

High Throughput Proteome Screening for Biomarker Detection*

Sheng Pan‡, Hui Zhang‡, John Rush§, Jimmy Eng‡¶, Ning Zhang‡, Dale Patterson||, Michael J. Comb§, and Ruedi Aebersold‡**‡‡

Mass spectrometry-based quantitative proteomics has become an important component of biological and clinical research. Current methods, while highly developed and powerful, are falling short of their goal of routinely analyzing whole proteomes mainly because the wealth of proteomic information accumulated from prior studies is not used for the planning or interpretation of present experiments. The consequence of this situation is that in every proteomic experiment the proteome is rediscovered. In this report we describe an approach for quantitative proteomics that builds on the extensive prior knowledge of proteomes and a platform for the implementation of the method. The method is based on the selection and chemical synthesis of isotopically labeled reference peptides that uniquely identify a particular protein and the addition of a panel of such peptides to the sample mixture consisting of tryptic peptides from the proteome in question. The platform consists of a peptide separation module for the generation of ordered peptide arrays from the combined peptide sample on the sample plate of a MALDI mass spectrometer, a high throughput MALDI-TOF/TOF mass spectrometer, and a suite of software tools for the selective analysis of the targeted peptides and the interpretation of the results. Applying the method to the analysis of the human blood serum proteome we demonstrate the feasibility of using mass spectrometry-based proteomics as a high throughput screening technology for the detection and quantification of targeted proteins in a complex system. *Molecular & Cellular Proteomics* 4: 182–190, 2005.

The comprehensive, quantitative analysis of proteomes is informative and challenging. It is informative because the comparative analysis of proteomes or fractions thereof identifies proteins that are present at different quantities in the samples compared. Such differences in turn have been used

to identify cellular functions and pathways affected by perturbations and disease (1–6), have been used to identify new components and changes in the composition of protein complexes and organelles (7–12), and have led to the detection of putative disease biomarkers (13, 14). Comprehensive proteome analysis is challenging because of the enormous complexity of the proteome. In comparison to the number of open reading frames in a genome the number of unique protein species expressed by it is vastly expanded by the action of post-transcriptional processing mechanisms including protein modifications, alternative splicing, and proteolytic processing. Consequently, to date, neither the complexity of a proteome nor its actual composition has been determined for any species.

Over the last few years a number of mass spectrometry-based quantitative proteomic methods have been developed that identify the proteins contained in each sample and determine the relative abundance of each identified protein across samples (15–20) or the absolute abundance of specific proteins in a sample (21, 22). Generally the proteins in each sample are labeled to acquire an isotopic signature that identifies their sample of origin and provides the basis for accurate mass spectrometric quantification. Samples with different isotopic signatures are then combined and analyzed typically by multidimensional chromatography tandem mass spectrometry. The resulting CID spectra are then assigned to peptide sequences, and the relative abundance of each detected protein in each sample is calculated based on the relative signal intensities for the differentially isotopically labeled peptides of identical sequence. Therefore, in a single operation the identity of the proteins contained in the samples and their relative abundance are determined. While the methods differ in the way the stable isotopes are incorporated into the polypeptides and the precise analytical (separation, mass spectrometry, and data processing) methods used (15), they have in common that in every experiment results are only obtained from those peptides for which in the tandem mass spectrometry (MS/MS)¹ experiment precursor ions are selected, successfully fragmented, and conclusively assigned to a peptide sequence. Therefore, in every proteomic experiment of this kind the proteome is rediscovered without taking advantage of the data collected from prior experiments. Furthermore it has become apparent that this type of proteomic

From the ‡Institute for Systems Biology, Seattle, Washington 98103, §Cell Signaling Technology, Inc., Beverly, Massachusetts 01915, ¶Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, ||Applied Biosystems, Framingham, Massachusetts 01701, and **Institute of Biotechnology, ETH Zurich and Faculty of Natural Sciences, University of Zurich, CH-8093 Zurich, Switzerland
Received, October 14, 2004, and in revised form, December 28, 2004

¹Published, MCP Papers in Press, January 5, 2005, DOI 10.1074/mcp.M400161-MCP200

¹ The abbreviations used are: MS/MS, tandem MS; F, fasted; S, saturated.

analysis is quite inefficient in that the number of successfully identified and quantified peptides is about an order of magnitude lower than the number of detectable peptides present in the sample (23) and that it is biased toward the proteins of higher abundance.

In many studies it is necessary to analyze a large number of proteomes and to compare the obtained results. In biomarker discovery studies for example, large numbers of samples are required to detect protein patterns that consistently associate with a specific condition within a large background of proteins that may randomly fluctuate within the population tested (24–26). In the emerging field of systems biology a key element is the quantitatively accurate and comprehensive measurement of the components that constitute the system in differentially perturbed states and the synthesis of these data into a model describing the system (27). Therefore, it is essential that quantitative proteomic experiments can be carried out at high throughput.

Recently we have argued that genomics-style biology can be separated into two distinct phases: a discovery phase in which all the possible elements of one type are discovered and a browsing or screening phase in which the list of all possible or known elements is searched for those that may be of interest in a particular study (28). The transition from a discovery to a browsing mode of operation has already been implemented for genomic sequencing, gene expression array analysis, and the analysis of single nucleotide polymorphisms. In this work we describe a method and its implementation in a platform to also transform quantitative proteomics from a discovery into a browsing mode of operation. We demonstrate the performance of the system by analyzing proteins contained in human blood serum. Based on the characteristics of the method, which includes vastly simplified data analysis, high throughput, absolute quantification of proteins in complex samples, reduced redundancy, the ability to search for and quantify specific proteins, and the potential for standardization of results between laboratories, the method is expected to become widely applicable in quantitative proteomic studies.

EXPERIMENTAL PROCEDURES

Preparation of Formerly *N*-Glycosylated Peptides from Serum—The procedure for the selective isolation of *N*-glycosylated peptides from serum was described previously (29). Proteins from 50 μ l of serum were exchanged into coupling buffer (100 mM NaAc and 150 mM NaCl, pH 5.5) using a desalting column (Bio-Rad) and oxidized by adding 15 mM sodium periodate at room temperature for 1 h. After removal of the oxidant using a desalting column, the sample was conjugated to hydrazide resin (Bio-Rad) at room temperature for 10–24 h. Non-glycosylated proteins were then removed by washing the resin six times with 1 μ l of urea solution (8 M urea, 0.4 M NH_4HCO_3 , pH 8.3). After the last wash and removal of the urea solution, the resin was resuspended in 4 \times diluted urea buffer (2 M urea, 0.1 M NH_4HCO_3 , pH 8.3). Trypsin was added at a concentration of 1 mg of trypsin/200 mg of serum protein and digested at 37 °C overnight. The peptides were reduced by adding 8 mM Tris(2-carboxyethyl)phosphine (Pierce) at room temperature for 30 min and alkylated by adding 10 mM iodoacetamide at room temperature for 30 min. The trypsin-released

peptides were removed, and the resin was washed three times with 1.5 M NaCl, 80% acetonitrile, 0.1% TFA, 100% methanol and six times with 0.1 M NH_4HCO_3 . *N*-Linked glycopeptides were released from the resin by addition of 0.5 μ l of peptide-*N*-glycosidase F (New England Biolabs, Beverly, MA) and incubation at 37 °C overnight. The released peptides were dried and resuspended in 25 μ l of 0.4% acetic acid solution for mass spectrometry analysis.

Synthesis of Stable Isotope-labeled Peptides—Fmoc (*N*-(9-fluorenyl)methoxycarbonyl)-derivatized stable isotope monomers containing one ^{15}N and five to nine ^{13}C atoms were from Cambridge Isotope Laboratories (Andover, MA). The precise sequences to be synthesized were selected from prior data generated by the analysis of peptides isolated from serum samples by ESI-MS/MS. Preloaded Wang resins were from Applied Biosystems (Foster City, CA). The synthesis scale was 5 μ mol. Amino acids activated *in situ* with 1-*H*-benzotriazolium,1-[bis(dimethylamino)methylene]-hexafluorophosphate(1-),3-oxide:1-hydroxybenzotriazole hydrate were coupled at a 5-fold molar excess over peptide. Each coupling cycle was followed by capping with acetic anhydride to avoid accumulation of one-residue deletion peptide byproducts. After synthesis, peptide resins were treated with a standard scavenger-containing trifluoroacetic acid-water cleavage solution, and the peptides were precipitated by addition to cold ether. Peptides were purified by reverse phase C_{18} HPLC using standard TFA/acetonitrile gradients and characterized by MALDI-TOF (Biflex III, Bruker Daltonics, Billerica, MA) and ion trap (LCQ DecaXP, ThermoFinnigan, San Jose, CA) MS. The purified synthetic peptide stocks were quantified by amino acid analysis using a PicoTag station (Waters, Milford, MA) for acid hydrolysis and an AccQ-Fluor reagent kit (Waters) for amino acid derivatization. The quantity of each reference peptide used per assay is indicated in Table I.

LC/Probot Fractionation and MALDI-TOF/TOF Analysis—6 μ l of the formerly *N*-glycosylated peptide mixture (corresponding to an isolate from 12 μ l of serum) was separated by reverse phase C_{18} column and spotted on a MALDI plate. The separation was performed using an Ultimate HPLC system (LC Packing/Dionex, Sunnyvale, CA) coupled with a Famos microautosampler (LC Packing/Dionex). A 100-min gradient with solvent B ramping from 5 to 40% in 70 min was used for peptide separation using an in-house packed C_{18} column (150- μ m inner diameter \times 12.5 cm). The solvents A and B were 0.1% TFA, HPLC grade water and 0.1% TFA, acetonitrile, respectively. The eluent from the capillary column was mixed with the α -cyano-4-hydroxycinnamic acid matrix solution (Agilent Technologies, Palo Alto, CA) in a 1:1 ratio in a mixing tee before spotting onto the MALDI plate. The fractions were automatically collected in 30-s intervals and spotted on a 192-well MALDI plate (Applied Biosystems) using a Probot microfraction collector (LC Packing/Dionex). The samples were analyzed by a MALDI-TOF/TOF tandem mass spectrometer (ABI 4700 Proteomics Analyzer, Applied Biosystems). Both MS and MS/MS data were acquired with a Nd:YAG (neodymium doped yttrium aluminum garnet) laser with 200-Hz sampling rate. For MS spectra, 1000 laser shots per spot were used to assure appropriate ion statistics for quantification. MS/MS mode was operated with 1-keV collision energy. The CID was performed using air as the collision gas. Typically 2000 laser shots were used for MS/MS acquisition. Both MS and MS/MS data were acquired using the instrument default calibration.

Data Base Searching of MS/MS Data—MS/MS data were searched against the human protein data base from International Protein Index (IPI) human protein data base version 2.28 from the European Bioinformatics Institute (EBI) and a standard peptide data base containing the spiked peptides. The mass tolerance of the precursor peptide was set at ± 0.4 Da, and the data base search was set to expect the stable isotope labeling and the following modifications: carboxymethylated cysteines, oxidized methionines, and an enzyme-catalyzed conver-

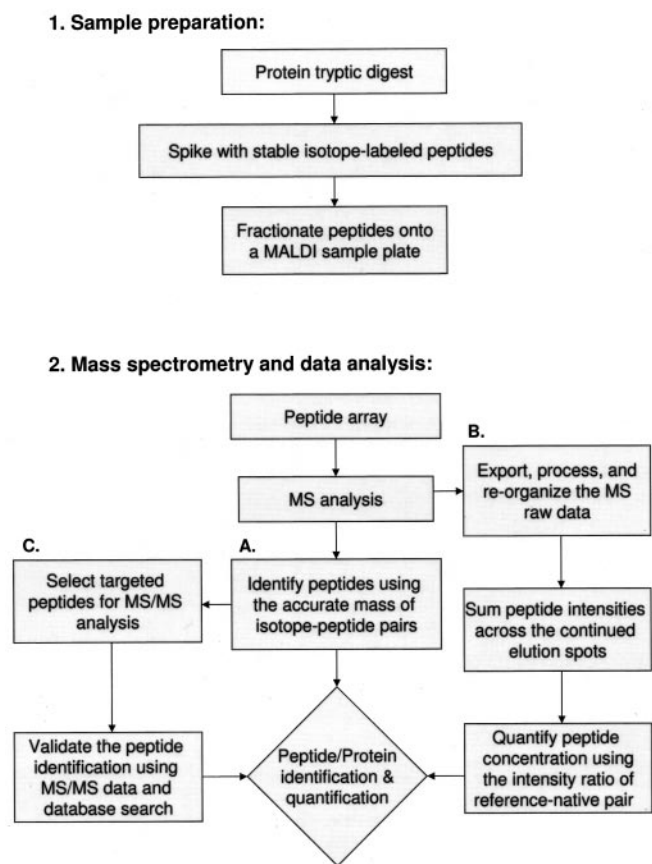


FIG. 1. A schematic illustration of an off-line LC-MALDI-TOF/TOF-based platform for proteome screening technology. Step A, high speed MS scanning; step B, peptide quantification; step C, optional confirmation of peptide identity by MS/MS.

sion of asparagine to aspartic acid at the site of carbohydrate attachment. No other constraints were included in the SEQUEST search.

Quantification—Binary files of MS survey scans were exported using 4700 Explorer software. Each file corresponded to a single MS spectrum. The peak information including spot number, mass, and intensity was extracted from the binary files and converted to text files. The individual files were then combined into a single text file that contained the peak information from all the spots. The file was scanned for peptides that had eluted across more than one sample spot. The signal intensities of these peptides from each adjacent spot were summed together to determine an accurate intensity over the entire peptide elution profile. The quantification of targeted peptides was achieved using the abundance ratio of a native peptide to the corresponding spiked stable isotope-labeled peptide for which the amount was known. The quantification of each identified peptide was manually checked to verify the validity of the results.

RESULTS

The method is schematically outlined in Fig. 1. It is conceptually simple and consists of two main steps, the production of peptide arrays and their interrogation by MALDI tandem mass spectrometer in MS and MS/MS mode. For the production of ordered peptide arrays, protein samples (untagged proteins or proteins labeled with specific stable isotope tags) are subjected to tryptic digestion and combined with a mix-

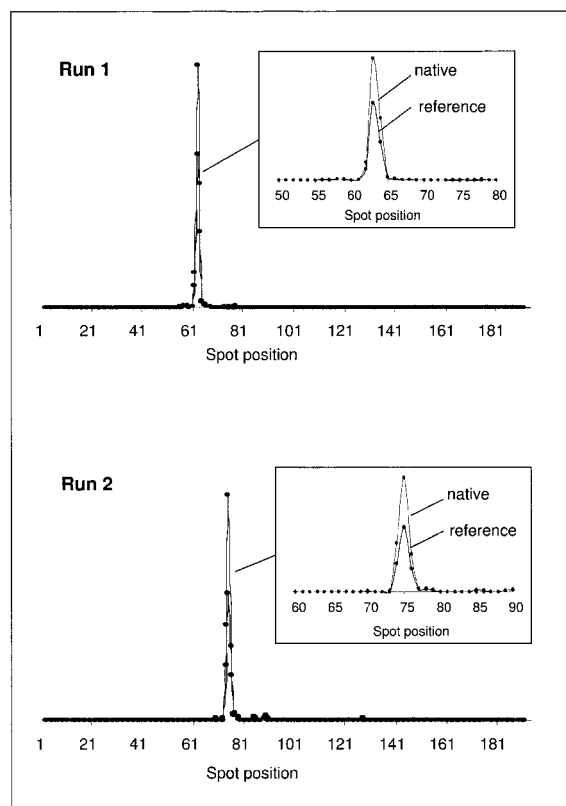


FIG. 2. Search and identification of a specific reference-native peptide pair in a complex background of serum-derived peptides. The native peptide was consistently identified in different runs using the stable isotope-labeled reference peptide as a search criterion even though the peptides were deposited on different spot positions in different runs.

ture of defined amounts of isotopically labeled reference peptides, each of which uniquely identifies a particular protein or protein isoform (proteotypic peptides). The reference peptides are generated by chemical synthesis and contain heavy stable isotopes. The decision that peptides should be synthesized is based on information obtained from prior experiments. The combined peptide mixture is separated by capillary reverse phase liquid chromatography, and the eluting peptides are deposited on a MALDI sample plate to form an ordered peptide array in which each array element contains peptides that are derived from the digested sample proteins and/or from the mixture of reference peptides. For the detection and quantification of the target polypeptides (*i.e.* those proteins for which a reference peptide was added to the sample) the sample is analyzed using a MALDI tandem mass spectrometer, carrying out the following sequential steps. In step A, high speed MS scanning, MALDI-MS spectra are acquired from each array element, generating two types of signals, one representing the signals of the peptides for which no reference peptide was added, appearing as single peaks, and the other representing the signals for those peptides for which a reference peptide was added, appearing as paired signals

TABLE I
The list of the stable isotope-labeled reference peptides shown in Fig. 4B

Peptide	Swiss-Prot/ TrEMBL accession no.	Protein annotation	Synthesized stable isotope-labeled peptide sequences ^a	Amount of reference peptide added in 6- μ l sample
				<i>pmol</i>
1	P08185	Corticosteroid-binding globulin precursor	AQLLQGLGFDLTER	9.4
2	P55058	Phospholipid transfer protein precursor	IYSDHSALESALIPLQAPLK	2.8
3	P10909	Clusterin precursor	LADLTQGEDQYYLR	4.8
4	P51884	Lumican precursor	LGSFEGLVDLTFIHLQHNR	0.7
5	P02750	Leucine-rich α -2-glycoprotein precursor	LPPGLLADFTLLR	8.0
6	P04004	Vitronectin precursor	DGSLFAFR	11.6
7	P04004	Vitronectin precursor	NDATVHEQVGGPSLTSDLQAQSK	3.5
8	Q13201	Endothelial cell multimerin precursor	FNPGAESVWLSDSLTK	5.0
9	P04278	Sex hormone-binding globulin precursor	LDVDQALDR	12.4
10	P04114	Apolipoprotein B-100 precursor	YDFDSSmLYSTAK	2.5
11	P80188	Neutrophil gelatinase-associated lipocalin precursor	SYDVTSLVFR	6.2
12	P54289	Dihydropyridine-sensitive L-type, calcium channel α -2/ δ subunits precursor	IDVNSWIEDFTK	3.8
13	P40225	Megakaryocyte-stimulating factor	DGTLVAFR	16.0
14	Q13876	Quiescin, bone-derived growth factor (fragment)	DGSGAVFPVAGADVQTLR	3.1
15	P40189	Interleukin-6 receptor β chain precursor	ETHLETDFTLK	5.8
16	P13473	Lysosomal-associated membrane protein 2 precursor, lysosome-associated membrane glycoprotein 2 precursor	WQMDFTVR	8.8
17	Q96CX1	Similar to RIKEN cDNA 2610528G05 gene (fragment)	LHEITDETETR	4.4
18	Q07954	Low density lipoprotein receptor-related protein 1 precursor	FDSTEYQWVTR	7.9
19	Q16853	Membrane copper-amine oxidase	IQmLSFAGEPLPQDSSmAR	2.4
20	P01033	Metalloproteinase inhibitor 1 precursor	FVGTPEVDQTTLYQR	5.4
21	Q92859	Neogenin precursor	TLSDVPSAAPQDLSLEVR	2.1

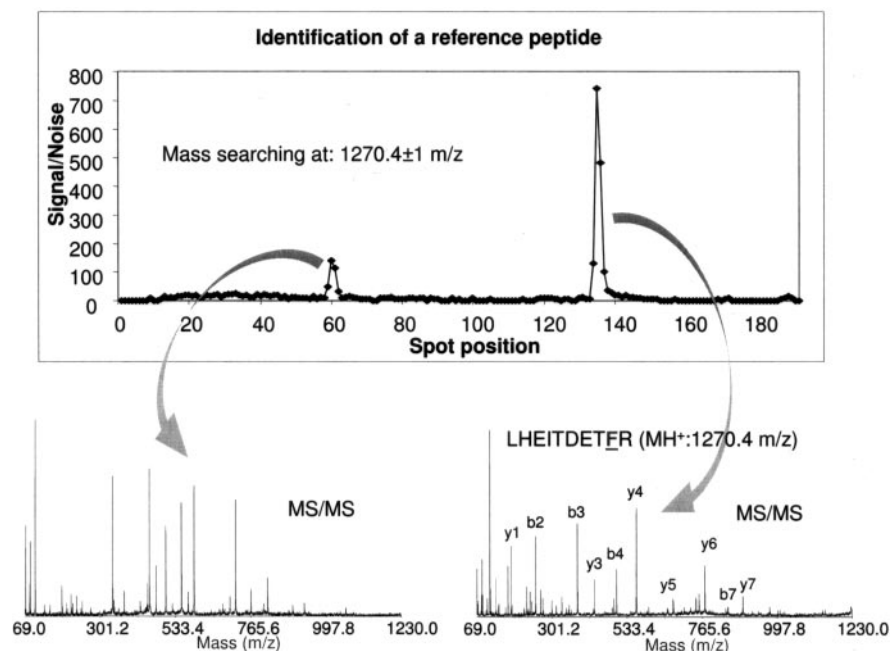
^a m, methionine oxidation; _, amino acid labeled with ¹⁵N and ¹³C.

with a mass difference that precisely corresponds to the mass difference encoded in the stable isotope tag. In step B, peptide quantification, the signal intensities of the isotopically heavy and light forms of a signal pair are determined and can be used to calculate the absolute abundance of the peptide derived from the protein sample. As reverse phase chromatography could split a specific pair of isotopic peptides across several consecutive spots on the MALDI plate, it is necessary to process the data prior to quantification. A specifically developed software tool scans the MS data files for peptides that eluted across more than one sample spot, sums the signal intensities of the corresponding signals from adjacent spots, and uses the integrated value for quantification, thus ensuring higher quantitative accuracy. In step C, optional confirmation of peptide identity by MS/MS, proteins are primarily identified by correlating the array position and the accurately measured mass of each isotopically labeled peptide pair in the array with a list of added reference peptides of known mass. Optionally peptide sequences could be confirmed by subjecting selected peptides to CID and searching the resulting spectra against a sequence data base (30) or a library of previously acquired MS/MS spectra representing the sequences of the reference peptides.

To test the robustness of peptide identification, reference

peptides were added to a complex mixture of formerly *N*-glycosylated tryptic peptides extracted from human serum (29) and spotted onto the sample plate under slightly different chromatography conditions. The plates were then analyzed, and the peptides were identified in the sample mixture based on their accurate masses, the paired nature of the signal, and the location on the peptide array. Fig. 2 shows the extracted ion traces over the chromatographic separation range for two consecutive runs. It is apparent that the stable isotope-labeled peptide LADLTQGEDQYYLR (mass, 1690.8 Da; stable isotope labeling on Leu (underlined); amount of added peptide indicated in Table I) and its corresponding native peptide were unambiguously identified in the complex background even though the targeted peptide pair was found in different spot positions in the two runs. The accurate mass together with the paired nature of the signal were sufficient for the identification of the target peptides within the complex sample mixture. With increasing complexity of the analyzed sample the chance that these criteria are insufficient for unambiguous peptide identification also increases. In these cases, peptide identities were confirmed by the fragment ion spectra of the precursors that are isobaric to the targeted peptide. An example of peptide confirmation by CID is illustrated in Fig. 3. Two peaks that corresponded to the mass of the stable iso-

FIG. 3. Complementarity of peptide identification using specific mass matching and peptide sequencing. The search of a specific mass (MH^+ , 1270.4 m/z for peptide LHEITDETFR; stable isotope labeling on Phe (underline)) resulted in more than one precursor ion locating at different spot positions. Both of the precursor ions were subjected to MS/MS analysis. The one with the higher intensity, distributing across spots 133–138, was identified as the targeted peptide.



tope labeled reference peptide LHEITDETFR (mass, 1269.4 Da; stable isotope labeling on Phe (underline)) were detected within the mass search window. The expected signal was discriminated from the unexpected one based on the CID spectrum. The SEQUEST search results of the obtained spectra indicated that the precursor ion with higher intensity, eluting across spot 133 to spot 138, was the target peptide. By limiting the number of sequencing operations using this approach, the platform not only provided for high confidence peptide identifications but also operated in a high throughput mode. For instance, with a laser sampling rate at 200 Hz available in the 4700 MALDI-TOF/TOF instrument, a 192-well sample plate could be analyzed in less than 1 h by MS scan of 192 spots followed by 200 MS/MS scans for selected peptide sequence validation.

To assess the performance of the system for rapid profiling of selected proteins in complex mixtures we analyzed *N*-glycoproteins in human serum. The serum-derived peptides were generated from serum proteins by using the solid-phase glycopeptide capture and release method as described under “Experimental Procedures.” The serum-derived sample was added with a mixture of isotopically labeled reference peptides. The combined sample was separated by capillary reverse phase chromatography and spotted onto the sample plate in 192 spots and analyzed by MALDI-TOF/TOF. As indicated in Fig. 4, the added reference peptides could be detected and identified over a broad range of the chromatographic separation range in a very complex sample. Fig. 4A shows the number of ions detected in each spot in MS mode, and Fig. 4B shows the distribution of the reference peptides detected in the sample over the chromatographic separation range. The distribution profile of the reference peptides de-

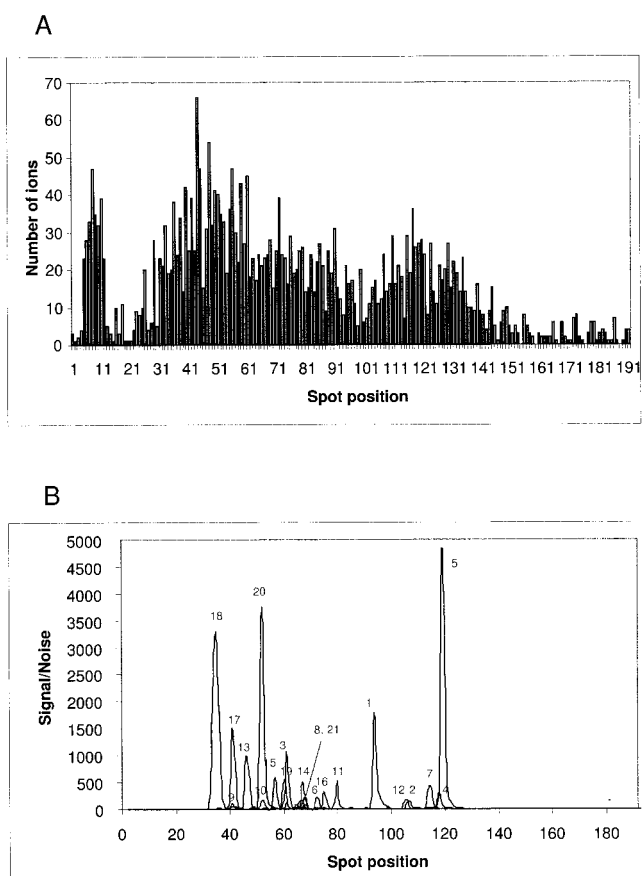


FIG. 4. A, the number of ions detected in each spot in MS mode. **B,** the elution profile of the stable isotope-labeled reference peptides added to the complex formerly *N*-glycosylated peptide mixture. The quantity of each reference peptide added is indicated in Table I.

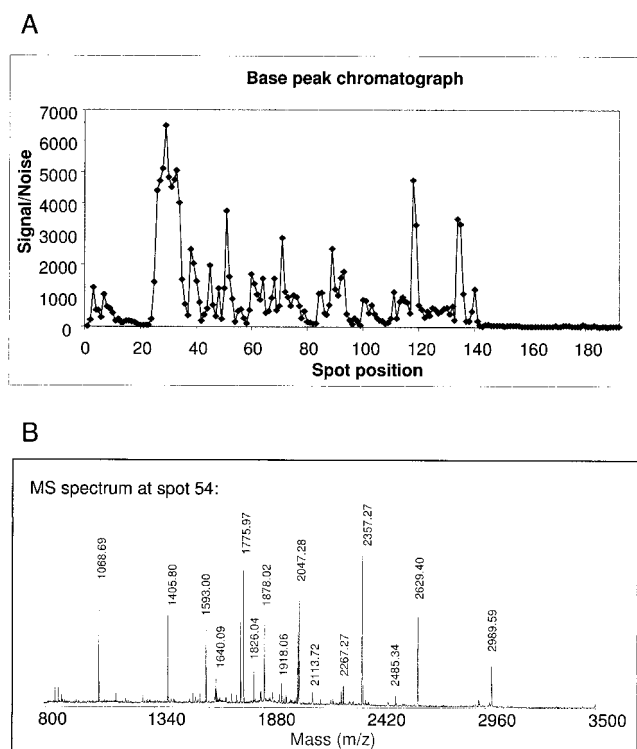


FIG. 5. A, the base peak chromatogram of a formerly *N*-glycosylated peptide mixture spiked with stable isotope-labeled peptides. The sample was fractionated in 192 wells on a MALDI plate. Each point on the *x* axis indicates a spot position. The elution of the majority of the peptides was between spots 25 and 140. B, the MS spectrum of a representative spot.

tected was extracted from the complex background. The sequence and quantity of the reference peptides discussed in Fig. 4B are listed in Table I. Fig. 5A shows the base peak chromatogram of the detected peptides, indicating that peptides were detected over the whole separation range with the majority of peptide signals concentrated between fractions 25 and 140. Fig. 5B shows the mass spectrum of a representative spot, indicating the complexity of the sample analyzed. In total more than 2500 unique ion signatures were detected in MS mode. To identify and quantify the target peptides we used the computer-driven selective peptide analysis method described above. Fig. 6 shows that the peptides could be identified and quantified even though they represented relatively minor peaks in a complex spectrum. Data for peptide NDATVHEQVGGPSLTSDLQAQSK, which was derived from vitronectin precursor and ^{13}C - and ^{15}N -labeled on residue leucine 18, are shown. Using the specific mass to search the MS data, the spot (or spots) containing the expected peptide pair was located. By examining the MS spectrum, the paired peaks (reference and native) were identified. The MH^+ of the reference peptide and native peptide were 2389.2 and 2382.2 Da, respectively. MS/MS analysis and sequence data base searching further confirmed the identification of the peptides. Since the amount of the reference peptide was known, the

concentration of the native peptide could be calculated based on the signal intensity ratio of the paired peptide signals. Consequently the identification and quantification of the related proteins in a complex serum sample was accomplished. The concentration of a protein in a serum sample can be calculated according to Equation 1,

$$C = \frac{(A_n/A_s) \times M_s}{V_a} \times (V_b/V) \quad (\text{Eq. 1})$$

where A_n and A_s are the integrated peak area of the native and reference peptide in the MS spectrum, respectively. M_s is the amount of stable isotope-labeled peptide spiked in the formerly *N*-glycosylated peptide mixture used for MALDI-TOF/TOF analysis. V_a is the volume of the formerly *N*-glycosylated peptide mixture used for MALDI-TOF/TOF analysis. V_b is the total volume of the formerly *N*-glycosylated peptide mixture extracted from the serum sample. V is the total volume of the serum used for the formerly *N*-glycosylated peptide extraction. It is important to note that the accuracy of the result estimated from the above formula depends on many factors including the purity of reference peptides, sample preparation process, formerly *N*-glycosylated peptide extraction efficiency, and data processing, etc.

To demonstrate the capacity of the system to rapidly and quantitatively profile selected serum proteins, formerly *N*-glycosylated peptide isolates from four human serum samples were analyzed. The samples were isolated from two individuals (indicated as 1 and 3, respectively) in a fasted (F) or saturated (S) state. The reference peptides were spiked into the samples, and the mixture was analyzed by off-line LC-MALDI-TOF/TOF platform. The proteins and the corresponding signature peptides for which both reference and native signals were detected by the platform are listed in Table II. The results are presented in the form of a peptide map in Fig. 7. The *x* axis represents the mass of the targeted native peptides, and the *y* axis indicates the abundance ratio of a native peptide to the corresponding isotope-labeled peptide, providing the quantitative information describing the corresponding protein. The peptides with masses 1542.7, 1559.8, 1662.9, 2278.3, and 2381.2 Da, respectively, did not show significant changes between the individuals and between the saturation states for the same individual. The data for the peptides with masses 1683.8, 1897.0, and 2195.2 Da, however, showed different patterns. For instance, for the peptide with mass of 2195.2 Da there was not a significant difference between the two individuals in the fasted state. However, in the saturated state the level of the peptide was increased significantly for individual 1, while only a minimal change was observed for individual 3. The result indicates that, even in very complex samples with enormous number of proteins that may fluctuate within a population, the key elements that indicate the state of a specific biological condition can be effectively extracted and expressed quantitatively by this approach.

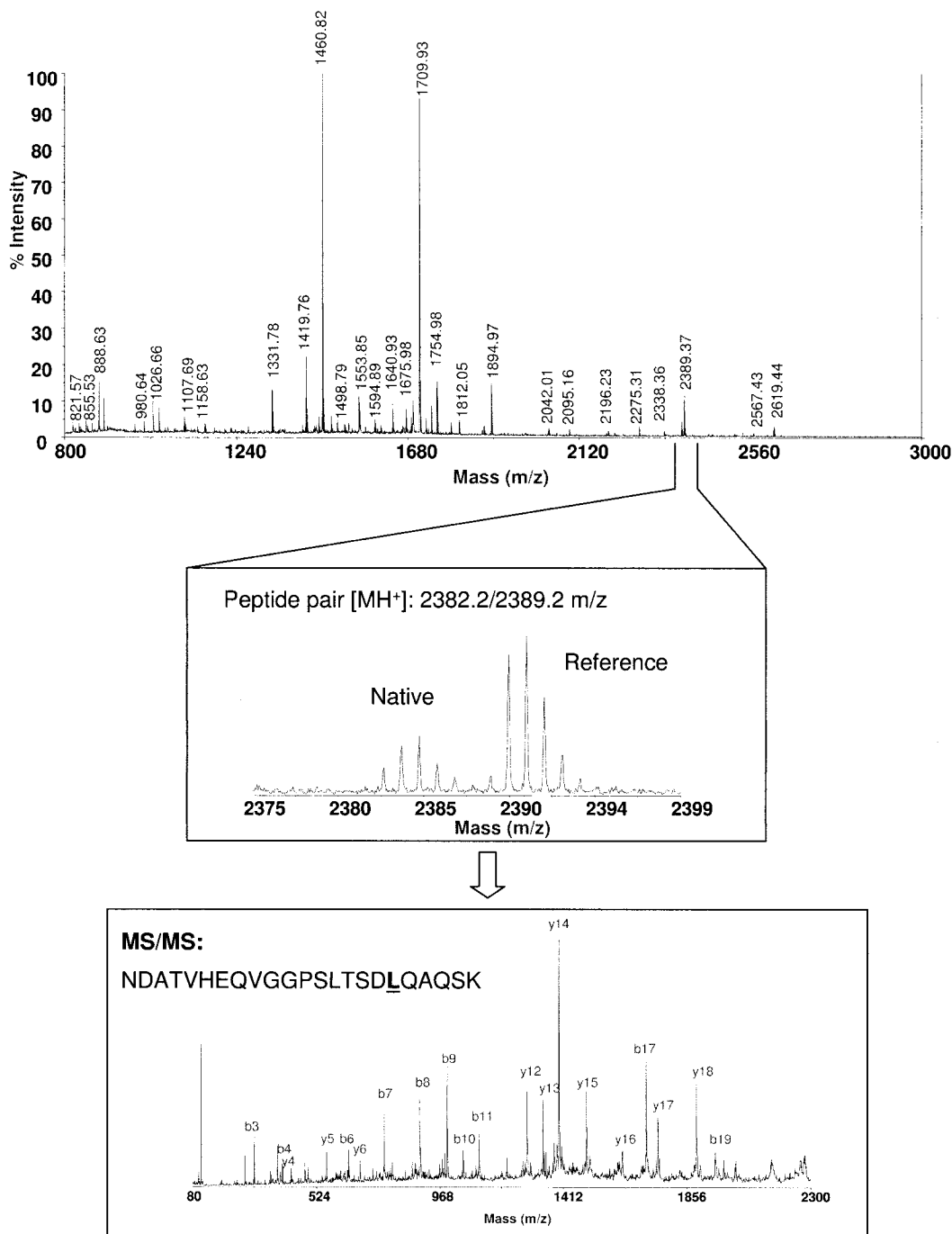


FIG. 6. Identification of a targeted peptide pair in a complex formerly *N*-glycosylated peptide mixture. The pair of the reference and native peaks was located and identified using MS data based on specific mass matching and the pair nature of the signal. The validation of the peptide sequence was accomplished using MS/MS analysis and data base searching.

DISCUSSION

In this study we describe a method for proteome screening and an experimental platform that supports the method. The method has the potential to reach very high throughput because the redundancy common to LC-MS/MS-based proteomic experiments is eliminated, and the analysis is focused on specific, information-rich analytes. It is an important ques-

tion how the candidate reference peptide sequences are identified in the first place. In the present study the peptides to be synthesized were selected from the data from prior ESI-LC-MS/MS experiments using formerly *N*-glycosylated peptides isolated from serum. We have also more generally addressed the question by generating a data analysis system and a data base that integrates proteomic data from different platforms,

TABLE II
The list of proteins and the corresponding signature peptides discussed in Fig. 7

Protein	Peptide sequence ^a	Mass (m/z)
Apolipoprotein B-100 precursor	YDFN*SSM#L [@] YSTAK	1542.7
Corticosteroid-binding globulin precursor	AQLLQGLGFN*L [@] TER	1559.8
Endothelial cell multimerin precursor	FNPGAESVLSN*STL [@] K	1662.9
Clusterin precursor	LAN*LTQGEDQYYL [@] R	1683.8
Neogenin precursor	TLSDVPSAAPQN*LSLEV [@] R	1897.0
Lumican precursor	LGSFGLVN*LTFIHL [@] QHNR	2195.2
Phospholipid transfer protein precursor	IYSN*HSALESALAIPLQAPL [@] K	2278.3
Vitronectin precursor	NN*ATVHEQVGGPSLTS [@] DL [@] QAQSK	2381.2

^a *, enzyme-catalyzed conversion of asparagine to aspartic acid at the site of carbohydrate attachment; #, methionine oxidation; @, amino acid that was labeled with ¹⁵N and ¹³C in the corresponding reference peptide.

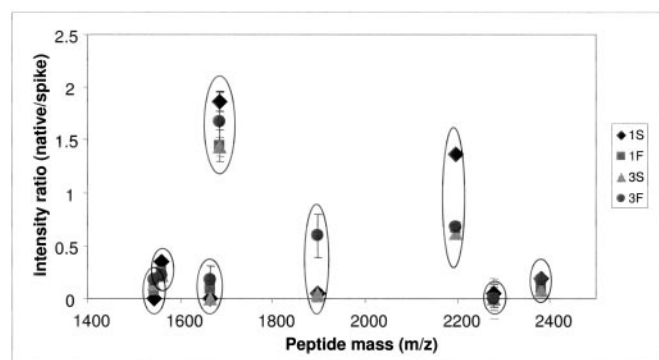


FIG. 7. Quantitative profile of the selected peptides detected in four different serum samples (1F, 1S, 3F, and 3S). The x axis represents the peptide mass. The y axis indicates the abundance ratio of a native peptide to the corresponding stable isotope-labeled reference peptide. The peptides and their corresponding proteins are listed in Table II.

laboratories, and experiments (31). This data base is a useful resource for the selection of sets of reference peptides. The off-line LC-MALDI-TOF/TOF-based platform provides several advantages for such an approach including high mass range and accuracy, selective MS/MS analysis based on MS information, and easy to interpret data structure. The generation of predominantly singly charged peptides by MALDI simplifies the quantitative analysis. Peptide identification can be performed on the same MALDI plate afterward by MS/MS if the information is needed. The ability to reexamine and verify the same sample set can be very beneficial for quantitative applications. In the present study the assignment of spectra to their corresponding peptide sequences was accomplished by sequence data base searching. Alternatively the spectra could also be searched against a library of spectra previously recorded for the reference peptides (library search). It is important to note that not all of the spike-in peptides behaved the same in a complex sample. In selection of reference peptides, criteria such as biological significance, sensitivity for mass analysis, good mass range, lack of potential mass overlap with other peptides, etc. need to be satisfied based on the type of mass spectrometer used. The development of proteome screening technology indicates an important tran-

sition of quantitative proteomics from a sole discovery mode into a multiphase technology. The implementation of the browsing/screening mode allows us to utilize the extensive genomic and proteomic knowledge that has been accumulated by biology and medicine and focus on analyzing the key elements that uniquely represent a specific biological condition. Technically, since the identification and quantification of targeted proteins is based on searching and identifying the corresponding signature peptide pairs directly, the approach significantly reduces sample complexity thereby improving throughput and identification confidence. It provides greater analytical dynamic range and facilitates the detection of low abundant proteins. The ability to describe specific protein patterns associated with certain biological conditions within a complex background in an absolute quantitative way provides the feasibility for data standardization. The proteome screening technology described in this report opens new opportunities for quantitative proteomic analysis and can potentially be developed into a high throughput technology for clinical diagnosis at the proteome level.

* This work was funded in part with federal funds from the NHLBI, National Institutes of Health, under Contract No. N01-HV-28179 and a research agreement from Applied Biosystems. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡‡ To whom correspondence should be addressed: Inst. of Biotechnology, Swiss Federal Inst. of Technology, ETH Hönggerberg HPT E 78, CH-8093 Zurich, Switzerland. Tel.: 41-1-633-31-70; E-mail: aebersold@biotech.biol.ethz.ch.

REFERENCES

1. Wright, M. E., Eng, J., Sherman, J., Hockenbery, D. M., Nelson, P. S., Galitski, T., and Aebersold, R. (2003) Identification of androgen-coregulated protein networks from the microsomes of human prostate cancer cells. *Genome Biol.* <http://genomebiology.com/2003/5/1/R4>
2. Guina, T., Purvine, S. O., Yi, E. C., Eng, J., Goodlett, D. R., Aebersold, R., and Miller, S. I. (2003) Quantitative proteomic analysis indicates increased synthesis of a quinolone by *Pseudomonas aeruginosa* isolates from cystic fibrosis airways. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2771-2776
3. Shio, Y., Donohoe, S., Yi, E. C., Goodlett, D. R., Aebersold, R., and Eisenman, R. N. (2002) Quantitative proteomic analysis of Myc oncoprotein function. *EMBO J.* **21**, 5088-5096

4. Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O., Bergamini, G., Coughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., Hopf, C., Huhse, B., Mangano, R., Michon, A. M., Schirle, M., Schlegl, J., Schwab, M., Stein, M. A., Bauer, A., Casari, G., Drewes, G., Gavin, A. C., Jackson, D. B., Joberty, G., Neubauer, G., Rick, J., Kuster, B., and Superti-Furga, G. (2004) A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nat. Cell Biol.* **6**, 97–105
5. Everley, P. A., Krijgsveld, J., Zetter, B. R., and Gygi, S. P. (2004) Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol. Cell. Proteomics* **3**, 729–735
6. Durr, E., Yu, J., Krasinska, K. M., Carver, L. A., Yates, J. R., Testa, J. E., Oh, P., and Schnitzer, J. E. (2004) Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat. Biotechnol.* **22**, 985–992
7. Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T. H., Crabtree, G. R., Aebersold, R., and Groudine, M. (2004) Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat. Struct. Mol. Biol.* **11**, 73–80
8. Ranish, J. A., Hahn, S., Lu, Y., Yi, E. C., Li, X. J., Eng, J., and Aebersold, R. (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nat. Genet.* **36**, 707–713
9. Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O., Goodlett, D. R., Eng, J., and Aebersold, R. (2003) The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* **33**, 349–355
10. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* **21**, 315–318
11. Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651
12. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574
13. Pusch, W., Flocco, M. T., Leung, S. M., Thiele, H., and Kostrzewa, M. (2003) Mass spectrometry-based clinical proteomics. *Pharmacogenomics* **4**, 463–476
14. Wulfkühle, J. D., Liotta, L. A., and Petricoin, E. F. (2003) Proteomic applications for the early detection of cancer. *Nat. Rev. Cancer* **3**, 267–275
15. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
16. Patterson, S. D., and Aebersold, R. H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.* **33**, (suppl.) 311–323
17. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. *Chem. Rev.* **101**, 269–295
18. Ong, S. E., Foster, L. J., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124–130
19. Flory, M. R., Griffin, T. J., Martin, D., and Aebersold, R. (2002) Advances in quantitative proteomics using stable isotope tags. *Trends Biotechnol.* **20**, S23–S29
20. Tao, W. A., and Aebersold, R. (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 110–118
21. Lu, Y., Bottari, P., Turecek, F., Aebersold, R., and Gelb, M. H. (2004) Absolute quantification of specific proteins in complex mixtures using visible isotope-coded affinity tags. *Anal. Chem.* **76**, 4104–4111
22. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6940–6945
23. Li, X. J., Pedrioli, P. G., Eng, J., Martin, D., Yi, E. C., Lee, H., and Aebersold, R. (2004) A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. *Anal. Chem.* **76**, 3856–3860
24. Hanash, S. (2003) Disease proteomics. *Nature* **422**, 226–232
25. Hanash, S. (2004) Integrated global profiling of cancer. *Nat. Rev. Cancer* **4**, 638–644
26. Domon, B., and Broder, S. (2004) Implications of new proteomics strategies for biology and medicine. *J. Proteome Res.* **3**, 253–260
27. Ideker, T. (2004) A systems approach to discovering signaling and regulatory pathways—or, how to digest large interaction networks into relevant pieces. *Adv. Exp. Med. Biol.* **547**, 21–30
28. Aebersold, R. (2003) Constellations in a cellular universe. *Nature* **422**, 115–116
29. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660–666
30. Eng, J., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **5**, 976–989
31. Desiere, F., Deutsch, E. W., Nesvizhskii, A. I., Mallick, P., King, N. L., Eng, J. K., Aderem, A., Boyle, R., Brunner, E., Donohoe, S., Fausto, N., Hafen, E., Hood, L., Katze, M. G., Kennedy, K. A., Kregenow, F., Lee, H., Lin, B., Martin, D., Ranish, J. A., Rawlings, D. J., Samelson, L. E., Shii, Y., Watts, J. D., Wollscheid, B., Wright, M. E., Yan, W., Yang, L., Yi, E. C., Zhang, H., and Aebersold, R. (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.* <http://genomebiology.com/2004/6/1/R9>