Table 1), which is the presence of histidine and lysine in their sequences. Initially, the mobile proton should be localized in the histidine residue[13,16] because of its highly basic properties in the gas phase, but upon collisionally induced dissociation (CID) this proton is delocalized along the peptide backbone inducing additional fragmentations. In spite of this, the b_1 fragment is still favored and detected as important signal using lower collision energies than the one used in the first group.

The losses of PITC and the aniline group of the amino group of lysine are also detected with the collision energy used (less than 25 eV). In a previous work, Gaskell reported that the PITC group in the ε-amino group of lysine is more labile than in the α-amino group of the peptide.[7] Thus, the y_{n-1} fragment and the losses $y_{n-1} - 93$ and $y_{n-1} - 135$ were also observed, in addition to the b_1 fragment. On the contrary, in the peptide DTHK (no. 10, Table 1) in spite of having a mobile proton at the histidine residue, the b_1 fragment was not observed. The MS/MS spectra of the other peptides in the second group have a similar pattern among them with the b_1 fragments and their complementary y_{n-1} fragments, with a MEC lesser than 25 eV.

Selective Isolation of Multiply Charged Peptides Blocked with PITC by Using Strong Cation Exchange Chromatography. In our group, quantitative blocking of primary amino groups of tryptic peptides and strong cation exchange chromatography (SCX) have been used to isolate with high selectivity the multiply charged peptides (named as RH peptides).[10] This procedure simplifies the complexity of peptides mixture considerably by isolating 3−5 RH peptides/protein. The RH peptides are well-represented among proteins of different proteomes and their analysis guaranteed high proteome coverage.[10] *In silico* analysis of protein sequences of several proteomes showed that more than 90% of the RH peptides have at least one histidine residues and more than 80% do not contain lysine in their sequences (data not shown).

In RH peptides, the sum of arginine and histidine residues in each sequence is greater than 1 (no. of R + no. of H > 1) and most of them produce multiply charged ions ($z \geq 2$) and contain mobile protons since they contain at least one histidine (see the column (R + H) in Table 1).

Considering that (1) RH peptides are isolated as N-terminal blocked species, (2) they are a family of peptides that generally contain a mobile proton upon CID, and (3) once they are derivatized with PITC their MS/MS spectra show an abundant b_1 ion, we decided to combine the PITC derivatization of tryptic peptides and separation by SCX to selectively isolate the RH peptides and evaluate whether the combined information of accurate molecular mass of peptides and the N-terminal amino acid extracted from the b_1 ion would be sufficient to perform reliable identification of proteins in a sequence database.

The ESI-MS spectrum of the tryptic peptides of rSK modified with PITC is shown in Figure 3A, and the signal assignment is summarized in Table 1. The six RH peptides of rSK present in this mixture (marked with asterisks in Figure 3A) were selectively retained into the SCX column, eluted, and analyzed successfully by ESI-MS (Figure 3B). After SCX chromatography, none of the non-RH peptides were observed in the mass spectrometric analysis of the retained fraction (see Figure 3B), indicating the high specificity of the proposed method for isolating only the RH peptides.[10]

Identification of rSK in a Protein Sequences Database Using the Information of Molecular Mass Accuracy and b_1 Ion. *Manual Identification.* All the ESI-MS/MS spectra of the RH

(13) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. *J. Am. Soc. Mass Spectrom.* **1996**, *118*, 8365–8374.
(14) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* **1999**, *121*, 5142–54.
(15) Breci, L. A.; Tabb, D. L.; Yates, J. R., III; Wysocki, V. H. *Anal. Chem.* **2003**, *75*, 1963–1971.
(16) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77*, 5800–13.

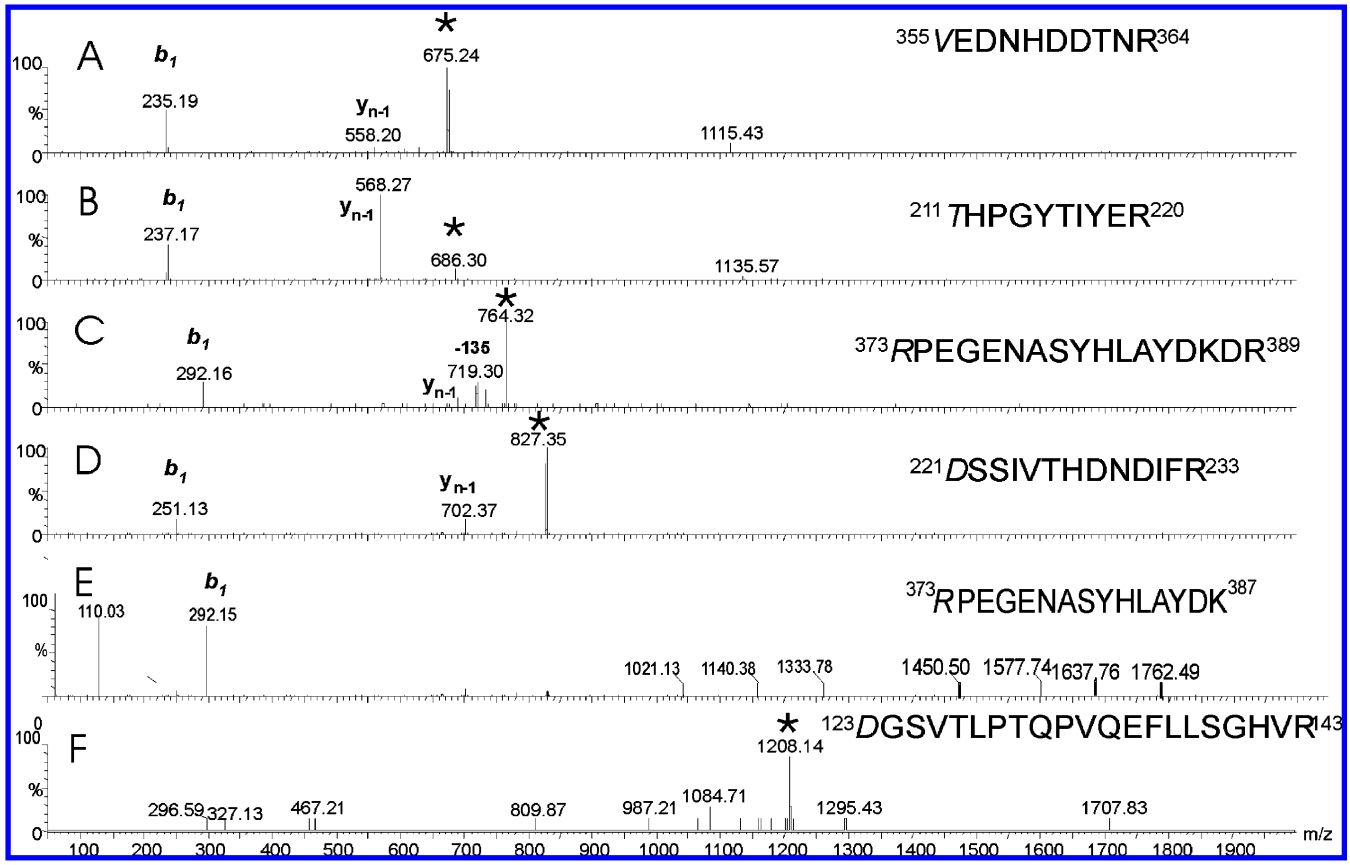

Figure 4. ESI-MS/MS spectra of the six multiply charged PTC-derivatized peptides after automatic precursor selection: (A) VEDNHDDTNR, 2+; (B) THPGYTIYER, 2+; (C) RPEGENASYHLAYDKDR, 3+; (D) DSSIVTHDNDIFR, 2+; (E) RPEGENASYHLAYDK, 2+; and (F) DGSVTLPTQPVQEFLLSGHVR, 2+. The b_1 fragment was not observed for the peptide shown in part F. The signals labeled with asterisks represent the precursor ions of selected PTC-derivatized RH peptides.

peptides, except one (see Figure 4F), showed the b_1 fragment. Each spectrum was completed after two scans of 2 s each, during the automatic selection of the precursor ions.

The mass values considering 30 ppm of mass accuracy and the N-terminal amino acids for all the five peptides were supplied through the sequence query option to the MASCOT program for protein database identification. Four out of five peptides were automatically identified and the protein correctly matched with the expected one. These four MS/MS spectra showed b_1 and y_{n-1} fragment ions. The precursor ion m/z 1208.14 only showed the fragment b_1 by manual selection and longer acquisition time (Table 1, peptide no. 23).

Table 2 shows the results of the mass peptide matching against the firmicute database containing proteins derived from all gram positive bacteria, (53 028 sequences). It is important to notice that considering the characteristics of the selected peptides (RH peptides, where no. of R + no. of H > 1), it was possible to considerably reduce the number of peptides that matched with a given molecular mass and the N-terminal amino acid information considering one missed cleavage site for trypsin. As mentioned before, four out of five peptide masses were identified as unique peptides in the whole database and assigned to the rSK protein using a very modest mass accuracy (30 ppm), if we take into consideration the state-of-the-art accurate mass measurements in instruments like FTICR-MS and Orbitrap.

Only the peptide of m/z = 686.30 (THPGYTIYER) is not unique for 30 ppm accuracy matching with 2 and 3 peptides that

Table 2. Number of SWISS-PROT (Firmicute) Database Matches for Candidates Peptides, Based on the Molecular Mass and the N-Terminal Information of the Streptokinase RH Peptides Modified with PITC

	N-term AA[a]	no. of peptides[b]	no. of RH peptides[c]	no. of peptide (1 MCS)[d]	no. of RH peptide (1 MCS)[e]
827.35	D	4	1	9	1
675.24	V	1	1	2	1
686.30	T	5	2	8	3
764.32	R	0	0	1	1
1010.43	R	1	1	1	1

[a] The N-terminal amino acid obtained in the analysis of the MS/MS spectrum. [b] Number of candidate peptides obtained in the normal database after matching. [c] Number of candidate peptides obtained in the RH database after matching. [d] Number of candidate peptides obtained in the normal database after matching, considering 1 missed cleavage site (MCS). [e] Number of candidate peptides obtained in the RH peptides database after matching considering 1 missed cleavage site (MCS).

are also classified as RH peptides. This kind of peptide could be further filtered by improving the mass accuracy or considering additional characteristics of peptides such as their p*I* or retention time.[17,18] A more detailed *in silico* analysis was made with the inclusion of other methods for selective isolation of peptides and other properties (manuscript submitted).

Automatic Identification. Manual identification of peptides using the MASCOT software was effective but obviously a procedure developed for high-throughput proteomics could be as automatic as possible with almost no intervention of users. Therefore, we

developed a program that allowed the identification of peptides using the mgf file generated from the MS/MS spectra. Usually, during a standard proteomic experiment the peptides do not necessarily produce fragments at high yields due to multiple fragment pathways by CID activation. In this proposal, RH peptides mainly generate two main fragments (b_1 and y_{n-1}) with high yields in a sequence-independent manner.[9] The mgf file with 10 queries was submitted to the Nterminal in house developed software, and the result obtained is presented in the Supporting Information, Supplementary Material 1. The result was similar to those obtained by manual identification where five tryptic peptides of rSK were successfully identified.

In both cases, the list of precursor peptides and their fragment ions were matched with the peptides from a filtered database containing only the RH peptides. The procedure was as follows: (1) the mass accuracy was fixed at 30 ppm and the candidate peptides for each precursor submitted were grouped in an initial list, (2) the masses of the N-terminal amino acids derivatized with PITC (as b_1 fragment) were calculated for the peptides generated *in silico* from the protein database. These mass values were matched with the peak masses submitted in the mgf files. Finally, (3) a list of candidates by each precursor ion was obtained, ordered by a simple score based on the intensity of the possible b_1 signal (explained in the Materials and Methods).

The efficiency of the assignments was around 50%; this means that a confident identification was possible (scores equal to 1 are considered as confident candidates) of the N-terminal amino acid for five peptides in the experiment carried out (Supporting Information, Supplementary Material 1); however, five out of six RH peptides were efficiently fragmented, which represent 83%. Obviously, other queries were processed, corresponding to low-intensity signals detected in the MS spectrum and some cases originated from the wrong selection of the precursor mass or its charge states during the automatic selection in the mass spectrometer. In these cases, none of them were assigned to peptides with confident scores in the database.

Application to the Analysis of Protein Mixture. Although the use of PTC-derivatized peptides to identify the first amino acid have been reported previously,[6] they have never been used in LC−MS/MS experiments for the analysis of real proteins samples. This is a challenge because it is necessary to combine the optimal conditions to achieve the fragmentation of peptides in a high-throughput experiment with some informatics tool that responds to the protein database identification. As a proof of concept, the procedure was applied to the analysis of cytosolic proteins from *E. coli*. Two LC−MS/MS experiments were accomplished for the analysis of PTC-derivatized peptides (1) without any selective isolation step of multiply charged peptides and (2) after the selective isolation of the multiply charged PTC-derivatized peptides. As previous MS/MS experiments performed in this work, the collision energies used were set according to the obtained MCE for peptides presented in Table 1 (collision energy ramps used are shown in the Supporting Information, Supplementary Material 2). A total of 116 unique peptides from 106 proteins at 20 ppm, which represents 39% of the total signals assigned to peptide sequences with scores equal to 1 (see the Supporting Information, Sup. 3), were identified after the analysis of PTC-derivatized peptides (see Table 3).

Table 3. Results Obtained after the Analysis with Three Different Experiments of a *E. coli* Protein Extract

experiment	unique peptides	% unique peptides	efficiency (%)	proteins identified
nonderivatized peptides			24	69
PTC-peptides	116	39	44	107
RH-PTC peptides	151	63	37	136

As expected, in the other experiment, the number of multiply charged PTC-derivatized peptides identified as unique by mass and N-terminal amino acid increased. In total, 151 peptides from 136 proteins were correctly identified, based on scores equals to 1, which represented the 63% of total signals assigned to possible sequences (Supporting Information, Sup. 3). The number of identified proteins is relatively low (136 out of 4386 for the *E. coli* proteome), but this is twice the proteins found using nonderivitized tryptic peptides.

This result was without proper fractionation at protein or peptide level. In the present work, the pool of RH peptides, with no additional fractionation, was analyzed in a single LC−MS/MS run. Obviously, an extensive fractionation, as described by other authors[19] as well as the use of more sensitive mass spectrometers will improve the number of identified proteins.[20]

As we mentioned above, the efficiency or percentage of MS/MS assigned confidently to peptide sequences is around 10−20% in standard applications. Using similar procedures (digestion and LC−MS/MS settings), we analyzed a nonderivatized tryptic digestion of *E. coli* in a single LC−MS/MS run to evaluate the efficiency of assignments with our analytical system. In this experiment were identified 180 peptides with an efficiency of 24% (755 queries) from 69 proteins, considering a FDR less that 2%. The fragmentation efficiency of PTC-derivatized peptides was greater than 35% (scores = 1), considering the total queries in both experiments, and slightly superior for the total PTC-derivatized peptides (Table 3).

Although the collision ramps used were set in accordance to the fragmentation patterns of model peptides, it could be improved, as well as the number and duration of scans chosen in the LC−MS/MS runs. During LC−MS/MS experiments, the accuracy in the molecular mass determination could vary and some signals may be excluded from the analysis because of the fixed mass window set for the difference between theoretical and experimental values. The use of some internal standard or the lock-spray system[21] could help to overcome this problem.

On the other hand, the reduced m/z range (from m(PTC-Gly) to m(PTC-Trp); m/z 192−321) necessary to obtain the information of b_1 ions is very narrow in comparison with the full MS/MS and it will be the same range for all peptides independent of their molecular mass. The size of multiple LC−MS/MS runs

(17) Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. *Proteomics* **2005**, *5*, 24–34.
(18) Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. *Mol. Cell. Proteomics* **2004**, *3*, 908–919.
(19) Corbin, R. W.; Paliy, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E., Jr.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S.; Soupene, E.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9232–9237.
(20) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212–217.
(21) Satomi, Y.; Kudo, Y.; Sasaki, K.; Hase, T.; Takao, T. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 540–546.

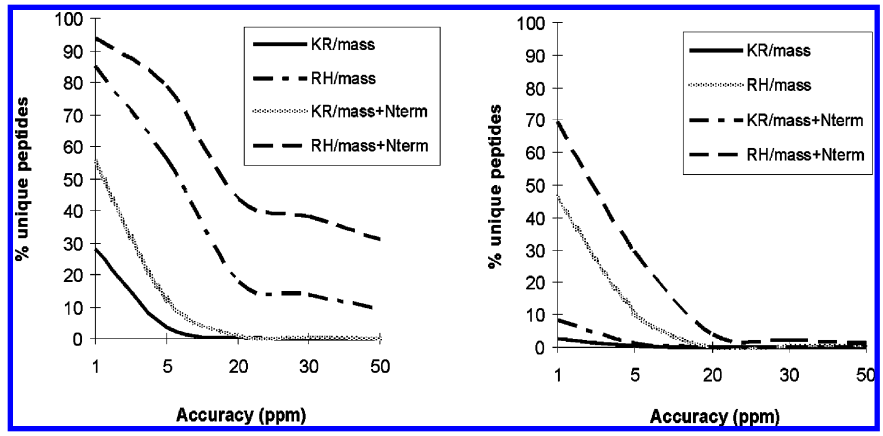

Figure 5. *In silico* analysis of unique peptides using the combination of several mass accuracies and the N-terminal information of all tryptic peptides (KR peptides) and the selectively isolated PTC-derivatized RH peptides: (left) *E. coli* proteome and (right) *H. sapiens* proteome.

will be considerably reduced and therefore more conveniently stored. Since the scan range can be reduced just to determine the molecular mass of b_1 ions, then using the same scan speed, an inherent improvement in sensitivity could be obtained.

Although the usage of ion traps in proteome analysis have gained in popularity because of their robustness, high-scan speed, sensitivity, versatility, accurate gain control in hybrid instruments, etc., some ion traps still have the limitation of a $^1/_3$ cutoff rule which in principle does not permit the detection of ions in the low-mass region where some b_1 ions are detected. Fortunately, more recent advantages of ion trap technology avoid the loss of this valuable information such as the iTRAQ reporter ions (m/z 114−119 and 121−122), and in principle it might occur similarly for b_1 ions.[22,23]

***In Silico* Analysis of Unique Peptides in *E. coli* and *Homo sapiens* Proteomes Based on the Selective Isolation of RH Peptides.** The identification of rSK is based on the information provided by the MS/MS spectra of RH peptides that were selectively isolated. These peptides at the same time were unique peptides in the sequence databases composed only by this type of peptide taking into account only the combination of two properties: molecular mass and N-terminal amino acid.

A previous *in silico* analysis revealed that selective isolation of RH peptides considerably simplifies the complex mixture of tryptic peptides by isolating just three RH peptides/protein, and at the same time they represent as average the 84% of the whole proteomes.[10] These data suggest that a considerable part of the proteins in the sequence database can be identified based on their MS/MS spectra of RH peptides because they are well represented and distributed.

The selective isolation of a particular type of peptides from the whole database might have two opposed effects. On one side, it might eliminate unique peptides present in the whole database that do not respond to the characteristics of the peptides selectively isolated (RH peptides), but on the other side, it might increase the number of unique peptides since this process could eliminate some peptides with properties very close or identical to a given RH peptide that could be selectively isolated.

Considering that RH peptides are very efficient to yield the desired Edman cleavage in the MS/MS spectrum (b_1 fragment ion) using very low-energy collisions and also taking into account that there are a wide variety of mass spectrometers available on the market with different performances, we calculate *in silico* the percentage of unique peptides in the whole database composed by all tryptic peptides of *E. coli* and *H. sapiens* and a subdatabase composed only by the RH peptides considering different mass accuracies and the information of the N-terminal amino acids. All these results were considering one missed cleavage of trypsin.

A detailed analysis is shown in Figure 5, where the percentage of unique peptides for several mass accuracy levels is shown. For the *E. coli* proteome, around 30% of the tryptic peptides are unique for low parts per million accuracy values (1 ppm), while for 5 ppm accuracy, these percentages decrease considerably. Thus for 20, 30, and 50 ppm, the results are very poor.

Considering also the N-terminal amino acid, additionally to the accurate molecular mass determination, the number of unique peptides substantially increases (Figure 5A). If RH peptides are selected, then it would be possible to obtain values close to 90% of unique peptides with 1 and 5 ppm of mass accuracy. In this case, even using very modest mass accuracies of 30 and 50 ppm, the results obtained become attractive for identification of proteins, unlike the analysis of general tryptic peptides analysis.

This behavior is similar to the analysis of the human proteome (Figure 5B); however, the percentage of unique peptides for low-mass accuracies (20, 30, and 50 ppm) is very low. Therefore it is recommended to use mass spectrometers such as the FTICR and Orbitrap to achieve highly accurate mass determination of the precursor ions (1−5 ppm range).

However, the methods for the selective isolation of peptides are not 100% effective, and a few percentage of unspecific peptides are detected. In a previous application of selective isolation of multiply charged peptides, we reported less than 5% of non-RH peptides in the analyzed fraction, specifically the contamination with peptides where the sum of arginine and histidine are equal to 1 (no. of R + no. of H = 1).[10] In the proposed procedure, the number of non-RH peptides analyzed by mass spectrometry in the fraction of interest could be reported as positive hits against the filtered RH database, in other words, false positive hits.

(22) Schwartz, J. C. High-Q pulsed fragmentation in ion traps. U.S. Patent 6,949,743, September 27, 2005.
(23) Schwartz, J. C.; Syka, J. E. P.; Quarmby, S. T. The 53rd ASMS Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, June 5−9, 2005.

Table 4. *In Silico* Analysis of False Positive Hits Considering 5% of non-Rh Peptides in the Selective Isolation of Multiply Charged Peptides

accuracy (ppm)	% false positive hits (R1, R2, R3)	% total false positive hits	average	CV (%)
1	(22, 21, 21)	(1.08, 1.05, 1.07)	1.06	1.2
20	(32, 30, 31)	(1.58, 1.49, 1.54)	1.54	3.3

In this way, we designed an *in silico* experiment to evaluate the possibility to obtain false positive hits with non-RH sequences, specifically those where no. of R + no. of H = 1. We evaluated 2728 non-RH peptides that represent the 5% of the total number of RH peptides contained in the *E. coli* database. Thus, we simulated the maximum number of non-RH peptides for a real experiment. The sequences were chosen randomly three times (replicates, R1, R2, R3) in an *E. coli* peptide database composed of peptides where the sum of arginine and histidine are equal to 1 (no. of R + no. of H = 1). These sequences were compared, based on its molecular mass and N-terminal amino acid information, with the filtered RH database at 1 and 20 ppm of mass accuracy. The positively matched were considered as potential false positives at the mass accuracy determined. The percentage obtained was around 20−30%, which represents percentages between 1 and 1.5 if we consider that the initial number of peptides represents 5% of the total peptides (Table 4).

The values were similar among the replicates (R1, R2, and R3) with CV less than 5%. These results suggest that due to the presence of non-RH peptides, the number of false positive hits, during the identification in the filtered RH database, may be less than 2% using 1 and 20 ppm mass accuracy.

CONCLUSIONS

All these results allowed us to conclude that PTC-derivatized peptides with fixed protons yield fragments different from the expected b_1 and y_{n-1} ions in the ESI-MS/MS spectra obtained in an hybrid type QTOF mass spectrometer because higher collision energy is require to fragment these peptides. On the contrary, peptides with mobile protons produce almost exclusively the desired fragments (b_1 and y_{n-1} ions) that provide information on the N-terminal amino acid. We also demonstrated that most of the multiply charged peptides derivatized with PITC efficiently produce the b_1 and y_{n-1} fragments.

Multiply charged PTC-derivatized peptides can also be selectively isolated from complex mixtures of tryptic peptides, combining the derivatization of all primary amino group and strong cation exchange chromatography. Accurate molecular mass measurement and the N-terminal amino acid of RH peptides is sufficient for a reliable protein identification in high-throughput proteomics, and it might have a positive impact to increase the number of protein identifications in filtered sequence databases containing only RH peptides.

ACKNOWLEDGMENT

A.S. and Y.P.-R. have contributed equally to this work. The authors would like to thank the INSPUR Company from China for its kind donation of the computer cluster TS10000 used for all calculations and the tools developed in this manuscript.

SUPPORTING INFORMATION AVAILABLE

Outputs of the automatic identification of peptides from rSK using the mgf files obtained during the selection of automatic precursors ions; collision energies ramps used for the fragmentation of PTC-derivatized peptides; and output files obtained from the analysis of PTC-derivatized peptides derived from *E. coli* proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review May 14, 2010. Accepted September 3, 2010.

AC1012738