Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1053 (2004) 269–278
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Free flow electrophoresis coupled with liquid chromatography–mass spectrometry for a proteomic study of the human cell line (K562/CR3)

Yonghui Wang[a], William S. Hancock[a,*], Gerhard Weber[b], Christoph Eckerskorn[b], Darryl Palmer-Toy[c,d]

[a] *Department of Chemistry and Chemical Biology, Barnett Institute, Northeastern University, Boston, MA 02115, USA*
[b] *FFEWeber GmbH, Martinsried D-82152, Germany*
[c] *Renal Unit, Massachusetts General Hospital, Boston, MA 02114, USA*
[d] *Harvard-Partners Center for Genetics and Genomics, Cambridge, MA 02139, USA*

Abstract

The requirement for prefractionation in proteomic analysis is linked to the challenge of performing such an analysis on complex biological samples and identifying low level components in the presence of numerous abundant housekeeping and structural proteins. The employment of a preliminary fractionation step results in a reduction of complexity in an individual fraction and permits more complete liquid chromatography/mass spectrometry (LC/MS) analysis. Free flow electrophoresis (FFE), a solution-based preparative isoelectric focusing technique, fractionates and enriches protein fractions according to their charge differences and is orthogonal in selectivity to the popular reversed phase high performance liquid chromatography (HPLC) fractionation step. In this paper, we explored the advantages of a combination of FFE and liquid chromatography/mass spectrometry to extend the dynamic range of a proteomic analysis of a complex cell lysate. In this study, the whole cell lysate of a chronic myelogeneous leukemia cell line, K562/CR3, was prefractionated by FFE into 96 fractions spanning pH 3–12. Of these, 35 fractions were digested with trypsin and then analyzed by LC/MS. Depending on the algorithm used for peptide assignment from MS/MS data, at least 319 proteins were identified through database searches. The results also suggested that pI could serve as an additional criterion besides peptide fragmentation pattern for protein identification, although in some cases, a pI shift might indicate post-translational modification. In summary, this study demonstrated that free flow electrophoresis provided a useful prefractionation step for proteomic analysis and when combined with LC/MS allowed the identification of significant number of low level proteins in complex samples.

Keywords: Proteomic; Electrophoresis; Liquid chromatography; Mass spectrometry; Prefractionation

1. Introduction

The enormous dynamic range of complex biological samples such as human plasma and human cell lines represents a significant challenge to proteomic research. For example, in human plasma, the dynamic range is over 10^{10} for common clinical protein analytes [1]. For any single technique, such as 1D-LC/MS, it is extremely difficult to achieve such a deep protein profile with reasonable coverage and thus prefractionation appears to be an important step in exploring disease biomarkers, many of which are low abundant proteins. Various prefractionation approaches has been used to solve this problem, such as affinity [2], ion-exchange [3] and reversed-phase [4] chromatography, and isoelectric focusing (IEF) [5–9]. All these methods fractionate proteins through different mechanisms with improved protein identifications, especially for low abundance proteins due to a reduction in complexity of a

Abbreviations: 2DE, two-dimensional electrophoresis; Arg, arginine; DTT, dithreitol; FFE, free flow electrophoresis; GPRW, a probable G protein-coupled receptor GPR32; hGH, human growth hormone; HPLC, high performance liquid chromatography; ID, identification; IEF, isoelectric focusing; LC, liquid chromatography; Lys, lysine; MS, mass spectrometry; MW, molecular weight; MY15, myosin; NUP214, nuclear pore complex protein NUP214; PBS, phosphate buffer saline; PTM, post-translational modification

* Corresponding author. Tel.: +1 617 373 4881; fax: +1 617 373 2855.
E-mail address: wi.hancock@neu.edu (W.S. Hancock).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.08.164

given fraction at the point of analysis by mass spectrometry (MS).

The two most popular approaches for proteomic analysis use either 2D gel electrophoresis to separate a protein mixture by a combination of MW and charge-based separations [10] and the shotgun sequencing approach, 1D- or 2D-LC/MS [11,12] where the corresponding tryptic peptides are analyzed by reversed phase LC/MS. In the latter approach, the protein mixture may not be separated before the enzymatic digestion step and the resulting highly complex peptide mixture provides an enormous analytical challenge. It is, therefore, of interest to study prefractionation approaches that can be readily combined with LC/MS. One promising approach, free flow electrophoresis (FFE) fractionates complex protein samples in the IEF mode [13]. Basically, proteins are deflected and focused in an electric field based on their pH-dependant charge density. This is achieved by the continuous transport of the sample in a thin, laminar, pH-graded buffer flow, and the perpendicular application of an electric field as a deflecting force. FFE is a high-resolution fractionation technique and it is compatible with most post-fractionation separations, such as 2D-electrophoresis (2DE) and 2D-LC. Recent applications in proteomics research of FFE include the isolation of rat liver peroxisome subpopulations [14–17], study of mammalian growth hormone [18], subcellular fractionation of *Saccharomyces cerevisiae* mitochondria [19], and separation of cytosolic proteins [20].

In this paper, we report the first combination of FFE and LC/MS for proteomic analysis. We have studied the human chronic myelogenous leukemia cell line, K562, transfected to express protein CR3 (CD11b/CD18). The K562 cell line is a widely used model system for study of the control of differentiation in chronic human myelogenous leukemia and normal myeloid development [21] and thus it is of great interest to study the protein profile of this cell line. Whole K562/CR3 cell lysates were first fractionated into 96 vials based on the net charge differences of the constituent proteins by FFE. After trypsin digestion, the proteome of each fraction was analyzed by LC/MS.

We found that the use of a prefractionation step prior to proteomic analysis resulted in more protein identifications (IDs) with higher sequence coverage and a protein profile with a greater dynamic range than was achieved with direct LC/MS analysis of the tryptic digest. Through a SEQUEST database search, 319 proteins were identified with criteria of a cross-correlation score (Xcorr) of at least 1.8, 2.5, and 3.5 for charge states +1, +2, and +3, respectively. We were able to use the measured pI values relative to the expected pI values for the fractions to screen the false-positive and -negative protein assignments. This analysis was informative in assessing different SEQUEST searching criteria. The proteomic analysis identified a number of proteins, such as integrins and semaphorins, which could be related to the properties of a transformed cell line.

2. Experimental

2.1. Substrates and chemicals

The trypsin (sequencing grade) was purchased from Promega (Madison, WI). Dithiothreitol, iodoacetamide, ammonium bicarbonate were all obtained from Sigma–Aldrich (St. Louis, MO). Formic acid, acetone and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ). The water used in all experiments was purified through a Milli-Q system, Millipore (Bedford, MA).

2.2. Cell lysis

The cell line K562 (human chronic myelogenous leukemia), transfected to express CR3 (CD11b/CD18 heterodimer), was expanded to $\sim 2 \times 10^7$ cells in DMEM/10% FCS/0.5 mg/mL G418. Cells suspended in media were mixed with an equal part of phosphate buffer saline (PBS)/0.25 M sucrose/35 mM sodium chloride/Roche Complete Protease Inhibitor Cocktail/5 mM sodium pyrophosphate/20 mM sodium fluoride/1 mM sodium vanadate (phosphatase inhibitors) at 4 °C, and centrifuged for 5 min at 500 × *g*. The cell pellet was taken up in 1 mL of the same PBS/protease inhibitor/phosphatase inhibitor buffer, and placed in a Parr apparatus at 4 °C. After 15 min equilibration at 1500 psi nitrogen, the cells were lysed by explosive decompression and mixed 1:1 with lysis buffer. Lysis buffer is 2 M thiourea/7 M urea/4% CHAPS/1% dithreitol (DTT)/2% pharmalyte. The protein concentration was approximately 2.5 mg/mL.

2.3. FFE separation

The separation of proteins from the raw homogenate was done prior the identification of individual components by mass spectrometry. Free-flow electrophoresis-based isoelectric focusing was performed using a prototype instrument, type OCTOPUS (FFEWeber GmbH, Planegg/Munich, Germany).

The crude homogenate was subjected to the FFE using the following settings: (i) the anodic and cathodic circuit electrolytes consisted of 100 mM sulfuric acid and 100 mM sodium hydroxide, respectively; (ii) the anodic and the cathodic electrolyte stabilisers (buffer inlet 1 and 6/7) were 7 M urea, 2 M thiourea, 250 mM manitol, 100 mM sulfuric acid, and 100 mM sodium hydroxide; (iii) the separation buffers (buffer inlet 2–5) consisted of 7 M urea, 2 M thiourea, and 250 mM manitol. Prolytes^{TM} (Tecan, Männedorf, Switzerland) were used to establish a linear pH-gradient from 3.0 to 11.5. 10 mM DTT, 0.6% β-octylglucopyranoside, and 0.1% Triton X-114 were added to the acidic and neutral separation buffers; (iv) counterflow buffer consisted of 7 M urea, 2 M thiourea, and 250 mM manitol.

Free-flow electrophoresis was performed in horizontal mode at 10 °C with a total flow rate of ∼51 g/h within the separation chamber at a voltage of 820 V (∼18 mA). The samples

were diluted in the separation medium, containing Prolyte #2 and applied to the separation chamber via the middle sample inlet with a flow rate of 1.5–2 mL/h, which corresponds to a protein throughput of 2–4 mg/h. Residence times in the separation chamber were ~23 min. Fractions were collected in 96-well plates, numbered 1 (anode) through 96 (cathode) and analysed by SDS-PAGE using an XCell SureLock™ Mini-Cell (Invitrogen, Carlsbad, CA, USA) in combination with precast NuPAGE® 4–12% Bis–Tris gels (Invitrogen). Staining of the proteins was carried out using a SilverQuest™ kit (Invitrogen) according to the manufacturer's instructions.

2.4. Tryptic digestion

To 100 uL of each FFE fraction, 25 μL 1 M NH_4HCO_3 buffer (pH 8.0) was added, followed by addition of 0.75 μL 1 M DTT. The mixture was incubated at 75 °C for 1 h. After it cooled down, 7.5 μL iodoacetamide (1 M, freshly prepared) was added and the mixture was incubated at room temperature in dark for 2 h. Proteins were precipitated by adding acetone (9:1, v/v) to the sample. After 15 min incubation at ambient temperature, the sample was spun at 10,000 × *g* for 15 min. The supernatant was removed, then the precipitate was reconstituted with 50 μL 0.1 M NH_4HCO_3, and treated with 1 μL trypsin (0.05 μg/μL) for overnight digestion. Another aliquot of 1 μL trypsin was added for 6 h to ensure complete digestion. The crude cell lysate was diluted 1:5 (v/v) with 200 mM NH_4HCO_3 and digested following the same procedure.

2.5. LC/MS analysis

The LC/MS was performed on a Surveyor LC system coupled with a LCQ DECA XP system (ThermoFinnigan, San Jose, CA). A ThermoHypersil C-18 column (BioBasic, 180 μm × 10 cm) was used for all LC/MS analyses. The flow rate was maintained at 100 μL/min before splitting and 2 μL/min after flow split. The mobile phases used for LC were 0.1% formic acid in water (A) and acetonitrile (B). The gradient was programmed as follows: 5% B hold for 1 min, then ramped to 40% B in 90 min, and to 80% B for another 20 min, finally hold at 80% B for 25 min. The longer gradient used for crude cell lysate digests was: 5% B hold for 1 min, ramped to 40% B in 180 min, to 80% B in another 65 min, and finally hold at 80% B for 25 min. The temperature of the ion transfer tube was set at 185 °C. The spray voltage was set at 3.3 kV and the normalized collision energy was set at 35% for MS/MS.

2.6. Bioinformatics

Protein identification was obtained through a database search using the SEQUEST algorithm incorporated into the BioWorks software (version 3.1). Theoretical pIs were obtained through the BioWorks interface. The SEQUEST search results were assessed by examination of the Xcorr scores. Different Xcorr scores and pI correlations were used to check the result of protein identifications from the LC/MS analysis of each FFE fraction.

3. Results and discussion

3.1. FFE separation

FFE is a fluid phase fractionation technique, which separates and focuses proteins in an inhomogeneous medium (pH gradient) with a high concentration of salts, detergents and reducing reagents. Proteins are focused and enriched based on their pI differences. In this work, the crude cell lysate was fractionated into 96 vials by FFE and afterwards the pH was measured for each fraction. The pH gradient of all FFE fractions ranged approximately from 3 to 12. The crude cell lysate and FFE fractions were analyzed by SDS-PAGE first, see Fig. 1. Most of the proteins were focused in fraction 27 through 61, with the pH ranging linearly from 4.75 to 9.60. In the basic fractions 57–61, proteins were focused into a few bands. On the other hand, many more proteins appeared in the

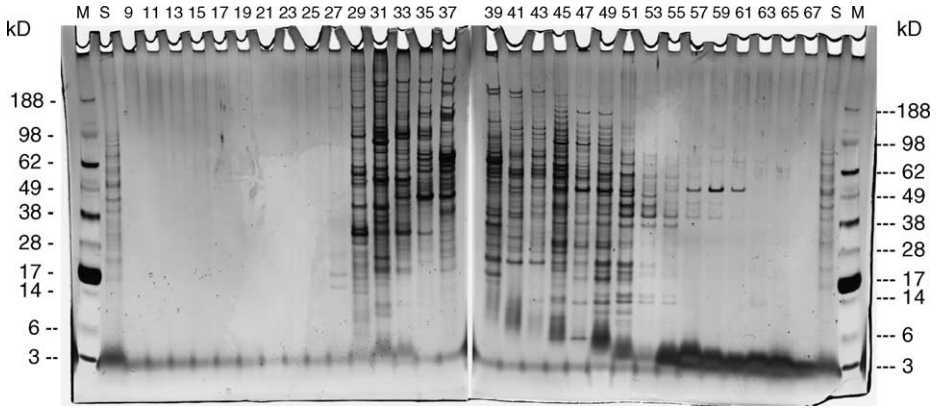

Fig. 1. The analysis of FFE fractions by SDS-PAGE. The SDS-PAGE analyses were performed with preCAST NuPAGE 4–12% Bis–Tris gels, and stained with SilverQuest.

acidic fractions. With the wide pH range that FFE achieved, very acidic and basic proteins were able to be identified and characterized, although smaller in number than the intermediate fractions. The fractions (27–61) with the most intense pattern of bands on silver-stained SDS-PAGE were chosen for tryptic digestion and LC/MS analysis.

To assess the ability of the FFE focusing approach to achieve tight focusing of protein fractions we examined the entire set of identified proteins using a conservative criteria of at least two peptide identifications (hits) in the same fraction (see later). A potential concern with the use of IEF-based prefractionation methods is the potential for high abundance proteins or proteins with significance charge heterogeneity to focus in several fractions. Such a result would have the effect of reducing the dynamic range of the proteomic measurement either through an abundant protein obscuring low level proteins or through a heterogeneous protein being present at low levels in many fractions and thus not being detected. In this study, however, we observed the majority of proteins were present in only one fraction (see Table 1). Some proteins were indeed present in several, mostly consecutive, fractions. In general, these proteins are highly abundant in the cell line, such as actins, α-enolase, β-enolase, and glyceraldehyde 3-phosphate dehydrogenase. These observations could either be due to the process of focusing high concentration proteins generating a broader band, or the presence of post-translational modifications (PTMs) increasing the pI distribution of a given protein.

For each fraction, the theoretical pIs of the proteins identified in a given fraction were averaged and compared with the measured pH for that fraction. Fig. 2 shows the comparison of the average pI and pH value for each fraction. The mean difference between the average pI and measured pH value of a given fraction was −0.41 units. In denaturing FFE separation media, with presence of a significant concentration of detergent and reducing reagents, the pI values of proteins can vary significantly (in some cases, up to 1.4 pH unit), from pI in native media [22], which is consistent with our data. Also, unknown post-translational modifications such as phosphorylation, glycosylation, and proteolysis, may cause deviations

Table 1
Protein distribution after FFE focusing—number of proteins present in different fractions

Number of fractions (*N*)	Number of proteins (≥2 hits in one fraction)
1	151
2	41
3	21
4	16
5	11
6	7
7	5
8	1
9	1
Total	254

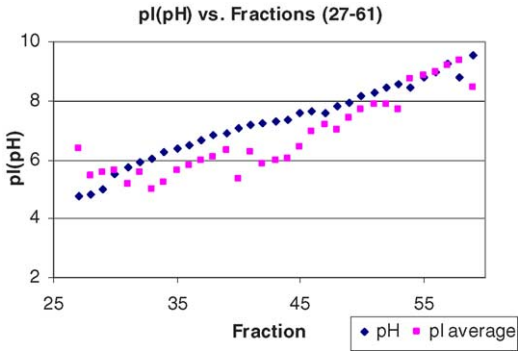

Fig. 2. A comparison of average pI values and measured pH for all 35 FFE fractions analyzed by LC/MS. The pI average was obtained by averaging the theoretical pIs of the identified proteins (≥2 hits in each fraction). The pH was measured with a pH probe with a special small tip.

in the measured pI values, although the observed pI distribution for the highly ranked proteins in each fraction was quite narrow. Fig. 3A shows the pI distributions for fraction 37 as a typical example.

3.2. Assessment of the criteria used for in-silico generation of peptide sequences from databases

The information contained in both genomic and proteomic databases can be used to generate peptide candidates using different enzyme specificities such as trypsin cleavage (Arg or Lys on both ends), partial tryptic cleavage (Arg or Lys on one end) and no-enzyme cleavage (no trypsin specificity). In this situation, one trades off the desire to produce as many candidate sequences for correlation with MS/MS data and the desire to minimize the number of false-positives.

There can be several reasons for the generation of peptides by other than classical trypsin cleavages. One cause is based on the sequence of a given protein, such as the presence of a C-terminus that does not contain Arg or Lys, or labile interior sites such as Asp-Pro, which can give acid or base catalyzed cleavage. Trypsin also has chymotryptic-like specificity, which generates peptides from cleavage at hydrophobic residues and studies have shown that the proportion of this type of cleavage can increase with high enzyme to substrate ratios. This situation will occur in complex protein mixtures where an adequate amount of trypsin for proteolysis of high level proteins will represent a large excess for low level proteins. An example of this altered specificity was observed in a study of the in vivo metabolism of human growth hormone (hGH) in the rat where a significant increase in chymotryptic-like cleavages was observed for hGH when present in low levels in plasma samples [23].

In view of the importance of the issue of which is the best list of candidate peptides for interrogating the MS/MS data, we investigated the number of false-positive identifications (as measured by aberrant pI values) with the use of different enzyme specificities. Fig. 3 shows the results for the analysis of one fraction, 37, which was typical of what was observed

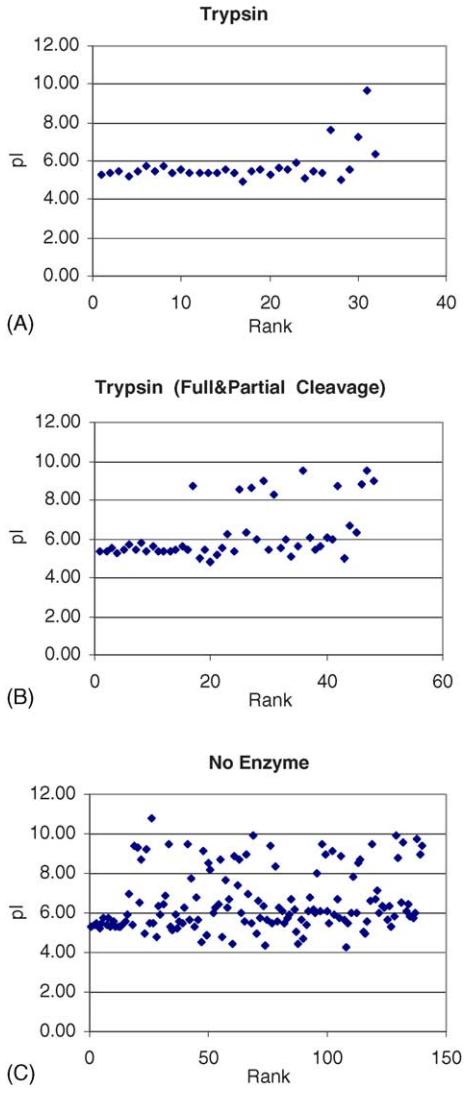

Fig. 3. The distributions of pI values of putative protein identifications made with different enzyme specificities in fraction 37 from the FFE prefactionation. The following search parameters were used: (A) trypsin; (B) full and partial tryptic cleavages; (C) no enzyme. Xcorr: 1.5, 2.0, 2.5.

in the other fractions (data not shown). It is clear that trypsin (Part A) gives a relatively small number of identifications (32) with a tight range of pI values except for those identifications with a low SEQUEST rank. The SEQUEST rank provides an approximate measure of relative protein concentration based on several criteria such as number of peptides and quality of MS data [24–26]. There is a significant increase in both the number of identifications (48) and in the scatter of the pI values with the incorporation of partial tryptic cleavages (Part B). Finally in Part C we see that the use of no enzyme specificity gives a large number of protein identifications (140) but at the expense of a large number of false identifications (as judged by the scatter of the pI values).

The data shown in Fig. 3 was achieved with Xcorr thresholds set at a moderate stringency level (1.5, 2.0, and 2.5 for 1+, 2+, and 3+, respectively) and resulted in 32, 48, and 140 putative protein identifications, depending on the enzyme specificity used. In order to optimize the number of correct protein identifications we then explored the most generous identifications made with fraction 37 that is using the no-enzyme cleavage specificity. We then attempted to reduce the number of false-positive identifications by the use of more stringent Xcorr thresholds. One can see that increasing the Xcorr thresholds from 1.5, 2.0, 2.5 to 1.8, 2.5, 3.0 and finally to 1.8, 2.5, 3.5 resulted in progressively fewer identifications (140 versus 26 versus 21). Thus, the results shown in Fig. 3 indicate that use of higher Xcorr thresholds gave a more consistent pI distribution.

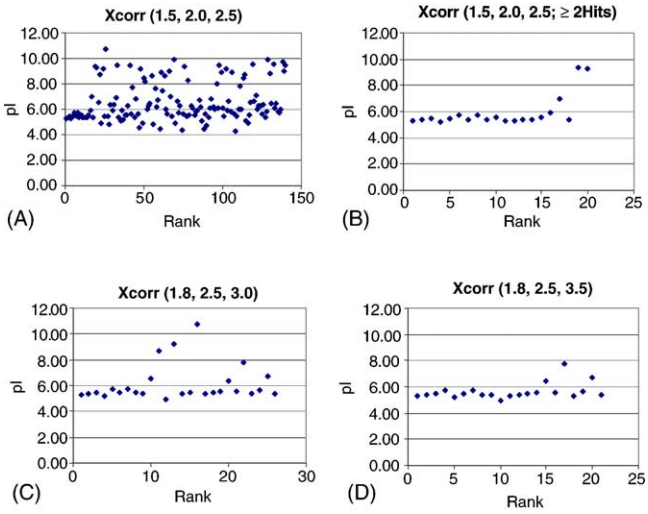

Fig. 4. The pI distribution of proteins identified in FFE fraction 37 using different Xcorr thresholds.

It has been noted that protein identifications based upon two or more peptide "hits" can be made with a higher confidence than the single "hits" used above [24,25]. Fig. 4B shows the pI distribution obtained for fraction 37 using the lowest of the three Xcorr thresholds (1.5, 2.0, 2.5), and no-enzyme cleavage specificity, but combined with the requirement for at least two peptide identifications. In this case twenty proteins were identified within a very narrow pI distribution. These results suggested that the use of Xcorr (1.5, 2.0, 2.5) with a requirement for two or more peptide identifications for each protein gave as good a correlation with pI as the most conservative setting of Xcorr (1.8, 2.5, 3.5) (Fig. 4D) and in the overall study resulted in a greater number of protein identifications (see later). Also, by comparing the protein identifications in Fig. 4B (20 proteins) and 4D (21 proteins), 15 proteins are in common with these two restrictive criteria. In addition, the three pI outliners shown in Fig. 4D with values above 6.0, acetyl-CoA carboxylase 2, bifunctional aminoacyl-tRNA synthetase, and hypothetical protein KIAA0918 were eliminated if one uses a more stringent search criteria, such as Xcorr (1.8, 2.5, 3.5), which further corroborated the value of pI measurements.

Table 2
Protein identifications in K562/CR3 cell line with different SEQUEST criteria

SEQUEST Xcorr score[a] (no enzyme)	35 FFE fractions	Cell lysate	
	135 min-run	135 min-run	270 min-run
(1.5, 2.0, 2.5)	(736) 2078	(31) 70	(70) 319
(1.8, 2.5, 3.0)	(169) 387	(17) 39	(41) 96
(1.8, 2.5, 3.5)	(155) 319	(17) 36	(41) 92

[a] The number in the bracket denotes the number of protein with two or more hits in the overall analysis. The Xcorr score is set as for +1, +2, and +3 charge state.

The trends shown here were consistently observed in other fractions and show that the use of pI to assess the number of false-positive identifications is useful in determining the appropriate strategy for processing large amounts of MS data. One must expect some scatter in pI values due to the presence of post-translational modifications or proteolytic processing of the intact protein, but the overall trends were clear in this study.

3.3. Assessment the use of pI values to increase the total number of protein identifications

After the FFE step, proteins in each fraction were identified by LC/MS analysis of the corresponding tryptic digest with the SEQUEST database search criteria described above. The number and quality of protein identifications, with and without FFE fractionation, were compared to assess the power of a prefractionation step in the proteomic analysis of a complex sample. Table 2 shows that indeed FFE prefractionation results in a significant improvement in protein identifications (736 by FFE-LC/MS versus 31 by standard LC/MS), with Xcorr thresholds (1.5, 2.0, 2.5) and a requirement for two or more peptide identifications). The low number of protein identification for the analysis without prefractionation demonstrates the problems of MS analysis of highly complex mixtures where overwhelming peptide complexity severely limits the power of the LC/MS approach. This conclusion was consistent with the observation of a significant increase in identifications, 30–70 proteins, with a longer LC gradient but without the FFE prefractionation (135–270 min).

As was demonstrated with a single fraction, we found that the total number of proteins identified in the entire study can vary significantly based on the use of different search criteria. In Table 2, the listing of a protein identification with two or more hits means that the protein has at least two unique peptide identifications in one or more fractions from the FFE separation. With Xcorr thresholds (1.5, 2.0, 2.5), after FFE separation, total number of protein IDs was 2078, of which 736 were found with two or more two peptide identifications. With considerably more stringent Xcorr thresholds (1.8, 2.5, 3.5) the number of protein identifications is reduced significantly to 319, among which 155 proteins were identified with two or more hits (Table 3). This table can

Table 3
Proteins identified in the K562/CR3 cell line using conservative criteria for the identification

	Protein identifications (Xcorr: 1.8, 2.5, 3.5; ≥2 hits)
#1	Actin, cytoplasmic 2 (gamma-actin)
#2	Fructose-bisphosphate aldolase A
#3	Glyceraldehyde 3-phosphate dehydrogenase, L
#4	Actin, alpha skeletal muscle
#5	Heat shock protein HSP 90-beta (HSP 84)[b]
#6	Peptidyl-prolyl *cis–trans* isomerase A
#7	Triosephosphate isomerase (TIM)
#8	Heat shock 70 kDa protein 1 (HSP 70.1)[b]
#9	Alpha enolase
#10	Transgelin 2 (SM22-alpha homolog)
#11	Keratin, type II cytoskeletal 1
#12	Beta enolase
#13	Alpha enolase, lung specific
#14	Profilin I
#15	Phosphoglycerate kinase 1
#16	Malate dehydrogenase, mitochondrial
#17	Heterogeneous nuclear ribonucleoproteins A2
#18	14-3-3 Protein zeta/delta[b]
#19	Hemoglobin gamma-A and gamma-G chains[a]
#20	L-Lactate dehydrogenase C chain (LDH-C)
#21	Heat shock 70 kDa protein 7[b]
#22	Pyruvate kinase, M1 isozyme
#23	Hemoglobin epsilon chain[a]
#24	Elongation factor 1-alpha 2 (EF-1-alpha-2)
#25	Tubulin beta-4 chain (tubulin beta-III)
#26	Elongation factor 2 (EF-2)
#27	54 kDa nuclear RNA- and DNA-binding protein
#28	Heat shock-related 70 kDa protein 2[b]
#29	Heat shock cognate 71 kDa protein[b]
#30	14-3-3 Protein gamma[b]
#31	Tubulin beta-5 chain
#32	T-complex protein 1, beta subunit
#33	Glutathione *S*-transferase P (GST class-PI)
#34	Fascin (singed-like protein)
#35	Heat shock protein HSP 90-alpha (HSP 86)[b]
#36	Flavin reductase (FR)
#37	Endoplasmin precursor
#38	Chloride intracellular channel protein 1
#39	Poly(RC)-binding protein 2 (alpha-CP2)
#40	Ubiquitin-conjugating enzyme E2-18 kDa UBC
#41	Tumor necrosis factor type 1 receptor[b]
#42	Transketolase (TK)
#43	Elongation factor 1-alpha 1 (EF-1-alpha-1)
#44	Macrophage migration inhibitory factor
#45	Heterogeneous nuclear ribonucleoprotein A1
#46	Fructose-bisphosphate aldolase C
#47	Nucleophosmin (NPM)
#48	10 kDa heat shock protein, mitochondrial
#49	Nuclear autoantigenic sperm protein (NASP)
#50	14-3-3 Protein epsilon[b]
#51	14-3-3 Protein beta/alpha[b]
#52	Antioxidant protein 2
#53	Placental ribonuclease inhibitor
#54	C-myc promoter-binding protein (MPB-1)[b]
#55	60 kDa heat shock protein, mitochondrial[b]
#56	14-3-3 Protein tau (14-3-3 protein theta)[b]
#57	T-complex protein 1, epsilon subunit
#58	40S ribosomal protein SA (P40)
#59	ATP synthase alpha chain, mitochondrial pre
#60	Hemoglobin zeta chain[a]
#61	40S ribosomal protein S21

Table 3 (*Continued*)

	Protein identifications (Xcorr: 1.8, 2.5, 3.5; ≥2 hits)
#62	Tubulin alpha-4 chain
#63	T-complex protein 1, ETA subunit
#64	Calreticulin precursor (CRP55) (calregulin)
#65	Putative nucleoside diphosphate kinase
#66	Keratin, type i cytoskeletal 14
#67	Proliferating cell nuclear antigen (PCNA)
#68	Protein disulfide isomerase A3[b]
#69	FK506-binding protein (FKBP-12)
#70	Y box binding protein-1
#71	Tropomyosin, cytoskeletal type (TM30-NM)
#72	Tubulin alpha-1 chain, brain-specific
#73	60S acidic ribosomal protein P2
#74	Superoxide dismutase [CU-ZN]
#75	Proto-oncogene tyrosine-protein kinase ABL1[b]
#76	Clathrin heavy chain 1 (CLH-17) (KIAA0034)
#77	Cofilin, non-muscle isoform
#78	Complement component 1
#79	Heterogenous nuclear ribonucleoprotein
#80	Tubulin beta-2 chain
#81	Stathmin (phosphoprotein P19) (PP19)
#82	L-Lactate dehydrogenase b chain (LDH-B)
#83	Moesin (membrane-organizing extension spike
#84	Splicing factor, proline-and glutamine-rich
#85	Protein disulfide isomerase[b]
#86	T-complex protein 1, gamma subunit
#87	Serum albumin[a]
#88	Hepatoma-derived growth factor (HDGF)
#89	Hemoglobin alpha chain[a]
#90	Tropomyosin alpha chain, smooth muscle
#91	Annexin II (LIPOCORTIN II)[b]
#92	Peroxiredoxin 2 (thioredoxin peroxidase 1)[b]
#93	Set protein (HLA-DR associated protein II)
#94	T-complex protein 1, theta subunit
#95	Peroxisome proliferator activated receptor
#96	Keratin, type I cytoskeletal 9
#97	60S ribosomal protein L22
#98	14-3-3 Protein eta (protein AS1)[b]
#99	Heterogeneous nuclear ribonucleoprotein M
#100	Tubulin beta-1 chain
#101	Proteasome subunit alpha type 5
#102	Collagen-binding protein 2 precursor
#103	GTP-binding nuclear protein ran (TC4)
#104	RRP5 protein homolog
#105	Transaldolase
#106	Hypothetical protein KIAA0153
#107	Inositol 1,4,5-trisphosphate receptor type
#108	Onconeural ventral antigen-1 (NOVA-1)
#109	Heterogeneous nuclear ribonucleoprotein A3
#110	T-cell acute lymphocytic leukemia-1 protein
#111	Glycyl-trna synthetase
#112	Histone H2B.S (H2B/S)[b]
#113	T-complex protein 1, delta subunit
#114	Phosphatidylethanolamine-binding protein
#115	Tryptophanyl-trna synthetase
#116	Nucleolin (protein C23)
#117	Elongation factor 1-delta (EF-1-delta)
#118	T-complex protein 1, alpha subunit
#119	Creatine kinase, B chain (B-CK)
#120	Eukaryotic initiation factor 4A-II
#121	Adenosylhomocysteinase
#122	Fragile x mental retardation syndrome
#123	Galectin-1[b]
#124	Transitional endoplasmic reticulum atpase

Table 3 (*Continued*)

	Protein identifications (Xcorr: 1.8, 2.5, 3.5; ≥2 hits)
#125	Guanine nucleotide-binding protein beta subunit
#126	Plasminogen precursor
#127	Lupus LA protein
#128	Keratin, type II cytoskeletal 6C
#129	Adenylyl cyclase-associated protein 1
#130	Keratin, type II cytoskeletal 6F
#131	Voltage-dependent anion-selective channel P
#132	ATP-dependent DNA helicase II, 80 kDa
#133	Poly(RC)-binding protein 1 (alpha-CP1)
#134	Initiation factor 5A (EIF-5A) (EIF-4D)
#135	T-complex protein 1, zeta subunit
#136	Baculoviral iap repeat-containing protein
#137	Vimentin
#138	Enhancer of rudimentary homolog
#139	Heterogeneous nuclear ribonucleoprotein K
#140	Matrin 3
#141	Stress-induced-phosphoprotein 1 (STI1)
#142	Ubiquitin-activating enzyme E1
#143	L-Lactate dehydrogenase A chain (LDH-A)
#144	Arginase 1 (liver-type arginase)
#145	Leukocyte tyrosine kinase receptor
#146	Protein bap28
#147	Cofilin, muscle isoform (cofilin 2)
#148	DNA polymerase alpha 70 kDa subunit
#149	ATP-binding cassette, sub-family A
#150	Myosin light chain alkali, GT-1 isoform
#151	Proactivator polypeptide
#152	Solute carrier family 2, facilitated glucose (GTR4)
#153	Solute carrier family 2, facilitated glucose (GTR1)
#154	Mage-like protein 2
#155	Ran-binding protein 17[b]

Database searching criteria: Xcorr: (1.8, 2.5, 3.5) for charge states +1, +2, +3; ≥2 hits.

[a] Possible contaminant from growth medium.

[b] The underlined proteins are of potential significance in the transformed cell line.

serve as a starting point in the description of the proteome of this cell line and future studies will be directed at increasing the number of identifications by improvements in the LC/MS approach, as well as the use of better bioinformatics tools.

Table 4 evaluates the success of the FFE prefractionation step by comparing the quality of sequence data for the top 20 proteins (as ranked by SEQUEST) with and without the prefractionation step. Several potentially significant proteins were identified with high sequence coverage, such as transgelin (71% coverage) and ribonucleoprotein A2 (43%) after FEE prefractionation, but were not detected at all in the direct LC/MS approach. In general, both the quantity and quality of the protein IDs were greatly improved with FFE prefractionation (see Table 4). In addition to reducing the complexity of peptide mixture at a given time point (which reduces ion suppression), FFE can enrich many low abundance proteins into fractions separate from high abundance proteins. For example, two high level proteins, actin, and α,β-enolase were largely focused into the pI range of 5–6 and 8–9, respectively.

Table 4
Sequence coverage achieved for the top 20 ranking proteins (Xcorr: 1.5, 2.0, 2.5) with and without FFE prefractionation

Top 20 protein identification		FFE-LC/MS (135 min)		LC/MS (135 min)		LC/MS (270 min)	
		No. of scan	Sequence coverage (/%)	No. of scan	Sequence coverage (/%)	No. of scan	Sequence coverage (/%)
1	Fructose-bisphosphate aldolase A	287	70.60	15	23.35	54	38.13
2	Actin, cytoplasmic 2	274	38.40	12	24.53	19	29.95
3	Glyceraldehyde 3-phosphate dehydrogenase	218	60.90	6	17.61	12	38.81
4	Actin, alpha skeletal muscle	137	28.91	9	11.41	10	14.32
5	Heat shock protein HSP 90-beta	118	43.65	10	16.02	30	29.28
6	Peptidyl-prolyl *cis–trans* isomerase A	115	74.55	2	8.48	6	31.52
7	Heat shock 70 kDa protein 1	82	29.80	3	3.74	9	10.45
8	Triosephosphate isomerase	78	63.05	1	8.03	4	12.05
9	Alpha enolase	74	40.09	4	11.06	8	14.98
10	Keratin, type II cytoskeletal 1	60	40.22	2	1.86	[a]	[a]
11	Transgelin 2	55	71.36	[a]	[a]	3	12.56
12	Profilin I	54	60.00	1	10.00	2	16.43
13	Heterogeneous nuclear 1ribonucleoproteins A2	53	43.06	[a]	[a]	1	7.08
14	Malate dehyfrogenase	51	46.45	[a]	[a]	1	6.51
15	Alpha enolase, lung specific	51	13.10	[a]	[a]	8	14.98
16	Beta enolase	49	16.13	[a]	[a]	8	12.67
17	Elongation factor 1-alpha 2	49	30.45	5	4.10	15	10.37
18	Tubulin beta-4 chain	43	18.89	1	2.22	8	8.67
19	Pytuvate kinase, M1 isozyme	42	27.87	[a]	[a]	[a]	[a]
20	Phosphoglycerate kinase 1	40	26.56	[a]	[a]	3	14.59

[a] Not available.

3.4. pI as a criterion for the determination of false-positive MS-based protein identifications

After FFE separation, proteins with similar pI values were focused into one or adjacent fractions. By studying the theoretical pI distribution of proteins within one fraction (#37), it was found that most high-ranking protein identifications had a similar pI (Fig. 4). Proteins in Fig. 4B and D were further assessed by pI correlation. In Fig. 4B, 17 out of 20 protein IDs ($\geq$2 hits) had very similar pI, (pI average = 5.48; R.S.D. = 3.28%). The outliers, nuclear pore complex protein NUP214, a probable G protein-coupled receptor GPR32 (GPRW) and myosin XV (MY15), with SEQUEST rank of 17, 19, and 20, had pI values well outside the range of 5.1–5.7 (6.99, 9.29, and 9.35). The identification of the proteins GPRW and MY 15 was confirmed in three replicate studies, with the proteins always detected with the same peptides (either one or two hits). In this situation, it is likely that the significant pI shifts observed with these proteins is due to post-translational modifications (including proteolysis). In addition, manual inspection of the MS/MS spectra for these identifications confirmed the SEQUEST assignment (see Fig. 5). The situation was quite different in the case of the third assignment where the identification of NUP214 was deemed to be a false-positive. The identification of NUP214 was made with two peptides but only in one run and the corresponding MS/MS spectra were of poor quality. As an example, Fig. 5 shows some typical MS/MS spectra used in the assignment of these three proteins and it can be readily seen that the SEQUEST algorithm at less conservative settings has difficulty in distinguishing noisy (C) from correct spectra (A, B).

In summary, this study shows the potential of pI as an experimentally determined parameter independent of MS-based identifications to support the protein identification and help eliminate false-positives.

3.5. Reproducibility study

The extraction procedure for a given FFE fraction was tested by performing three individual extractions of fraction 37. Fraction 37 was shown on 1D gel electrophoresis to be of moderate complexity with approximately 15 bands of reasonable intensity. The extractions were digested and analyzed consecutively by LC/MS using the standard protocol of this study. With Xcorr thresholds set at 1.5, 2.0, 2.5, the triplicate analysis identified a cumulative total of 53 proteins with at least two hits, of which 17 (32.1%) proteins were identified in all triplicate runs and 37 (69.8%) were identified in replicate runs. This reproducibility is significantly less than observed in consecutive replicate analysis of the same extract (fraction 37) where 80% of protein identifications with two or more hits were repeatable. An explanation for this higher variability may be related to the storage of the FFE samples in a 96-well plate and the presence of precipitate in each well, so that even after careful workup one could expect some variation in sampling of a given fraction. It can be expected that the trypsin digestion will be sensitive to such variations, as well as differences in the number of proteins extracted from the well. Another source of variability in the analyses performed in

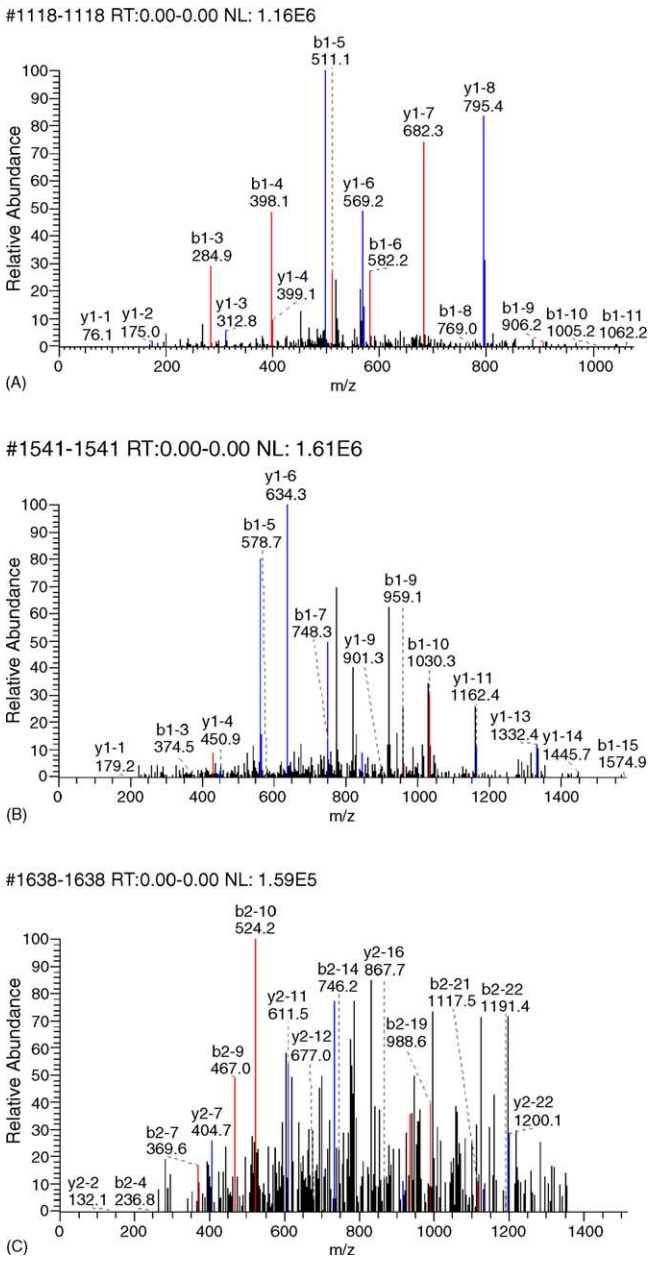

Fig. 5. The peptide MS/MS spectrums used in the identification of the following proteins: MY15, GPRW and NUP214. (A) MY15, peptide: GVQLLAVSHVG and correlated to residues 2786–2796 in the protein. The following criteria was used, Xcorr: 2.541, ΔCn: 0.230; (B) GPRW, peptide: FLLGFLFPLAIIFTC, residues 221–235 Xcorr: 2.437, ΔCn: 0.230; (C) NUP214, peptide: EKTLPPAPVLMLLSTDGVLCPF residues 380–401 Xcorr: 2.503, ΔCn: 0.025.

both these studies has been reported by others that use "data dependent acquisition" of fragmentation data [11,25]. The detection of low level peptides is compromised in a flowing system by the constraint of the ion trap mass spectrometer to have time to fragment a given low level peptide in a complex mixture. The authors recommended performing each analysis in triplicate and achieved reproducibility similar to what was found in this study [27].

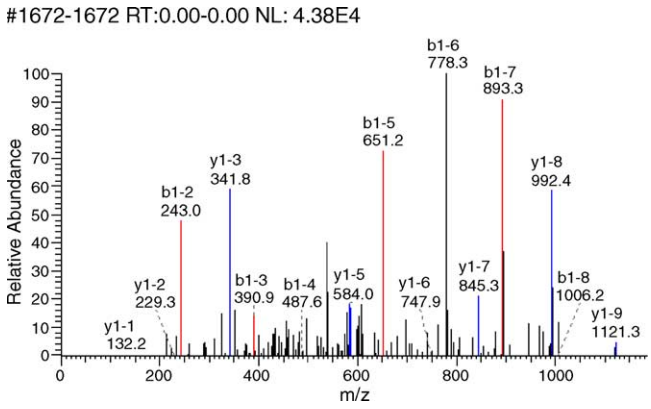

Fig. 6. The peptide MS/MS spectrum used in their identification of Integrin alpha-V. The peptide sequence was: IEFPYKNLPI and correlated to residues 959–968 in the protein. The following criteria was used, Xcorr: 2.73, ΔCn: 0.08.

3.6. Interesting proteins

While structural and housekeeping proteins are present in a cell with a high copy number, many functionally important proteins are of low abundance. With the deep dynamic range that FFE-LC/MS achieved, we were able to characterize a significant number of proteins, which can give insights about specific metabolic processes and pathways in a specific cell line. For example, this K562 cell line has been transfected to express the β2-integrin CD11b/CD18. Integrins are important cell adhesion receptor proteins, which are involved in cell-extracellular matrix and cell–cell interactions [28]. The β2-integrins are a family of leukocyte adhesion molecules consisting of three surface membrane heterodimeric glycoproteins CD11a/CD18, CD11b/CD18, and CD11c/CD18. In this subfamily of integrins, a same β subunit CD18 is shared by three α subunits CD11a, CD11b, and CD11c [29]. In this study, four integrins were identified with two or more hits using Xcorr threshold (1.5, 2.0, 2.5) and with pI values within expected values. Among these proteins, one protein, integrin alpha-V also met the most stringent criteria of Xcorr threshold (1.8, 2.5, 3.5). In each case, the MS/MS spectra were manually checked for clear b-ion and y-ion series and an example spectrum is shown in Fig. 6. In addition, integrin alpha-L (CD11a), alpha-X (CD11c), and alpha-M (CD11b), were all identified with one peptide with an Xcorr threshold (1.5, 2.0, 2.5) and with pI within the expected range and good quality MS/MS spectra.

A research interest of the laboratory is in the role of novel signaling molecules such as semaphorins in the transformation of normal into tumor cells. Semaphorins were originally identified as axon guidance factors functioning in the nervous system, recent studies have uncovered some immunological functions as well [30]. In this study, two semaphorins (semaphorin 3F, semaphorin 4F) were identified with two or more peptides with Xcorr (1.5, 2.0, 2.5) and appropriate pI values.

4. Conclusions

In this FFE-LC/MS study, we achieved a deep dynamic range in the measurement of the protein profile for the human chronic myelogenous leukemia cell line, K562/CR3. By comparing the LC/MS results, before and after FFE fractionation, it is clear that prefractionation made a significant improvement in protein identifications, both in quality and quantity. In this study, we found pI could be used as an additional criterion for protein identification in addition to MS/MS data and was able to reduce the rate of false-positive protein assignments. The study characterized a number of proteins of interest in the study of the cellular response to transformation and examples described here include integrins and semphorins.

Acknowledgments

The authors wish to thank ThermoFinnigan for instrument support and Thermo Keystone Hypersil for columns. The authors also wish to thank Dr. Billy Wu, Dr. Jo Tsai, and Mr. Haven Baker for helpful discussion.

References

[1] N.L. Anderson, N.G. Anderson, Mol. Cell Proteomics 1 (2002) 845.
[2] R.M. Birch, C. O'Byrne, I.R. Booth, P. Cash, Proteomics 3 (2003) 764.
[3] M.G. Harrington, J.A. Coffman, F.J. Calzone, L.E. Hood, R.J. Britten, E.H. Davidson, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 6252.
[4] V. Badock, U. Steinhusen, K. Bommert, A. Otto, Electrophoresis 22 (2001) 2856.
[5] P. Davidsson, S. Folkesson, M. Christiansson, M. Lindbjer, B. Dellheden, K. Blennow, A. Westman-Brinkmalm, Rapid Commun. Mass Spectrom. 16 (2002) 2083.
[6] A. Gorg, G. Boguth, A. Kopf, G. Reil, H. Parlar, W. Weiss, Proteomics 2 (2002) 1652.
[7] T.Q. Shang, J.M. Ginter, M.V. Johnston, B.S. Larsen, C.N. McEwen, Electrophoresis 24 (2003) 2359.
[8] X. Zuo, D.W. Speicher, Proteomics 2 (2002) 58.
[9] F. Giorgianni, D.M. Desiderio, S. Beranova-Giorgianni, Electrophoresis 24 (2003) 253.
[10] E. Klein, J.B. Klein, V. Thongboonkerd, Contrib. Nephrol. 141 (2004) 25.
[11] W.S. Hancock, S.L. Wu, P. Shieh, Proteomics 2 (2002) 352.
[12] W.H. McDonald, J.R. Yates 3rd, Dis. Markers 18 (2002) 99.
[13] L. Krivankova, P. Bocek, Electrophoresis 19 (1998) 1064.
[14] H. Mohr, A. Volkl, Electrophoresis 23 (2002) 2130.
[15] A. Volkl, H. Mohr, G. Weber, H.D. Fahimi, Electrophoresis 18 (1997) 774.
[16] A. Volkl, H. Mohr, G. Weber, H.D. Fahimi, Electrophoresis 19 (1998) 1140.
[17] A. Volkl, H. Mohr, H.D. Fahimi, J. Histochem. Cytochem. 47 (1999) 1111.
[18] W. Hymer, K. Kirshnan, W. Kraemer, J. Welsch, W. Lanham, Electrophoresis 21 (2000) 311.
[19] H. Zischka, G. Weber, P.J. Weber, A. Posch, R.J. Braun, D. Buhringer, U. Schneider, M. Nissum, T. Meitinger, M. Ueffing, C. Eckerskorn, Proteomics 3 (2003) 906.
[20] P. Hoffmann, H. Ji, R.L. Moritz, L.M. Connolly, D.F. Frecklington, M.J. Layton, J.S. Eddes, R.J. Simpson, Proteomics 1 (2001) 807.
[21] H.P. Koeffler, D.W. Golde, Blood 56 (1980) 344.
[22] G. Candiano, G.M. Ghiggeri, G. Delfino, C. Queirolo, G. Vecchio, J. Chromatogr. 374 (1986) 279.
[23] J.E. Battersby, V.R. Mukku, R.G. Clark, W.S. Hancock, Anal. Chem. 67 (1995) 447.
[24] D.A. Wolters, M.P. Washburn, J.R. Yates 3rd, Anal. Chem. 73 (2001) 5683.
[25] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates Jr., Nat. Biotechnol. 17 (1999) 676.
[26] Z. Yang, W.S. Hancock, J. Chromatogr. A (2004), in press.
[27] B.R. Wenner, M.A. Lovell, B.C. Lynn, J. Proteome Res. 3 (2004) 97.
[28] J.P. Xiong, T. Stehle, S.L. Goodman, M.A. Arnaout, R.R. Ingalls, R.L. Delude, S. Flaherty, R. Savedra Jr., D.T. Golenbock, T.A. Chatila, R.S. Geha, D.R. Freyer, M.L. Morganroth, C.E. Rogers, R.F. Todd 3rd, J. Thromb. Haemost. 1 (2003) 1642.
[29] M.A. Arnaout, T.A. Chatila, R.S. Geha, D.R. Freyer, M.L. Morganroth, C.E. Rogers, R.F. Todd 3rd, Blood 75 (1990) 1037.
[30] A. Kumanogoh, H. Kikutani, Adv. Immunol. 81 (2003) 173.