research articles Journal of proteome research

Optimized Peptide Separation and Identification for Mass Spectrometry Based Proteomics via Free-Flow Electrophoresis

Johan Malmström,[†] Hookeun Lee,[†] Alexey I. Nesvizhskii,[‡] David Shteynberg,[§] Sonali Mohanty,[||] Erich Brunner,[||] Mingliang Ye,[§] Gerhard Weber,[⊥] Christoph Eckerskorn,[⊥] and Ruedi Aebersold*[,†,#]

Institute for Molecular Systems Molecular Biology, Swiss Federal Institute for Technology Zürich (ETH), Hoenggerberg 8093 Zürich, Switzerland, Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA, The Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA, Zoological Institute, University of Zurich, 8057, Zürich, Switzerland, BD Diagnostics, Am Klopferspitz 19, 82152 Planegg, Germany, and Faculty of Natural Sciences, University of Zürich, Switzerland

Received February 23, 2006

Multidimensional LC−MS based shotgun proteomics experiments at the peptide level have traditionally been carried out by ion exchange in the first dimension and reversed-phase liquid chromatography in the second. Recently, it has been shown that isoelectric focusing (IEF) is an interesting alternative approach to ion exchange separation of peptides in the first dimension. Here we present an improved protocol for peptide separation by continuous free-flow electrophoresis (FFE) as the first dimension in a two-dimensional peptide separation work flow. By the use of a flat p*I* gradient and a mannitol and urea based separation media we were able to perform high-throughput proteome analysis with improved interfacing between FFE and RPLC−MS/MS. The developed protocol was applied to a cytosolic fraction from Schneider S2 cells from *Drosophila melanogaster*, resulting in the identification of more than 10 000 unique peptides with high probability. To improve the accuracy of the peptide identification following FFE−IEF we incorporated the p*I* information as an additional parameter into a statistical model for discrimination between correct and incorrect peptide assignments to MS/MS spectra.

Keywords: free-flow electrophoresis (FFE) • mass spectrometry • proteomics • multidimensional separation • isoelectric focusing (IEF)

Introduction

In a typical proteomics experiment, a sample of interest is separated either at the protein level or at the peptide level after enzymatic digestion of the proteins, followed by protein identification by tandem mass spectrometry (MS/MS). Various combinations of peptide/protein separation methods have been employed to generate multidimensional separation systems for comprehensive proteome analysis. The initial separation approach in proteomics was two-dimensional gel electrophoresis (2DE) where the proteins in the sample are separated by isoelectric focusing (IEF) in the first dimension and by molecular weight in the second, resulting in a two-dimensional map of proteins.[1,2] Protein spots of interest are then excised out of the gel, digested by an appropriate enzyme, typically trypsin, and identified by mass spectrometry. Although 2DE has a high separation power it suffers from limitations that restricts the usefulness of this technique.[3] The introduction of multidimensional LC−MS/MS based shotgun approaches at the peptide level in conjunction with stable isotope labeling allows higher throughput of protein identification and protein quantification.[4,5] In such a multidimensional LC−MS/MS based experiment a number of different combinations of peptide separation methods have been used. These include size-exclusion chromatography followed by reversed-phase liquid chromatography (RPLC),[6] RPLC followed by capillary electrophoresis,[7] strong cation-exchange chromatography (SCX) followed by RPLC,[5,8,9] SCX followed by avidin affinity chromatography to select specifically biotinylated peptides, followed by RPLC[10,11] or IEF followed by RPLC.[12−16] Among these, SCX followed by RPLC is the most widely used.

The highest resolution when working with multidimensional separation is achieved with two orthogonal methods,[17] as in 2DE, indicating that IEF and RPLC may be a good alternative to SCX and RPLC. Tan et al.[18] constructed a miniaturized multi chamber device for solution IEF separation of peptide mixtures. The total sample load achieved with this IEF device was as high as 10 mg of peptides. However, they did not find any improvement in the resolution of IEF for the separation of a co-digested

* To whom correspondence should be addressed: Tel: +41-44-633 3170. Fax: +41-44 633 10 51. E-mail: rudolf.aebersold@imsb.biol.ethz.ch.
[†] Institute for Systems Molecular Biology.
[‡] University of Michigan.
[§] The Institute for Systems Biology.
[||] Zoological Institute, University of Zurich.
[⊥] BD Diagnostics.
[#] Faculty of Natural Sciences, University of Zürich.

10.1021/pr0600632 CCC: $33.50

Published on Web 07/29/2006

nine-protein mixture compared with that achieved by SCX. Recently, a commercial Rotofor multichamber IEF device was applied to fractionate tryptic peptides from serum samples.[16] The identification of 437 proteins was achieved after the 20 collected IEF fractions were analyzed by μRPLC−MS/MS. However, in their study, no carrier ampholytes were added for the separation, compromising the resolution. IEF in polyacrylamide gels, typically used for the separation of proteins, was recently also used to separate peptides.[14,15,19] Using immobilized pH gradient gel (IPG) as the first separation dimension followed by RPLC/MS, Cargile and colleagues[15] identified more than 6000 peptides corresponding to more than 1200 proteins from the cytosolic fraction of E-coli when *pI* information was used to aid the identification. Another approach for IEF based peptide separation has been recently described using off-gel isoelectric focusing where *pI* and reversed phase retention time was used to validate the database search data, resulting in the identification of more than 1800 peptides from 15 fractions.[20] Shen et al.[21] performed IEF separation in fused silica capillaries and found that peptides with p*I* value differences as small as 0.01 could be resolved. Also, capillary IEF coupled online with μRPLC was used for proteome analysis of a yeast cell lysate.[12,13] It was found that because of the completely orthogonal separation mechanism the overall peak capacity of this system was significantly improved over SCX/RPLC. While high resolution was achieved, the practical application of capillary IEF in proteome research is seriously limited by the small sample loading capacity.

Continuous free flow electrophoresis (FFE) is a liquid-based IEF technique, first described by Hanning 35 years ago.[22] In FFE, samples are continuously injected into a carrier ampholyte solution flowing as a thin film between two parallel plates and by applying an electric field perpendicular to the flow direction, the analytes are separated according to their p*I*.[23] FFE offers improved sample recovery, high sample loading capacity and large flexibility when choosing separation method, which have been described in a recent review.[24] FFE has successfully been adapted to proteomics experiments as the first mode of separation for proteins in conjunction with LC peptide separation for proteome mapping of various cell lines.[25−27] Recently, FFE has also been used for separating peptides, where the high loading capacity and the p*I* information when performing database search was of advantage.[28,29] Despite this progress some experimental challenges remain, especially the interfacing difficulty between FFE and RPLC−MS which limits the efficiency of this approach.

In this study, we have optimized a protocol for improved peptide separation by FFE−IEF by using a flat p*I* gradient and by utilizing mannitol and urea in the separation media to simplify sample cleanup before LC−MS. We have also improved the statistical tool PeptideProphet[30] by incorporating the p*I* information as an additional parameter useful for discrimination between correct and incorrect peptide assignments to MS/MS spectra.

Experimental Section

FFE-Weber reagent basic kit (Prolyte 1, Prolyte 2, Prolyte 3 and Prolyte 4−7 and p*I* markers) were purchased from FFE-Weber Inc (now BD-Diagnostics, PAS). α-Hydroxyisobutyric acid (HIBA), DL-2-aminobutyric acid, nicotinamide, glycyl−glycine and ethanolamine were purchased from Sigma-Aldrich (Steinheim, Germany), AMPSO and HEPES from Roth (Karslruhe, Germany) and TAPS from ACROS (NJ). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). All other reagents and chemicals were purchased from either Sigma (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). Spin Columns for solid-phase extraction were purchased from Harvard Apparatus (Holliston, MA). Costar 96 well cell Culture Plates were purchased from Corning Incorporated (Corning, NY), and 96 well sample collection plates (deep-well, 2 mL/well) were purchased from Waters (Milford, MA).

Sample Preparation. Schneider S2 cells from *Drosophila melanogaster* were grown to high density in 150 cm^2 flasks in Schneider's Drosophila cell culture medium (Invitrogen AG, Basel, Switzerland) supplemented with 10% fetal calf serum (FCS). To harvest the cells, the suspension (a few cells adhered to the surfaces of the flask; these were resuspended by gently tapping the flasks periodically) was transferred to 50 mL Falcon tubes and centrifuged (1000 × *g*) for 7 min at 4 °C. The pelleted cells were washed 3 times with 50 mL of chilled PBS by using the initial centrifugation conditions. To begin the cell fractionation procedure the final washed cell pellet was resuspended in 5 volumes of fresh ice cold hypotonic cell lysis buffer (10 mM HEPES (pH 7.90), 1.5 mM $MgCl_2$, 10 mM KCl) supplemented just before use with 0.5 mM DTT, complete mini protease inhibitor cocktail (Roche, Basel, Switzerland) and incubated in this solution for 10 min on ice. Then the swollen cells were dounced 20 times or until all cells were visibly lysed. The efficiency of cell lysis was assessed by using phase microscopy to monitor the release of intact nuclei. The resulting lysate was transferred to corex tubes and centrifuged at 4 °C for 7 min at 1000 × *g*, which generated a nuclear pellet. The supernatant (combined cytoplasmic and membrane fractions) was removed and ultra centrifuged at 100 000 × *g* (SW 60 rotor, Beckman) at 4 °C for 30 min to generate separate membrane and cytoplasmic fractions. The supernatant (cytoplasmic fraction) was removed and precipitated by adding acetone (−20 °C) (6.5 times the sample volume) followed by incubation on ice for 5 min and centrifugation in an Eppendorf tube at 14 000 rpm for 10 min, at 4 °C. The precipitated proteins were resuspended in a buffer containing 6 M Urea, 5 mM EDTA, 10 mM Tris-Hcl pH 8.3, 0.05% SDS. The protein concentration of each fraction was measured using the RC DC Protein assay (BioRad, Hercules, CA); Bovine Serum Albumin (BSA) served as a standard. A 2-mg portion of protein was diluted 6-fold with H_2O, without alkylating the proteins after reduction since alkylation could alter peptide p*I,* and digested with 40 μg of trypsin at 37 °C for 18 h.

Free Flow Electrophoresis. Isoelectric focusing was performed using a FFE-instrument, type Prometheus from FFE Weber Inc (now BD-Diagnostics, PAS). For a detailed description of this system please see references.[23,26] A schematic illustration of the apparatus is also shown in Figure 1. All media contained 8 M Urea and 250 mM Mannitol. Except for the counter flow medium, ProLyte or other electrolytes (acid or base) were also added for the creation of a pH gradient. The anodic stabilization media contained 100 mM sulfuric acid, 100 mM α-Hydroxyisobutyric acid (HIBA), 150 mM DL-2-Aminobutyric acid, 100 mM Nicotinamide, and 15 mM glycyl−glycine, separation media 1 contained 23% prolyte mixture 4−7 and the cathodic stabilization media contained 100 mM sodium hydroxide, 75 mM Ethanolamine, 75 mM AMPSO (3-*N*-((α,α-dimethyl-hydroxyethyl)-amino-2-hydroxypropanesulfonic acid), 150 mM TAPS and 30 mM HEPES. Electrolyte anode circuit was 100 mM sulfuric acid and electrolyte cathode circuit was 100 mM sodium hydroxide.

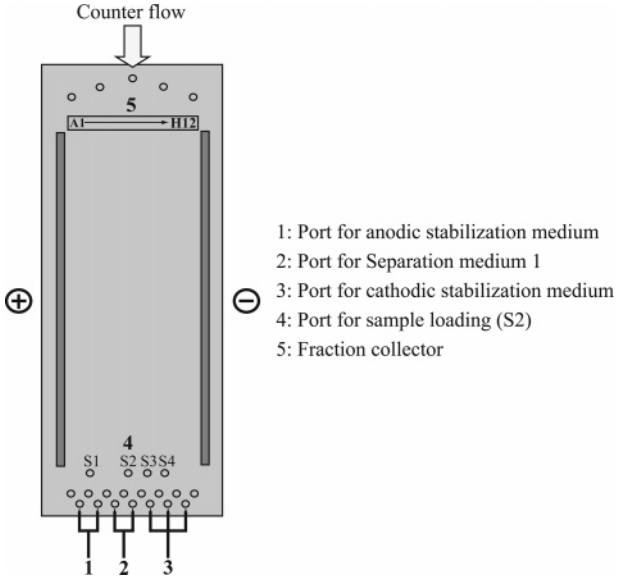

Figure 1. Schematic representation of the continuous FFE apparatus. Separation chamber: 500 × 100 × 0.4 mm; I1−I7: ports for stabilization and separation media (1−3 in the Figure); S1−S4: ports for sample delivery (S2 was used here) (4 in Figure). Counter flow medium: 8 M urea and 250 mM mannitol; anodic stabilization medium: 100 mM sulfuric acid, 100 mM α-hydroxyisobutyric acid (HIBA), 150 mM DL-2-aminobutyric acid, 100 mM nicotinamide and 15 mM glycyl-glycine, 8 M Urea and 250 mM mannitol, Separation media 1: 23% prolyte mixture 4−7, 8 M urea and 250 mM mannitol and the cathodic stabilization media: 100 mM sodium hydroxide, 75 mM ethanolamine, 75 mM AMPSO, 150 mM TAPS and 30 mM HEPES, 8 M urea and 250 mM mannitol. Flow rate for separation media: 57 mL/h; flow rate for sample delivery: 1 mL/h. Applied voltage 350 V (19 mA).

All media were prepared fresh daily. The media were pumped into the FFE separation chamber via the different ports as shown in Figure 1. The flow rate for the separation medium was set at 57 mL/h and the flow rate for delivering the sample solution was 1 mL/h. The sample solution was pumped into the separation chamber via the middle sample port (S2). To visualize the focusing process, a small amount of FFE-Weber p*I* marker solution (2−5% of total sample volume) was added to the sample. Electrophoresis was performed at 10 °C with a constant voltage of 350 V. The separated sample was collected from the 96-channel tubing outlets to a 96-well plate. For the separation of *Drosophila melanogaster* peptides and evaluation of FFE, the fractions were collected into a deep-well 96-well sample collection plate (maximum volume 2 mL for each well). In other cases, the fractions were collected into a Costar 96-well cell culture plate, and light absorbance was measured in each well at 420 nm using a PowerWave X Microplate spectrophotometer (Bio-Tek Instruments, Inc., Winooski, VT). For the fractionation of *Drosophila melanogaster* peptides, the peptides obtained from digesting two mg total protein were dissolved in 1 mL separation medium, and 20 μL of FFE-Weber *pI* marker solution was added to the sample solution. This sample was loaded continuously for 1 h at 1 mL/h. The fraction collection was started when the red dye (one of components in the FFE-Weber p*I* marker solution) reached the chamber end. When the red dye was no longer detectable at the outlets of the collection tubes, sample collection was continued for 10 min. Total collection time was 65 min and the volume of each collected fraction was about 0.5−1 mL. A Thermo Orion needle tip micro pH electrode (Thermo Electron Corporation, Beverly, MA) was used to measure the pH value of each fraction. Peptides from the FFE fractions 21 to 58 were purified by C18 reversed-phase spin columns according to the manufacture's instructions (Harvard Apparatus MA). After spin column purification, the eluted peptides where dried down and dissolved in approximately 20 μL of 0.1% formic acid. For comparing different cleanup schemas Amicon ultrafiltration was used. The Amicon ultrafiltration columns were purchased from Millipore (Concord, MA) and the peptides purified as described in the reference by Xie et al. 2005.[28]

RPLC−MS Analysis. The setup of the μRPLC system was as described previously.[31] ESI-based LC−MS/MS (LTQ Thermofinnigan San Jose, CA) analysis were carried out using an Agilent 1100 series (Agilent Technologies, Paolo Alto, CA) on a 75 μM × 10.5 cm fused silica microcapillary reversed phase column. For the capillary column, one end of polyimide-coated fused-silica capillary (Polymicro Technologies, Phoenix, AZ) was manually pulled to a fine point ~5 μm with a microflame torch. The columns were in-house packed with C18 resin (5 μm, 200 Å Magic C18AQ, Michrom BioResources, Auburn, CA) using a pneumatic pump (Brechbuehler, Spring, TX) at constant helium gas pressure of 1500 psi. Sample volumes of 2 μL were loaded onto the precolumn at a flow rate of 4 μL/min in 5 min. After sample loading, the sample was separated and analyzed at a 200 nL/min flow rate with a gradient of 5% B to 35% B over 50 min as described.[32] Mobile phase A was 0.1% formic acid and mobile phase B was 100% acetonitrile. Peptides eluting from the capillary column were selected for CID by the mass spectrometer using a protocol that alternated between one MS scan and three MS/MS scans. The three most abundant precursor ions in each survey scan were selected for CID, if the intensity of the precursor ion peak exceeded 10 000 ion counts. The electrospray voltage was set to 2.0 kV and the specific m/z value of the peptide fragmented by CID was excluded from reanalysis for 2 min.

Data Processing. MS/MS spectra were searched using the SEQUEST search tool[33] against a *Drosophila melanogaster* sequence database. The protein sequence database was extracted from the ABCC nonredundant protein database (National Cancer Institute, NIH, latest version as of Feb. 22, 2005), and contained all *Drosophila melanogaster* protein sequences as well as known contaminants such as porcine trypsin and human keratins. Reference: NRP (Non-Redundant Protein) Database: National Cancer Institute Advanced Biomedical Computing Center, 2004. ftp://ftp.ncifcrf.gov/pub/nonredun. The search was performed allowing for unconstrained enzyme search, specifying monoisotopic mass and mass tolerance of 3 Da, and with methione oxidation as a variable modification. The database search results were first processed using the PeptideProphet program[30] modified to utilize the p*I* information. Peptides were assembled into proteins and the probabilities at the protein level were computed by the protein inference program ProteinProphet.[34,35] The resulting list of proteins was filtered using ProteinProphet probability of 0.9 or greater, which in this dataset corresponded to an estimated false positive error rate of 0.8% and a sensitivity of 91%. To evaluate the improvement achieved by the introduction of the p*I* information in the statistical model of PeptideProphet, the dataset was also processed without using the p*I* information, and the numbers of peptide and protein identifications were compared at the fixed 0.9 probability threshold. The identified peptides and corresponding proteins were stored and analyzed

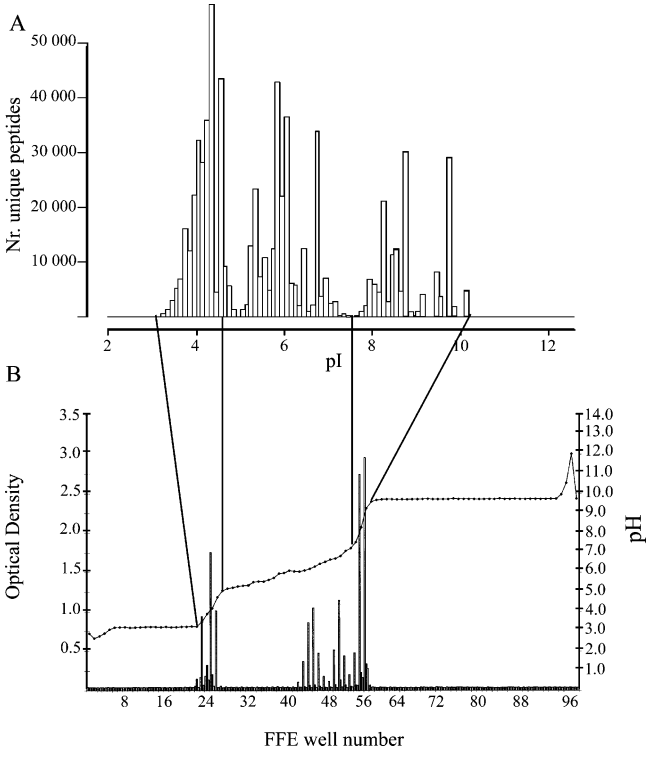

Figure 2. Selection of p*I* separation range. (A) The graph displays the p*I* and frequency of an in silico digest of the NCBI NR *Drosophila melanogaster* proteins (drosophila.nci.20041218) of peptides with a molecular weight between 800−4000 Da. (B) Separation of p*I* markers using FFE-Weber prolyte solution 4−7 in the separation medium. The optical intensity of the p*I* markers are displayed as vertical bars and the pH in corresponding well, measured by a micro pH electrode, is shown as spots connected by the line. The resulting p*I* gradient has a steeper increase in the region from 3−5 and 7.5−10, where the linear gradient from 5−7.5 is stretched out. The peptides covered by these three regions are indicated by the lines pointing to the graph with the in silico digested peptides. Two flat regions, one at low pH (∼pH 3.0) representing the pH of the anodic stabilization medium and the other at high pH (∼pH 10) representing the pH of cathodic stabilization.

in a MySQL data model, which was used to accommodate further analysis of the data.[36]

Results and Discussion

Selection of p*I* Gradient. Most of the previous FFE based IEF experiments have used a p*I* gradient between 3 and 10. However, after an in silico digest of the NCBI NR *Drosophila melanogaster* protein database (drosophila.nci.20041218) and selection of peptides with a molecular weight potentially detectable in a mass spectrometer (800−4000 Da) it becomes apparent that the majority of these peptides have a p*I* between 4 and 8 (Figure 2A). The number of peptides are especially high in the p*I* range from 4 to 7 whereas relatively few peptides have a p*I* above 9. To maximize the resolution of the 4−8 p*I* range we selected a novel nonlinear p*I* gradient (Figure 2B). This was accomplished by using the FFE-Weber prolyte solution 4−7 and the resulting p*I* gradient, determined by pH measurement of all wells, displayed a steeper increase in the gradient from 3 to 5 and 7.5 to 10 whereas region from 5 to 7.5 is stretched out. The bars in Figure 2B represent the intensity of the colored p*I* markers used to determine the resolution of the separation. As indicated by the lines in Figure 2, which connect the graph of the in silico digested peptides to the graph with the observed pH gradient, it is shown that all peptides could be found within the selected p*I* gradient, i.e., no peptides are lost. Consequently, the peptide rich region from 5 to 7.5 is stretched out, whereas the peptides in the steeper parts of the gradient are comprised into fewer wells. This results in better separation of the peptide rich regions and thereby increasing the likelihood of identifying more peptides in these crowded areas as well as reducing the number of wells in the peptide poor regions. Since the separation is carried out in 8 M urea, which has previously been shown to increase the p*K* up to one p*K* unit for acidic analytes, it is likely that more of the peptides predicted to be outside the linear part of the gradient is actually within.[37,38] Ideally, however, the linear part of the gradient should be decreased down to 4 to also stretch out the abundant region from 3.5 to 5 since poor resolution in this peptide rich region could lead to under sampling in the mass spectrometer resulting in missed peptide identifications. However, such a prolyte mixture is not yet available. Two flat regions, one at low pH (∼pH 3.0) representing the pH of the anodic stabilization medium and the other at high pH (∼pH 10) representing the pH of cathodic stabilization medium were observed. It should also be pointed out that the fact that FFE−IEF is carried out in a continuously flowing solution allows the processing of sequential samples without system regeneration, a feature that is attractive for high-throughput proteomic applications. In chromatography, the analysis time depends mainly on the elution time of the last eluting analyte and the time required for column regeneration. Also, the amount of sample is limited by the column capacity. In contrast, in FFE samples are loaded continuously and all the separated analytes are collected from the separation chamber concurrently. Because the sample is continuously injected and collected, there is no upper limit of the total sample amount being processed, the processing time does depend, however, on the analyte concentration in the sample. In this regard, the use of continuous isoelectric focusing allow the separation of unusually high amounts of sample and is thus an attractive tool for pre-fractionating the sample before other downstream separation schemas.

Type of Separation Media. The optimization of the IEF-process under free flow conditions aims at controlling band broadening effects like electrodispersion and electroendosmosis. By using appropriate additives electroendosmosis can be decreased. A suitable candidate for this is hydroxypropylmethylcellulose (HPMC), which has traditionally been used in several applications.[26,28] The rather large molecular weight range of HPMC and its chemical properties are associated with difficulties when recovering the sample for MS-analysis. To simplify the workflow we optimized a protocol without HPMC, using manntitol and urea in the separation media.

In an attempt to compare the background peaks from the mannitol/urea protocol to the background peaks from the traditionally used FFE separation media containing HPMC, three types of separation media were prepared: (I) 8 M Urea, 250 mM Mannitol, 20% Prolyte solution 4−7, (II) 0.2% HPMC, 20% Prolyte solution 4−7 and (III) 0.2% HPMC, 25% Glycerol and 20% Prolyte solution 4−7 and the solutions were cleaned up by using either spin columns or 5 kDa ultrafiltration columns. After cleanup, the samples were resuspended in buffer A and 10% of the material was analyzed by LC−MS/MS. Figure 3 shows the base peak chromatogram from the different samples, demonstrating that all protocols containing

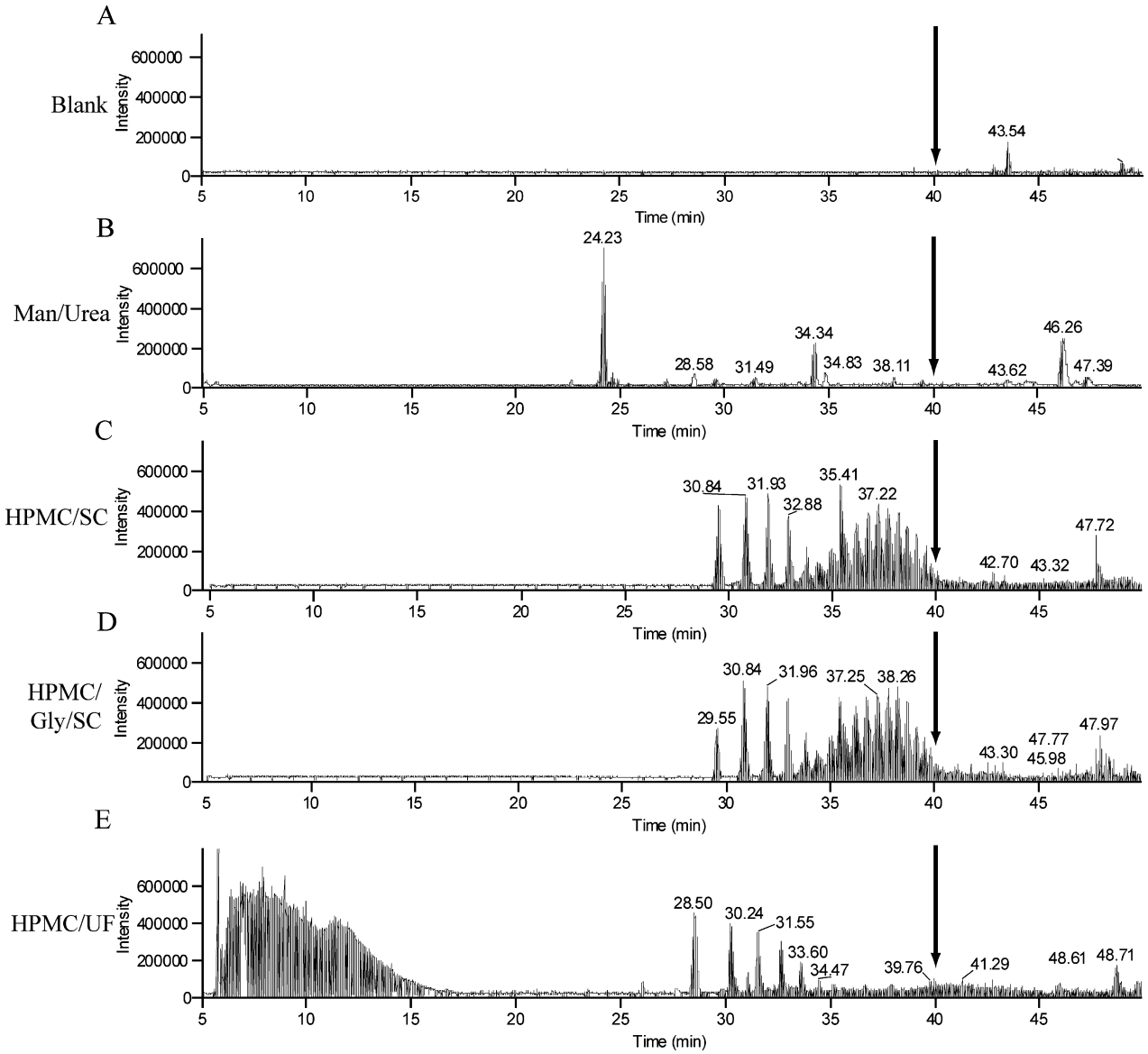

Figure 3. Base peak chromatogram of background peaks from different FFE separation medias and purifications schemes. Separation media from three different FFE separation protocols were prepared (I) 8 M Urea, 250 mM Mannitol, 20% Prolyte solution 4−7, (II) 0.2% HPMC, 20% Prolyte solution 4−7 and (III) 0.2% HPMC, 25% Glycerol and 20% Prolyte solution 4−7 followed by cleanup by ultrafiltration (5 kDa) or C_{18} spin columns to monitor background peaks from the separation media. (A) Blank run, (B) Separation media I and spin columns, (C) Separation media II and spin columns, (D) Separation media III and spin columns, and (E) Separation media II and ultrafiltration.

HPMC resulted in higher background base peak intensities (Figure 3C−E) compared to a blank run (Figure 3A) and the mannitol urea protocol (Figure 3B). The background was especially severe in the HPMC based separation media cleaned-up by C_{18} (Figure 3C,D). The mannitol/urea based separation media, cleaned-up with C_{18} spin columns, contained one major peak at around 24 min (Figure 3B), corresponding to an acetonitrile concentration of 11%. This peak seems to be protocol specific and likely origin of this peak is from the chemicals used. The exact source of this peak was not determined, however, since it elutes off at low acetonitrile concentrations and will not significantly interfere with peptide identification. Cleanup of the HPMC based protocols with spin columns resulted in contaminating peaks starting to elute at around 16% acetonitrile concentration and continued to elute for the entire gradient, which would interfere significantly with peptide identification (Figure 3C,D). In contrast, clean up of the HPMC based protocol with ultrafiltration significantly reduced the contaminants during peptide elution, but instead potential hydrophilic polymers with small molecular weight were detected early in the elution (Figure 3E). Separation media I and III were not compatible with ultrafiltration and were not analyzed by LC−MS/MS. These experiments were repeated twice and gave in our hands reproducible results.

One MS-scan from the five base peak chromatograms was selected around 40 min elution, corresponding to 25% acetonitrile concentration, indicated by an arrow in Figure 3. The background masses observed in these MS-scans was relatively small in the blank run and in the mannitol/urea based protocol (Figure 4A,B). However, all samples containing HPMC generated a significant increase in background features (Figure 4C−

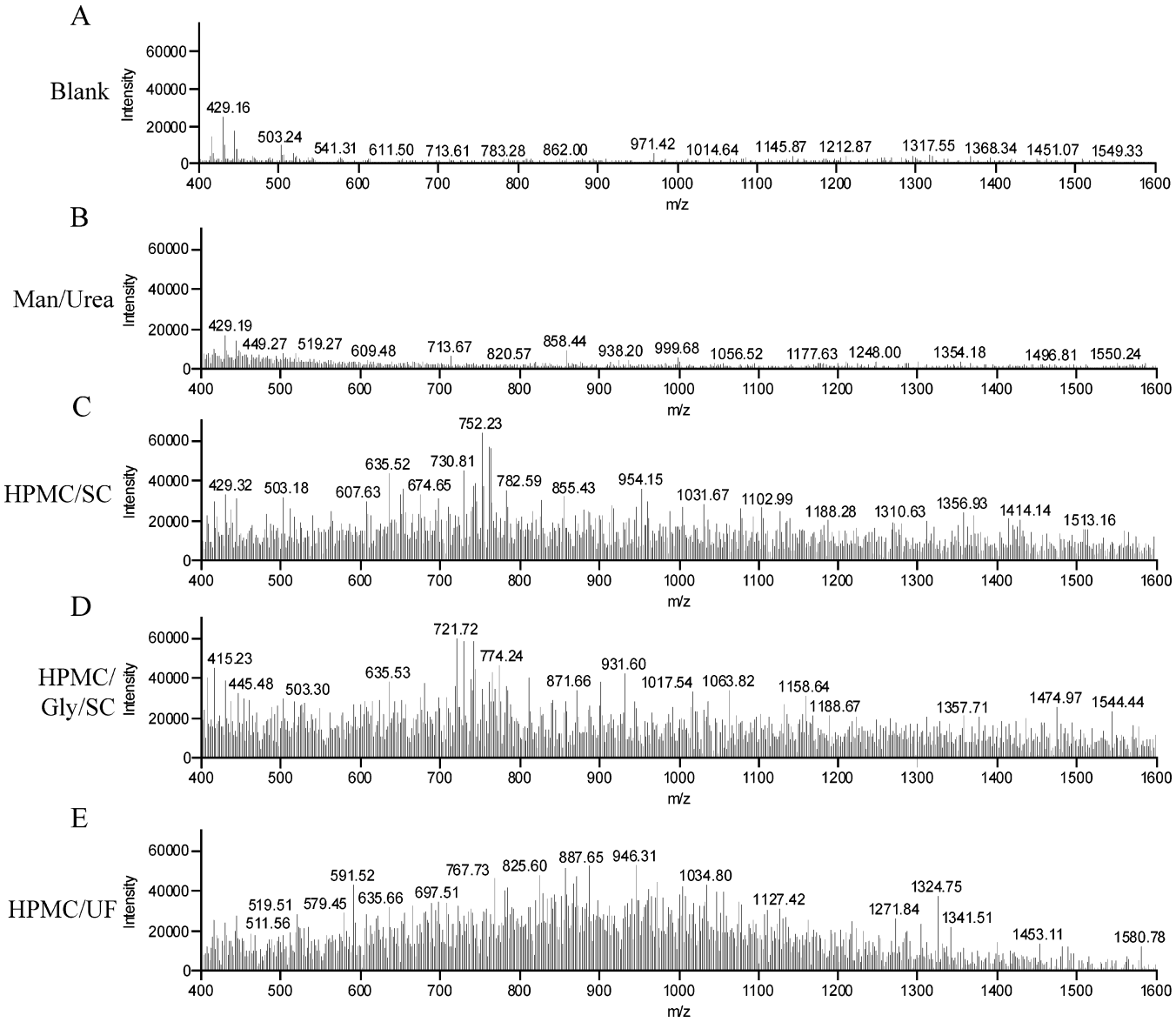

Figure 4. Representative mass spectra of background peaks from different purifications schemes. One MS scan around 40 min in the base peak chromatograms in Figure 3, corresponding to 25% acetonitrile concentration, was selected. (A) Blank run, (B) separation media I and C_{18} spin columns, (C) separation media II and C_{18} spin columns, (D) separation media III and C_{18} spin columns, and (E) separation media II and ultrafiltration.

E), even though the MS-scans were selected in a region with relatively low base peak chromatogram intensity (see Figure 3).

These data show that mannitol, urea, and prolytes 4−7 can efficiently be removed by a C_{18} clean up step, resulting in simultaneous sample clean up and concentration. The use of C_{18} spin columns is substantially faster than using ultrafiltration columns and in addition, by the use of C_{18} 96-well plate the sample preparation after FFE cleanup can be performed with high-throughput.

Fractionation of *Drosophila melanogaster* Cytoplasmic Sample. To asses the performance of the developed protocols, a tryptic digest of a cytosolic protein fraction from *Drosophila melanogaster* was analyzed. Two mg of total tryptic peptides were separated by FFE−IEF as described in the Experimental Section. From the following RPLC−MS/MS analysis of the separated fractions, a total number of 10 020 unique peptides corresponding to 1848 unique proteins above a ProteinProphet probability of 0.9 were identified, out of which 367 proteins were single hits. Figure 5 shows a two-dimensional density plot[39] of detected masses from FFE well 29, where the mass distribution across the reversed phase elution profile can be seen. No obvious contaminating peaks can be observed in the displayed peptide rich area of the chromatogram in the region corresponding to 12−35% acetonitrile concentration. It is noteworthy that the background contaminating peak from the mannitol/urea protocol is outside the selected area (see Figure 3B). After processing and analyzing the remaining fractions, we could conclude that peptides were present in every well containing separation media (fraction 24−59) ranging from measured pH from 3.5 to 9.75 as expected (Figure 6A). The most populated fractions were present in the pH range of 4−6 and correlates to the abundance level of the theoretical digest (Figure 2A), whereas fewer peptides were found in the pH range of 6−6.5. In the lower pH range, the more populated region, the standard deviation (S. D.) of the calculated p*I* of the

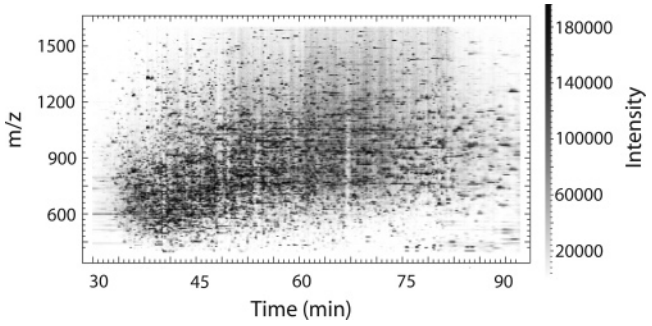

Figure 5. Two-dimensional density plot of observed masses from FFE fraction 29. Peptides from FFE well number 29 were purified by C_{18} spin columns and analyzed by LC−MS/MS. Mass over retention time of detected features were plotted, corresponding to acetonitrile concentration 12−35%.

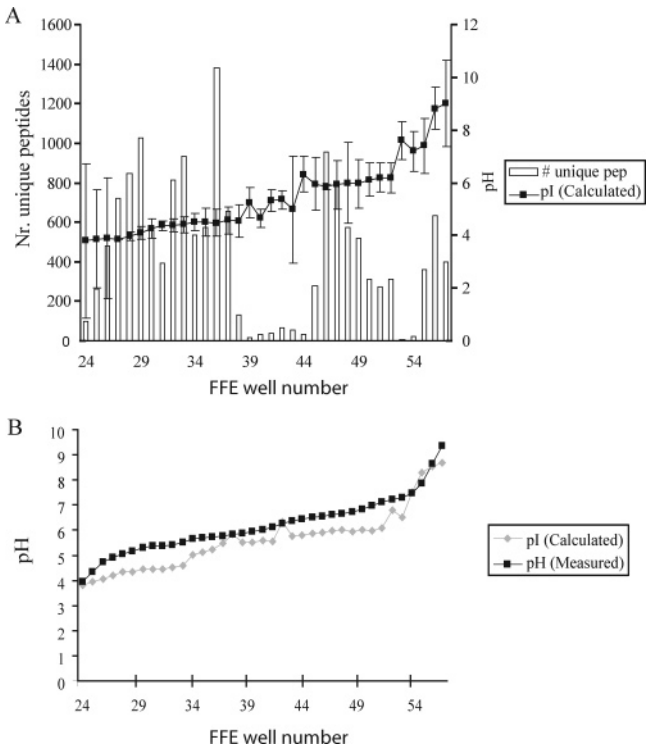

Figure 6. Identified peptides and their average theoretical p*I*. Peptides were separated by the mannitol/urea based protocol using FFE Weber prolyte solution 4−7, followed by peptide identification by LC−MS/MS and database search using Sequest, and peptide identity verification by PeptideProphet.[30] Only peptides with a peptide prophet score above 0.9 were included in the graph. (A) The sum of identified peptides across FFE wells, from one LC−MS/MS run per well, is shown as vertical bars in the graph and the calculated average p*I* of the peptides in corresponding well is shown as dots connected by the line. Error bars represents calculated p*I* standard deviation. (B) Comparison of the calculated theoretical p*I* (♦) to the measured pH (■) in every well.

identified peptides with a peptide probability of >0.9 is relatively small (±0.3), whereas in the region with less peptide identification the S. D. is larger (±0.75) (Figure 6A). This observation was reproducible for repeated runs and for other peptide samples. Partly this could possibly be explained by the fact that fewer peptides were identified in the wells with high S. D., but there seems also to be a trend that the small S. D. is located in the acidic range. When plotting the molecular weights of the identified peptides with high probability over FFE well number, larger peptides are found in the acidic end (data not shown), which also could suggest that larger peptides focus better in the FFE separation chamber. This could be due to the fact that larger peptides are less prone to diffusion during the separation.

The effectiveness of isoelectric focusing by FFE was confirmed by calculating the p*I* values of the identified peptides and comparing the calculated values to the measured ones. Figure 6B shows a plot of the average p*I* of the peptides identified in the fraction and the measured pH of the fractions vs fraction numbers. The linear correlation ($r^2 = 0.96$) confirms that isoelectric focusing by FFE effectively separates peptides with a high degree of predictability. The average difference of the pH values between the calculated p*I* and the measured pH was 0.57. This difference corresponds approximately to the shift created by the addition of urea to the separation media, which has previously been shown to increase the p*K* of the analytes.[37,38] This means that the predicted p*I* values of peptides are close to the measured pH values. In addition, to evaluate the resolution of the FFE−IEF separation, the number of wells every peptide was present in was determined. About 87% of the peptides identified with very high confidence (PeptideProphet probability above 0.99) were confined to one or two fractions and 70% of the peptides with a PeptideProphet probability of 0.9. The average pH difference between consecutive FFE fractions using the 4−7 linear p*I* gradient is 0.1 pH units indicating that the resolution for this fractionation technique is around 0.2 pH units for 70% of the peptides with a PeptideProphet probability of 0.9.

Use of p*I* Information in the Validation of Peptide Assignments to MS/MS Spectra. The p*I* information from FFE fractionation can be used when validating the protein database search results of MS/MS spectra as previously been shown.[15,20,28] To take full advantage of this we incorporated p*I* information into PeptideProphet.[30] The statistical model implemented in PeptideProphet is based on the use of the expectation maximization (EM) algorithm to derive a mixture model of correct and incorrect peptide assignments from the data. It takes as input the observed information about each assigned peptide in the dataset (database search scores and the properties of the assigned peptide), models the distributions of scores for the entire dataset and, at the end, computes for each assignment a probability of being correct. In addition to the database search score (or multiple scores in the case of SEQUEST), PeptideProphet typically utilizes the difference between the measured and theoretical peptide mass, the number of termini consistent with the type of enzymatic cleavage used, and the number of missed cleavage sites. In this work, the statistical model of PeptideProphet was extended to include the p*I* value. The p*I* information is incorporated in the model as follows. First, the probabilities of peptide assignments to spectra are computed without taking the p*I* values into consideration. Second, for each FFE fraction, the average p*I* of the correctly identified peptides in that fraction is calculated by summing the theoretical (determined from the sequence) p*I* values of each peptide assignment and using the peptide probabilities obtained at the first step as weighting factors. The standard deviation of p*I* values in each FFE fraction is computed in an analogous manner. Third, a standardized score (p*I* score) is calculated for each peptide assignment by subtracting the average p*I* value and dividing it by the standard deviation (using the average value and the standard deviation for the corre-

sponding FFE fraction). For simplicity, p*I* scores are discretized by binning. The probabilities of having a particular p*I* score among correct and incorrect assignments are estimated for each p*I* score bin, and then used to adjust the initial probability that peptide assignments are correctly computed without taking p*I* information into account. Inclusion of p*I* information in the model results in increased probability of lower scoring peptides if they have a p*I* close to the average p*I* observed for high scoring peptides from the same fraction, and in decreased score of peptides which have a p*I* far away from the average p*I*. The actual amount of adjustment is dataset dependent and learned from the data by the statistical model. Since the p*I* score is based on the average calculated p*I* of the identified peptides, the model is independent of changes in buffer conditions and temperature which influences the p*K* of the analytes. Thus, the p*I* function in the PeptideProphet is applicable to any type of IEF based separation independent of the separation conditions used.

In an attempt to measure the impact of including the p*I* information in the statistical model, all LC−MS/MS runs were run twice separately through the PeptideProphet and ProteinProphet: once using the p*I* information and the second time without. The use of the p*I* information resulted in a 4.6% increase in the number of proteins identified with probability of 0.9 and greater (1848 vs 1767 proteins) and a similar increase at the level of unique peptides (10020 and 9577 peptides identified with ProteinProphet adjusted peptide probability of 0.9 or greater with and without using the p*I* information, respectively). The overall relatively small increase in the number of identified peptides and proteins compared to other studies[14,28] indicates that PeptideProphet performs well even without the p*I* information.

Separation Properties and Analysis of Separation. To identify as many proteins as possible in multidimensional separation proteomics experiments is it necessary that the different types of separation modes are orthogonal. The peptides identified in this study, separated by IEF-FFE followed by RPLC, provide a data set that can be used to study the relationship between physical properties of the peptides and the separation process. For example, when the peptides elute off the RPLC column and are ionized in the mass spectrometer every peptide is assigned a scan number. The later the peptides are eluted of the RPLC the higher scan number will be assigned to the peptide. Since there is a relationship between hydrophobicity and elution time, there is thus is a correlation between peptide hydrophobicity and scan number. The higher the scan number, the more hydrophobic is the peptide. It has previously been shown that there is a relationship between peptide molecular weight and hydrophobicity.[40] Figure 7A displays this relationship, where molecular weight is plotted on the *y*-axis and scan number on the *x*-axis ($r^2 = 0.15$). However, when the scan number is plotted against FFE well it becomes obvious that there is no significant correlation (Figure 7B), indicating a lack of relationship between peptide p*I* and hydrophobicity. Therefore, a two-dimensional separation system using the two methods, IEF and RPLC, results in an orthogonal separation with high peak capacity.

Conclusion

IEF and RPLC are two orthogonal separation techniques. A combination of these two techniques is useful for shotgun proteomics multidimensional peptide separation experiments. FFE−IEF has the advantage of very high sample load, short separation times and no regeneration time between consecutive runs, which is an attractive feature when doing shotgun proteomics experiments. With the modified protocol, containing urea and mannitol in the separation media, sample processing before LC−MS/MS is improved, resulting in fewer interfering peaks. Another advantage of using FFE−IEF is that peptides with similar p*I* elute in the same FFE fraction, which can be then used when validating peptide identity. In this study, we have integrated this feature in the statistical model of PeptideProphet. Since IEF is a focusing method, the peak capacity could be enlarged with an adapted pH gradient. A useful future development would be to extend the gradient so that the peptide rich region around pH four would be within the linear part of the gradient.

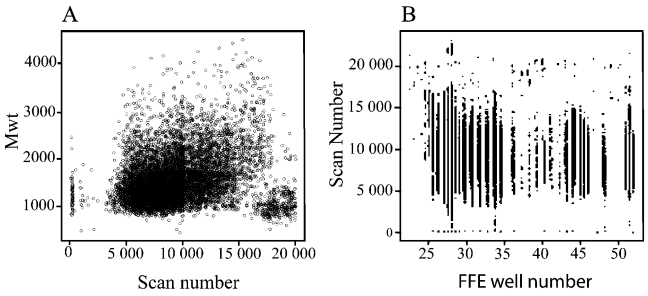

Figure 7. Analysis of separation. Identified peptides with a peptide prophet score above 0.9 were used. (A) Molecular weight was plotted against scan number (retention time) and (B) Scan number (retention time) was plotted against FFE Well.

Abbreviations. FFE, free-flow electrophoresis; IEF, isoelectric focusing; 2DE, two-dimensional gel electrophoresis; MS, mass spectrometry; SCX, strong cation-exchange chromatography; RPLC, reversed-phase liquid chromatography; HPMC, hydroxypropylmethylcellulose.

Acknowledgment. This work was supported by SNF Grant (No. 3100A0-107679/1) and with Federal (US) Funds from the National Heart, Lund and Blood institute, national institute of health, under contract No. N01-HV-28179. J.M. was supported by a postdoctoral fellowship from Wennergren foundations and the Swedish Society for Medical Research (SSMF).

References

(1) O'Farrell, P. Z.; Goodman, H. M. *Cell* **1976**, *9* (2), 289.
(2) Klose, J. *Humangenetik.* **1975**, *26* (3), 231.
(3) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (17), 9390−9395.
(4) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17* (10), 994−999.
(5) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd. *Nat. Biotechnol* **2001**, *19* (3), 242−247.
(6) Opiteck, G. J.; Jorgenson, J. W. *Anal. Chem.* **1997**, *69* (13), 2283−2291.
(7) Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. *Anal. Chem.* **1997**, *69* (8), 1518−1524.
(8) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *J. Proteome Res.* **2003**, *2* (1), 43−50.
(9) Shen, Y.; Jacobs, J. M.; Camp, D. G., 2nd; Fang, R.; Moore, R. J.; Smith, R. D.; Xiao, W.; Davis, R. W.; Tompkins, R. G. *Anal. Chem.* **2004**, *76* (4), 1134−1144.
(10) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol* **2001**, *19* (10), 946−951.
(11) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. *J Proteome Res* **2002**, *1* (1), 47−54.
(12) Chen, J.; Balgley, B. M.; DeVoe, D. L.; Lee, C. S. *Anal. Chem.* **2003**, *75* (13), 3145−3152.
(13) Chen, J.; Lee, C. S.; Shen, Y.; Smith, R. D.; Baehrecke, E. H. *Electrophoresis* **2002**, *23* (18), 3143−3148.

(14) Cargile, B. J.; Bundy, J. L.; Freeman, T. W.; Stephenson, J. L., Jr. *J. Proteome Res.* **2004**, *3* (1), 112−119.
(15) Cargile, B. J.; Talley, D. L.; Stephenson, J. L., Jr. *Electrophoresis* **2004**, *25* (6), 936−945.
(16) Xiao, Z.; Conrads, T. P.; Lucas, D. A.; Janini, G. M.; Schaefer, C. F.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D. *Electrophoresis* **2004**, *25* (1), 128−133.
(17) Giddings, J. C. *J. High Resolut. Chromatogr. Commun.* **1987**, *10*, 319−323.
(18) Tan, A.; Pashkova, A.; Zang, L.; Foret, F.; Karger, B. L. *Electrophoresis* **2002**, *23* (20), 3599−3607.
(19) Lu, Y.; Bottari, P.; Turecek, F.; Aebersold, R.; Gelb, M. H. *Anal. Chem.* **2004**, *76* (14), 4104−4111.
(20) Heller, M.; Ye, M.; Michel, P. E.; Morier, P.; Stalder, D.; Junger, M. A.; Aebersold, R.; Reymond, F.; Rossier, J. S. *J. Proteome Res.* **2005**, *4* (6), 2273−2282.
(21) Shen, Y.; Berger, S. J.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **2000**, *72* (9), 2154−2159.
(22) Zeiller, K.; Pascher, G.; Hannig, K. *Hoppe Seylers. Z. Physiol. Chem.* **1970**, *351* (4), 435.
(23) Weber, G.; Bocek, P. *Electrophoresis* **1998**, *19* (10), 1649−1653.
(24) Krivankova, L.; Bocek, P. *Electrophoresis* **1998**, *19* (7), 1064.
(25) Wang, Y.; Hancock, W. S.; Weber, G.; Eckerskorn, C.; Palmer-Toy, D. *J. Chromatogr. A* **2004**, *1053* (1−2), 269−278.
(26) Hoffmann, P.; Ji, H.; Moritz, R. L.; Connolly, L. M.; Frecklington, D. F.; Layton, M. J.; Eddes, J. S.; Simpson, R. J. *Proteomics* **2001**, *1* (7), 807−818.
(27) Moritz, R. L.; Ji, H.; Schutz, F.; Connolly, L. M.; Kapp, E. A.; Speed, T. P.; Simpson, R. J. *Anal. Chem.* **2004**, *76* (16), 4811−4824.
(28) Xie, H.; Bandhakavi, S.; Griffin, T. J. *Anal. Chem.* **2005**, *77* (10), 3198−3207.
(29) Xie, H.; Rhodus, N. L.; Griffin, R. J.; Carlis, J. V.; Griffin, T. J. *Mol. Cell. Proteomics* **2005**, *4* (11), 1826−1830.
(30) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74* (20), 5383−5392.
(31) Yi, E. C.; Lee, H.; Aebersold, R.; Goodlett, D. R. *Rapid Commun. Mass Spectrom.* **2003**, *17* (18), 2093−2098.
(32) Lee, H.; Yi, E. C.; Wen, B.; Reily, T. P.; Pohl, L.; Nelson, S.; Aebersold, R.; Goodlett, D. R. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2004**, *803* (1), 101−110.
(33) Eng, J.; McCormack, A. L.; Yates, J. R. (r) *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(34) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2003**, *75* (17), 4646−4658.
(35) Keller, A.; Eng, J.; Zhang, N.; Aebersold, R.; li, X. J. *Molecular Systems Biology* **2005**, 1, msb4100024-E1−msb4100024-E8.
(36) Malmstrom, L.; Marko-Varga, G.; Westergren-Thorsson, G.; Laurell, T.; Malmstrom, J. *BMC Bioinformatics* **2006**, *7*, 158.
(37) Bjellqvist, B.; Hughes, G. J.; Pasquali, C.; Paquet, N.; Ravier, F.; Sanchez, J. C.; Frutiger, S.; Hochstrasser, D. *Electrophoresis* **1993**, *14* (10), 1023−1031.
(38) Gianazza, E.; Artoni, G.; Righetti, P. *Electrophoresis* **1983**, *4*, 321−326.
(39) Li, X. J.; Pedrioli, P. G.; Eng, J.; Martin, D.; Yi, E. C.; Lee, H.; Aebersold, R. *Anal. Chem.* **2004**, *76* (13), 3856−3860.
(40) Horváth, C. *High-Performance Liquid Chromatography: Advances and Persectives*, ed.; Academic Press: New York, 1983; Vol. 1, p 88−155.

PR0600632