Journal of Chromatography B, 875 (2008) 478–486

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Proteomic analysis optimization: Selective protein sample on-column retention in reverse-phase liquid chromatography☆

Witold M. Winnik*, Pedro A. Ortiz

National Health and Environmental Effects Research Laboratory, Environmental Carcinogenesis Division, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, United States

ARTICLE INFO

Article history:
Received 5 August 2008
Accepted 26 September 2008
Available online 2 October 2008

Keywords:
Proteomic profiling
Microsomal proteins
Cytochrome P450
Reverse-phase liquid chromatography
Fractionation
Sample injection
Flow-through
Selectivity
DMSO
1D gel electrophoresis
SDS PAGE
Nanospray LC–MS/MS
Mass spectrometry

ABSTRACT

In an effort to optimize reverse-phase liquid chromatography (RPLC) for proteomics, we studied the impact of composition of the sample injection solution on protein on-column selection and retention. All the proteins studied were retained on-column when injections were made in 50% formic acid, 0.1% TFA or 8.3 M urea. When formic acid was increased to 80%, the superoxide dismutase standard (MW 26,159) and 58 mouse microsomal proteins that possessed low-range molecular weights, high pIs or basic amino acid clusters were non-retained, resulting in retention selectivity during sample injection. Introducing to the 80% formic acid injection solution an organic solvent such as acetonitrile or acetonitrile–DMSO induced further retention selectivity, and increasing levels of organic solvents reduced on-column retention. The proteome was split into the proteins that were retained on-column which eluted at higher retention times (RTs), vs the proteins that collected in the injection flow-through which normally eluted at lower RTs. This protein selectivity was confirmed after fraction collection, 1D-GE and nano-LC–MS/MS. The significance of this procedure is that it can be exploited for fast extraction of small basic proteins from the bulk of the proteome and for on-column enrichment of hydrophobic proteins.

Published by Elsevier B.V.

1. Introduction

The usefulness of HPLC-based analytical methods in protein analysis [1–3] has benefited from several optimizations of pre-HPLC sample preparation techniques. As an example, the fractionation of membrane proteins by hydrophilic interaction chromatography (HILIC) was preceded by solvent extraction of bovine mitochondria where buffered mixtures of organic solvents were used for selective extraction of membrane proteins prior to chromatography [4]. Another example was the use of immunodepletion to improve otherwise poor LC resolution of blood plasma samples resulting from the deposition of abundant plasma proteins on a mRP-C18 column [5,6].

As for the column choice, RPLC separation of hydrophobic proteins has relied mainly on standard C-4 LC columns [7]. Subsequently, the polymeric large-pore LC materials (such as used in monolithic LC columns) were introduced to provide superior resolution and recoveries of proteins and peptides [8,9]. In addition, complex protein separations have been performed in multiple LC dimensions, using, for example a combination of RP column and an ion exchange column [1,10]. For instance, a pH-gradient ion exchange chromatofocusing technique was used to provide superior resolution in ion exchange separations of proteins [1,10] and peptides [11] in a 2D LC separation scheme using RP as a second LC dimension prior to mass spectral analysis. Mini-Rotafor™ was used for breast cancer protein separation in the first dimension followed by non-porous RP LC chromatography and protein identification by MS/MS [12]. It is well known that the multidimensional applications of separation in proteomics have the tendency to generate an abundance and a variety of solvents which are potentially incompatible with RP and MS. For example, solvents containing ion-pairing reagents, amines, and involatile salts may

☆ The research described in this article does not necessarily reflect U.S. Environmental Protection Agency policy and no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

* Corresponding author at: U.S. Environmental Protection Agency, Mail Code: B143-06, Environmental Carcinogenesis Division, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, United States. Tel.: +1 919 541 3509; fax: +1 919 685 3191.
E-mail address: winnik.witold@epa.gov (W.M. Winnik).

1570-0232/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.jchromb.2008.09.029

result in electrospray MS signal suppression, loss of sensitivity and MS instrument inlet clogging. Also, there is an inherent sample complexity in protein analysis and protein solubility which often imposes limitations on the solvent choice. Therefore due to these two facts, the introduction of protein samples onto LC columns deserves more attention and investigation.

In this study, we investigated the impact of solution composition of the sample injection volume on protein on-column selection and retention. The aim of the study was to prevent unintended sample losses, and to provide the means for selective elution and capture of different types of proteins while minimizing protein sample handling. In preceding studies, Mann and co-workers [5] introduced urinary proteins in 6 M urea injection volume, whereas Martosella et al. [13] introduced human brain membrane raft proteins in 80% formic acid, which is an effective protein solvent [14]. Both studies used the recently developed macroporous mRP-C18 column. This column was also utilized for an efficient fractionation of proteins in human serum and plasma at an elevated temperature (80 °C) providing optimal chromatographic resolution [6,15,16]. In the present work, samples were injected into an mRP-C18 column in a number of injection volumes: (1) 0.1% TFA (used for HPLC standards), (2) 50% formic acid, or (3) 80% formic acid, with different amount of DMSO, acetonitrile and water. Formic acid was chosen for preparation of LC injection solutions because it can solubilize a wide range of proteins, such as the membrane proteins used in this study among others. Also, it is volatile and compatible with MS, and has been shown to be compatible with the mRP-C18 column at high concentration [13]. The RP chromatography was followed by fraction collection, in-solution digestion, and nano-LC–MS/MS or gel electrophoresis (1D-GE), in-gel protein digestion and nano-LC–MS/MS.

2. Materials and methods

2.1. Materials and reagents

Unless stated otherwise, reagent grade chemicals and protein standards were purchased from Sigma–Aldrich. HPLC grade solvents were obtained from Honeywell Burdick and Jackson (Morristown, NJ). Formic acid, 99%, was purchased from Chem Service (West Chester, PA). Beckman Coulter (Fullerton, CA) Gold HPLC System with a diode-array UV detector and a fraction collector and a column heater was used for protein fractionations. LC–MS/MS was performed using a ThermoFisher Scientific (Waltham, MA) system including a LXQ linear ion trap, a dual LC Surveyor pump system, a Surveyor autosampler, and a nanospray ESI ion source. A Centrivap™ (Thermo Fisher Scientific) and a Labconco (Kansas City, MO) Freezone™ lyophilizer were used in sample preparation.

2.2. Microsomal protein precipitation

Proteins from mouse liver microsomes (Male CD-1, BD Gentest) were solubilized in a SDS buffer containing 5% SDS, 5% 2-mercaptoethanol, 10% Glycerol and 60 mM Tris (pH 6.8) by boiling for 5 min. Protein precipitation was performed using the chloroform/methanol method as previously described [17]. The precipitation conditions did not discriminate between the hydrophilic and hydrophobic proteins. Briefly, four volumes of methanol were added to the protein solution and vortexed followed by the addition of one volume of chloroform. The solution was vortexed and three volumes of deionized (DI) water were added. The mixture was vortexed again then centrifuged at 13,000 rpm (15,740 × *g*) for 5 min. The aqueous upper layer was discarded leaving the protein interface intact. Four volumes of ethanol were added, and the mixture was briefly vortexed and centrifuged at 13,000 rpm for 5 min. The supernatant was removed and the protein pellet was air dried. The protein content was determined at 1.0 mg (±0.1 mg)/protein pellet by the EZQ protein quantitation kit (Molecular Probes, Eugene, OR).

2.3. LC sample injection in urea

One milligram of a microsomal protein pellet was dissolved in 1 ml of 8.3 M urea (0.5 g/ml) containing 1% acetic acid (a modified procedure from [16]) and 10% acetonitrile. Urea was dissolved in water, acetic acid and acetonitrile and the solution volume was adjusted with DI water to 1 ml. An aliquot of 0.5 ml of this sample solution was injected on the LC column.

2.4. LC sample injection in 80% formic acid without organic solvents

One milligram of a microsomal protein pellet was dissolved in 1 ml of 80% formic acid and 20% water in a glass vial. An aliquot of 0.333 ml this sample solution was injected on the LC column. *Attention: proper laboratory safety precautions need to be taken while working with concentrated formic acid at high temperature and pressure.*

2.4.1. LC sample injection in 80% formic acid with organic solvents

One milligram of a microsomal protein pellet was placed in a glass vial and an aliquot of DMSO was added: 0.10 ml, 0.07 ml, or 0.05 ml, followed by an equal amount of acetonitrile: 0.10 ml, 0.07 ml, or 0.05 ml, respectively. The sample was vortexed and sonicated for 2 min and DI water was added up to 0.20 ml: 0 ml, 0.06 ml, or 0.10 ml, respectively. Finally, 0.80 ml of formic acid was added and the 2-min vortexing and sonication activity was repeated upon which the sample pellet was dissolved.

2.5. LC microsomal protein sample injection in 50% formic acid

One milligram of a microsomal protein pellet was dissolved in 90% (v/v) formic acid in a glass vial. One third of this solution was diluted with DI water until the concentration of formic acid was 50% and it was injected on the LC column.

2.6. LC sample injection of four protein standards

A stock solution containing 0.5 mg/ml of each of the following proteins: superoxide dismutase (SOD) from bovine liver, glutathione-*S*-transferases (GST) from equine liver, alpha glycerophosphate dehydrogenase (GPD) from rabbit muscle, and carbonic anhydrase (CarbAnh) from bovine erythrocytes, was prepared in 0.1% formic acid. Out of the stock the following working solutions were prepared: (1) an aliquot of 0.050 ml of the stock solution was added to 0.950 ml of 0.1% TFA solution in water; (2) an aliquot of 0.050 ml of the stock solution was added to 0.800 ml of formic acid and 150 ml of water; (3) an aliquot of 0.100 ml of the stock solution was added to 0.800 ml of formic acid. Then 1 mg of the microsomal protein pellet was added and the mixture was vortexed until it was completely dissolved. 0.300 ml of the solution was mixed with 0.240 ml of water to give the 1:1 formic acid:water ratio; (4) an aliquot of 0.100 ml of the stock solution was added to 0.800 ml of formic acid. Then 1 mg of the microsomal protein pellet was added and dissolved, and 0.300 ml of this solution was mixed with 0.037 ml of water to give the 8:2 formic acid:water ratio. These working solutions (1–4) were individually injected on the LC column to determine the effect of increasing concentrations of formic acid (0, 50 and 80%, Fig. 1).

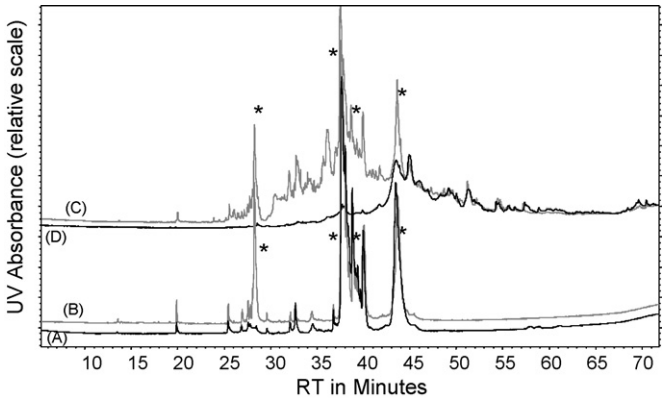

Fig. 1. Four LC chromatograms of the protein samples injected on the LC column in different sample solutions and chromatographed using the same LC solvent gradient conditions (Sections 2.6, 2.7): (A) 0.025 mg of GST, SOD, GPD, CarbAnh was injected on the LC column in 80% formic acid and 20% water (v/v); (B) 0.025 mg of GST, SOD, GPD, CarbAnh was injected on the LC column in 0.1% TFA and 99.9% water (v/v); (C) 0.333 mg of microsomal protein and 0.016 mg of GST, SOD, GPD, and CarbAnh was injected on the LC column in 50% formic acid and 50% water (v/v); (D) 0.333 mg (1 mg/ml) of microsomal protein and 0.016 mg of GST, SOD, GPD, and CarbAnh were injected on the LC column in 80% formic acid, 10% acetonitrile, and 10% water (v/v). Protein standard peaks are indicated by a "*" sign.

2.7. Protein fractionation by LC

The sample solution was manually injected at the flow rate of ca. 0.3 ml/min onto a macroporous 4.6 mm × 5 cm mRP-C18 high-recovery LC column (Agilent) equilibrated with solvent "A" (0.1% TFA, 2% acetonitrile, 98% water). The column temperature was maintained at 78 °C during the sample injection and throughout the LC run. A 3.5 ml volume of the injection flow-through, including the solvent "A" column wash, was collected. The following LC/UV (220 nm) method was used during the LC run: flow rate 0.75 ml/min, solvent "B" (0.08% TFA in acetonitrile) 0% for 5 min, linear gradient to 80% "B" at 60 min, linear gradient to 98% B at 65 min, 98% B isocratic to 75 min. Twenty-five 2-min fractions were collected starting at 25 min RT and stored at −80 °C awaiting processing. The frozen fractions were thawed at room temperature and 0.01 ml of 0.3 M ammonium bicarbonate was added. Acetonitrile was removed from the samples in a Centrivap™ evaporator and the remaining aqueous solutions were lyophilized to dryness and stored. To keep mRP LC column in good operating condition the column was not heated without maintaining a LC flow, and it was flushed with 30% formic acid in acetonitrile (3 ml) and isopropanol (3 ml) after each run.

2.8. Re-fractionation of the injection flow-through

Diluted sample injection flow-through from the 80% formic acid experimental runs (total volume of 3.5 ml of the flow-through and the LC gradient solvent "A" column wash) was mixed with formic acid to give the 1:1 water:formic acid ratio, and a 1.50 ml aliquot was re-fractionated by LC. Twenty-five fractions were collected and evaporated to dryness according to the procedure described in the preceding paragraph.

2.9. 1D gel SDS PAGE analysis of LC protein fractions and protein digestion

LC lyophilized fractions were re-suspended in 0.2 ml of a SDS sample buffer containing 50 mM Tris pH 6.8, 100 mM DTT, 1.66% SDS, 5% Glycerol, 0.002% bromophenol blue. A sample volume of 0.05 ml was loaded onto the SDS-PAGE gel (Protean II, BioRad, Hercules, CA). The gel was fixed, stained with Sypro Ruby (Invitrogen, Eugene, OR) and the proteins were detected by fluorescence on the Typhoon 9410 scanner (GE Healthcare, Piscataway, NJ).

In-gel digestion was performed as previously described [18]. Briefly, the gel bands were excised manually and sliced into 1–2 mm^3 size pieces. The gel pieces were dehydrated with acetonitrile, dried in a Centrivap™ evaporator and treated with 10 mM dithiothreitol (DTT) at room temperature for 30 min. The gel proteins were alkylated with 100 mM iodoacetamide at room temperature in the dark for 30 min. The gel pieces were dehydrated and then rehydrated with 100 mM ammonium bicarbonate, and then dehydrated and rehydrated again using enough volume of a 20 μg/μl trypsin solution in 50 mM ammonium bicarbonate only to swell the gel. The gel pieces were kept on ice for 30 min and then protein digestion was carried out at 37 °C overnight. The peptides were extracted by three sequential rinses with a 5% formic acid/50% acetonitrile solution. The extracted peptide solution volume was reduced to about 20 μl before the LC–MS/MS analysis.

2.10. In-solution digestion of the LC flow-through and the LC protein fractions

In-solution digestion was performed according to a modified procedure using acetonitrile as a solvent additive to improve solubility of hydrophobic proteins and peptides [19]. The lyophilized flow-through solution was reconstituted in 0.080 ml of 50 mM ammonium bicarbonate, 0.020 ml of acetonitrile, and 0.6 μg of a mass spectrometry grade trypsin stock solution (1 mg/ml). The solution was incubated overnight at 37 °C, and then an additional amount of 0.2 μg of trypsin was added before the mixture was incubated for additional 3 h. The digest was acidified with 0.004 ml of formic acid and 0.2 ml of 0.1% TFA in water before the solution was passed through the macro peptide desalting cartridge (Microm Bioresources Inc.). The peptides were eluted with 80% acetonitrile and 0.1% TFA in DI water. The eluant was reduced in volume in Speedvac (ThermoFisher) to ca 0.140 ml. 0.010 ml of this solution was injected in the nano-LC–MS/MS system. This experimental procedure was also applied to the fractionated protein samples requiring in-solution digestion and LC–MS/MS analysis.

2.11. Nanospray-ESI–MS/MS analysis

Before the nano-LC separation, peptides were introduced in 10 μl aliquots onto one of the two RP peptide micro-traps mounted in the 10-port switch valve [11] at the flow rate of 4 μl/min. The peptides were eluted in a sequential manner from the micro-traps using the second LC pump onto a nanospray column 5 μm BioBasic™ C18, 300 Å pore size, 75 μm × 10 cm, tip 15 μm (New Objective), at a flow rate of ca 250 nl/min. RP LC–MS/MS solvent "A" was 0.1% (v/v) formic acid in 98% water, 2% acetonitrile, and the RP solvent "B" was 98% acetonitrile, 2% water with 0.1% (v/v) formic acid. A linear gradient was applied to 50% solvent "B" at 45 min, 65% solvent "B" at 55 min, 80% "B" at 60 min, 90% "B" at 67 min, 90% "B" at 68 min, and 0% "B" at 72 min. Briefly, peptides were injected onto one of the two available peptide traps (10 min) and then let stand in wait with the negligible flow rate of a few nl/min while the preceding peptide sample was being eluted off the other micro-trap onto the nanospray RP column and was undergoing the separation. Then the solvent flow through the traps was toggled and the loading/separation/switching cycle repeated until all samples were analyzed. A "Most Intense Five" scan sequence was set up in Xcalibur™, including a survey MS scan from 300 to 1800 m/z followed by five data dependent MS/MS scans using the collision energy of 35%, the isolation width of 2.0 m/z, the activation Q parameter of 0.25, and the activation time of 30 ms. A Dynamic

Exclusion™ allowed one MS/MS scan per precursor ion followed by a 30 s exclusion period. The limit of detection for the MS analysis was ca. 100 attomol for an angiotensin I peptide standard.

The data set was processed in the Bioworks 3.3™ with the following filter settings: Xcorr > 1.5 for +1 charged ions, Xcorr > 2.0 for +2 charged ions, Xcorr > 2.5 for +3 charged ions, and a peptide probability score $P < 0.5$ which was sufficiently strict to minimize false positive hits without removing valid +1 charged peptide sequences. These filter settings were chosen to suppress false positive identifications monitored by the number of proteins retrieved from a reversed-sequence decoy database [20] appended to a forward analytical database *mouse.nr.fasta* and indexed together. The software P-MOD™ 2.0 was used to search the MS/MS datasets for unexpected protein modifications using the protein sequences retrieved during the Bioworks™ data processing. When necessary, peptide sequences were subjected to NCBI-Blast searches to finalize protein assignments.

3. Results and discussion

3.1. Effect of formic acid in the sample injection volume on protein selectivity in LC chromatography

The effect of the presence of formic acid in the sample injection volume on LC protein separation was initially studied using the GST, SOD, GPD, and CarbAnh protein standards. When the standards (1 ml sample volume) were injected onto the LC column in 80% formic acid and in 0.1% TFA, the results shown in Fig. 1 were obtained (Fig. 1, traces A and B, respectively). These traces were almost identical with respect to the peaks of CarbAnh at RT 37.9 min, GST proteins at 39–40.5 min, and GPD at 43.8 min. However the SOD peak at 28.4 min RT was missing in the 80% formic acid experiment. This finding was confirmed by re-injecting the SOD – containing diluted flow-through from the latter experiment (data not shown). This phenomenon was further studied using mouse liver microsomal protein sample spiked with GST, SOD, GPD, and CarbAnh. The sample was introduced onto the LC column in a more dilute formic acid solution (50%) (Fig. 1, trace C). The retention times of the four protein standards remained almost unchanged and the SOD peak did appear in the chromatogram. These experiments demonstrated that higher concentration (80%) of formic acid in the sample injection volume led to selective protein depletion while the lower concentration (50%) did not.

This protein selectivity during sample loading on a LC column was further studied using microsomal protein samples as a model including a wide range of different proteins. Fig. 2 compares three LC chromatographic traces obtained using different means of sample introduction onto the LC column: in 8.3 M urea solution containing 1% acetic acid and 10% acetonitrile (top trace, A), 80% formic acid (middle trace, B), and 50% formic acid (bottom trace, C). Most chromatographic features of these three chromatograms appear identical except for the early RT portion that is missing from the trace B. This apparent protein depletion effect is consistent with the previously observed absence of the early eluting SOD peak in the LC run separating four protein standards using 80% formic acid during sample loading (Fig. 1A).

The effort to identify the proteins in the 80% formic acid run included those that were not retained on-column and were collected in the flow-through. Therefore it was important to subject the flow-through from the 80% formic acid run to lyophilization to dryness, reconstitution in a digestion buffer, digestion with trypsin and analysis by nanospray-LC–MS/MS to identify the proteins. For comparison, twenty-five 2-min-long LC fractions were collected for digestion and LC–MS/MS analysis from the LC run represented in the trace C, Fig. 2, in which the sample was initially solubilized in

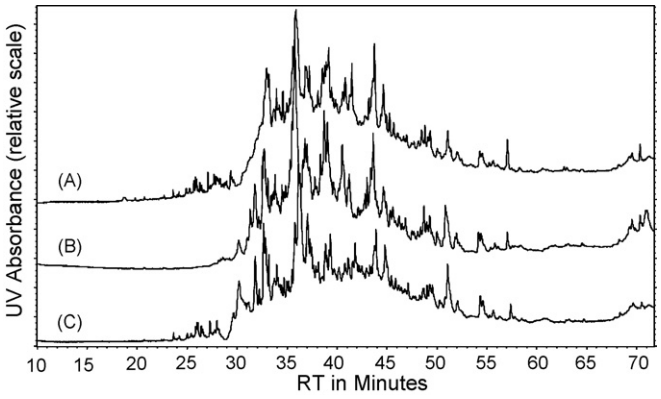

Fig. 2. Three LC chromatograms of protein samples injected on the LC column in different solutions and chromatographed using the same LC solvent gradient (Section 2.7): (A) (Section 2.3) 0.5 mg (1 mg/ml) of microsomal protein was injected on the LC column in 8.3 M urea with 1% acetic acid and 10% acetonitrile; (B) (experimental Section 2.4) 0.333 mg of microsomal protein was injected on the LC column in 80% formic acid and 20% water (v/v); (C) (experimental Section 2.5) 0.333 mg of microsomal protein was injected on the LC column in 50% formic acid and 50% water.

90% formic acid, then diluted with DI water and loaded onto the LC column under less acidic conditions using 50% formic acid (all identified proteins are listed in supplementary Table S-1b). Fifty-eight proteins were identified by LC–MS/MS in the LC flow-through digest (Table 1).

Most of the proteins identified in the flow-through were basic and were characterized by lower molecular weights. The average pI of the identified proteins was 9.43, calculated with a standard deviation of 1.83, and the average molecular weight was 26,776 Da, calculated with a standard deviation of 16,152 Da (calculated using the STDEVP Excel function). In contrast, the average molecular weight of all 414 non-redundant proteins identified by nano-LC–MS/MS of the protein digests in the 25 LC fractions of the LC run using 50% formic acid in the sample loading buffer was significantly higher (45,035 Da) and the 33,216 Da standard deviation was also larger demonstrating a bigger scatter of the data points across the molecular weight range. This result showed that smaller-size basic proteins were preferentially collected in the LC flow-through if the sample loading was performed in 80% formic acid. A closer inspection of the proteins identified in the flow-through showed that most of them possessed high pIs, or at least multiple basic amino acid residues (K, R, H), located near each other in their protein sequences (Table 1).

Out of the 58 proteins identified in the flow-through, 26 were basic ribosomal proteins with the pI in the range of 8.7–12.1. A minority of the proteins identified in the flow-through of the LC run involving 80% formic acid in the sample loading solution possessed pI below 7. However, these proteins all possessed clusters of basic amino acid residues making them potentially susceptible to the effects of acidity in the sample loading solution. For example, although progesterone receptor membrane component had a relatively low pI of 4.57, it possessed a RLKRR cluster of basic residues. Superoxide dismutase 1 soluble had a pI of 6.04 but it possessed a basic amino acid residues cluster HFNPHSKKH. Similarly, the bovine liver SOD protein standard that was used in this work had a pI of 6.59 and it contained a long sequence of basic residues RKKRRR near its C-terminus. This stands in contrast to the other three protein standards (GST, GPD, and CarbAnh) which were successfully deposited on the LC column in 80% formic acid and which did not possess such long sequences of basic amino acids. This result suggests that long sequences of basic amino acids, which are expected to be extensively protonated in 80% formic acid, favor far-reaching solvation in the solution rather than the

Table 1
Proteins identified by nano-LC-MS/MS in the LC flow-through from the experiment 2.4.

Amino acid sequence (highly basic)	Protein name	pI	MWt	Accession
RKKALKLGAKK	Argininosuccinate synthetase	8.29	46,555	6996911
KTRRR	ATP synthase, H + transporting, mitochondrial F1 complex, alpha subunit	9.22	59,716	6680748
KKR, KRYSYLKPR	ATP synthase, H + transporting, mitochondrial F1F0 complex, subunit e	9.74	8230	83715998
KNGKKIPK	B Chain B, Crystal Structure of murine class I Mhc H2-Db	7.34	11,811	3891454
KKAKKGILER	Betaine-homocysteine methyltransferase	7.82	44,992	7709990
KAKKK, HFIKLYKKKTGK	BiP	4.95	72,433	2598562
RRFK, KKDRVHH, RQWKGH	Calumenin isoform 1	4.34	37,041	6680840
HRHR, HRH	CATA_MOUSE Catalase	7.78	59,728	115704
RRVYK, KTKIKNYK	Cytochrome *c* oxidase, subunit VIb polypeptide 1	8.36	10,065	13385090
KKYIPGTKMIFAGIKKKG	Cytochrome *c*, somatic	10.01	11,598	6681095
KKAAGAGKVTKSAQKAQKAK	Eukaryotic translation elongation factor 1 alpha 1	9.04	50,140	13278382
KRVSKR	Fatty acid binding protein 1, liver	9.04	14,236	8393343
RRRIEILRRK	Fibrinogen, alpha polypeptide	7.21	61,288	33563252
TPVKKAKKKPAATPKKAKK, KRASKKK	H1 histone family, member 0	11.38	20,848	31560697
RK	Heat shock protein 1 (chaperonin 10)	9.06	10,956	6680309
KKVGRK, HRK	Heat shock protein 65	5.91	60,903	51455
KLNKRR, KKVLEKKFH	Heterogeneous nuclear ribonucleoprotein A/B isoform 2	8.11	30,812	6754222
RQSGKKR	Heterogeneous nuclear ribonucleoprotein A2/B1 isoform 2	9.04	32,440	32880197
RRRRRR	Heterogeneous nuclear ribonucleoprotein A3 isoform b	9.10	39,628	31559916
KAKTVKPKAAKPKTSKPKAAKPKKTAAKKK	Histone 1, H1e	11.10	21,964	13430890
KTNKRR	I49069 A + U-rich RNA-binding protein – mouse (fragment)	7.93	19,412	2137113
KVKKHLQDLSSRVSRARH	Low density lipoprotein receptor-related protein associated protein 1	7.95	42,189	63999380
RGKGGKGLGKGGAKRHRKVLR, HAKRK	mCG17007, H4 histone family	11.20	11,475	148700589
KKAVTPAPPMKR	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7 (B14.5a)	10.57	12,568	12963571
RRRR, RRYRRNFNYRRRR	Nuclease sensitive element binding protein 1	9.87	35,709	113205059
KKKKK	Phospholipase C-alpha	5.99	56,586	200397
KKKGPKVTVK	PPIB_MOUSE Peptidyl-prolyl *cis–trans* isomerase B precursor	9.93	22,699	118091
HLIKKAVAVRKH	PREDICTED: similar to ribosomal protein S13	10.55	17,238	82918395
RLKRR, KGRK, KYHHVGKLLK	Progesterone receptor membrane component	4.57	21,681	31980806
RKKAK	Ribosomal protein 10	10.16	24,516	56541260
YSAAKTKVEKKKKKEKVLAT	Ribosomal protein L6 (M-TAXREB107)	11.23	32,591	695638
KKRRTGCIGAKHRISK	Ribosomal protein L11	9.97	20,240	13385408
RFKVMKAKKMRNRIIK	Ribosomal protein L14	11.03	23,549	13385472
RKSRGLGKGHKFHHTIGGSRRAAWRRR	Ribosomal protein L15	12.11	24,131	13385036
KKKISQKKLKKQK	Ribosomal protein L17	10.20	21,383	161484662
KARKK, RKRR	Ribosomal protein L19	11.48	23,467	6677773.0
KKLVAKGGKKKK	Ribosomal protein L22	9.61	14,750	6677775
RKKVHPAVVIRQRKSYRRK	Ribosomal protein L23	10.51	14,856	12584986
RKRSSRHRGKVK, HKSKKK, KRR	Ribosomal protein L3	10.66	46,095	7305441
RGKKK, RRRLTKH	Ribosomal protein L35	11.53	14,544	13385044
RRK, KSVKIKKNK	Ribosomal protein L38	10.56	8199	12963655
RRWHRR, RRR, RKKIHRR	Ribosomal protein L4	11.49	47,124	30794450
KGKKAKGKK	Ribosomal protein L7a	11.02	29,958	7305443
KGKKAKGKK, RRH	Ribosomal protein L8	10.01	15,445	3851579
KKNR, RRSR	Ribosomal protein S10	10.53	18,904	13399310
RRKGGRRGRR	Ribosomal protein S14	10.14	16,291	3097244
QKKKRTFRKFTYR	Ribosomal protein S15	10.39	17,029	6677799
RRK, RLKKIRAHRGLRHFWGLRVRGQHTKTTGRRGRTVGVSKKK	Ribosomal protein S18	10.99	17,708	6755368
KLAKHK, KKH	Ribosomal protein S19	10.85	16,076	12963511
RIIAAKDH	Ribosomal protein S21	8.71	9136	21536222
RHGLYEKKKTSRKQRK, KKK	Ribosomal protein S24 isoform 3	11.38	15,187	46519160
RRLR	Ribosomal protein S28	10.70	7836	21426821
HKRRKTGGKRKPYHKKRK	Ribosomal protein S8	10.28	24,248	56541092
KNKRLTKGGKKGAKKK	RS3A_MOUSE 40S ribosomal protein S3a	9.75	29,866	3122836
RRRHSHSH, RRRH, RRR	Splicing factor, arginine/serine-rich 10	11.25	33,646	6677975
HFNPHSKKH	Superoxide dismutase 1, soluble	6.04	15,933	45597447
RMLRIKR	UCR6_MOUSE Ubiquinol-cytochrome *c* reductase complex 14 kDa	9.52	13,519	17380333
KKK, RCWRSKK	Zinc finger, CDGSH-type domain 1	9.17	12,089	19527228

surface of the RP LC column where the solvation is limited. In addition to the length of the basic amino acid sequence also the steric effects imposed by protein conformation and the protein size might decide about its appearing in the flow-through vs adhering to the column.

The proteins were cross-referenced between the two LC runs, and out of 58 proteins identified in the injection flow-through of the LC run in which sample injection was made in 80% formic acid, 50 were also identified among the proteins retained on-column when sample injection was made in 50% formic acid. The average RT for the proteins identified by LC–MS/MS in the flow-through was 32.4 min thus indicating a relatively early elution of these proteins.

3.2. Effect of adding a strong solvent to the sample injection volume on protein selection in LC RP chromatography

Solubilization of a hydrophobic analyte using organic solvents is a common sample preparation practice in chromatography. However, the presence of an organic solvent in a sample injection solution might result in an unexpected selectivity during analysis of complex protein mixtures. For example, the addition of 10%

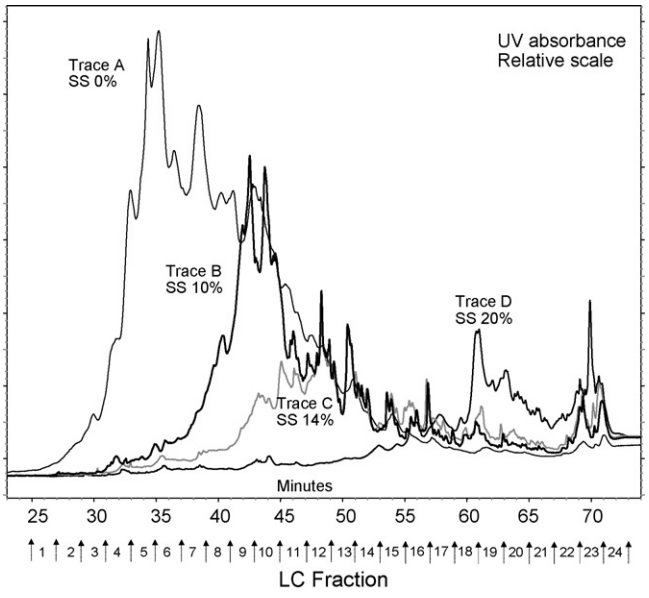

Fig. 3. An impact of an increasing amount of a strong solvent (SS; acetonitrile and DMSO 1:1, v/v) in 80% formic acid in the injection volume on the LC on-column analysis of 1 mg of a microsomal protein sample (Section 2.4.1). The *x*-axis is retention time in minutes. The LC fractions are shown as numbers separated by arrows. All traces have been normalized to the trace A.

acetonitrile to the sample buffer including 80% formic acid produced an altered LC chromatogram showing a dramatically reduced intensity of the proteins eluting before RT of 43 min (Fig. 1D). In contrast, the presence of 10% acetonitrile in the sample injection solution consisting of 8.3 M urea and 1% acetic acid did not result in such protein selectivity (Fig. 2, trace A). Protein selectivity during sample loading was also observed when samples were introduced in the mixture of two organic solvents (acetonitrile and DMSO) in 80% formic acid. The introduction of equal amounts of (5%) acetonitrile and (5%) DMSO in 80% formic acid resulted in a pronounced protein depletion, as evidenced in the LC chromatogram below RT 42 min (Fig. 3B).

DMSO was used in combination with acetonitrile because of the following properties. DMSO has been also identified as a denaturing reagent for membrane proteins and was found to assist in protein unfolding [21,22]. It has been also reported to compete for amide hydrogen atoms with carbonyl groups thus collapsing protein secondary structure supported by hydrogen bonding in native proteins [21]. In a study of beta-amyloid peptide components of Alzheimer's plaques, DMSO protein solutions exhibited unprotected monomeric protein entities comparable to these obtained from 8 M urea solutions [22]. Even at low DMSO concentrations, structural stability of several proteins was markedly reduced and it disrupted amide–water interactions by competing with the peptide carbonyl group for the bond donor capacity [23].

The three LC chromatograms on Fig. 3, A (acquired using 0% strong solvent; SS), B (10% SS) and C (14% SS) in 80% formic acid showed similar chromatographic features and they all converged

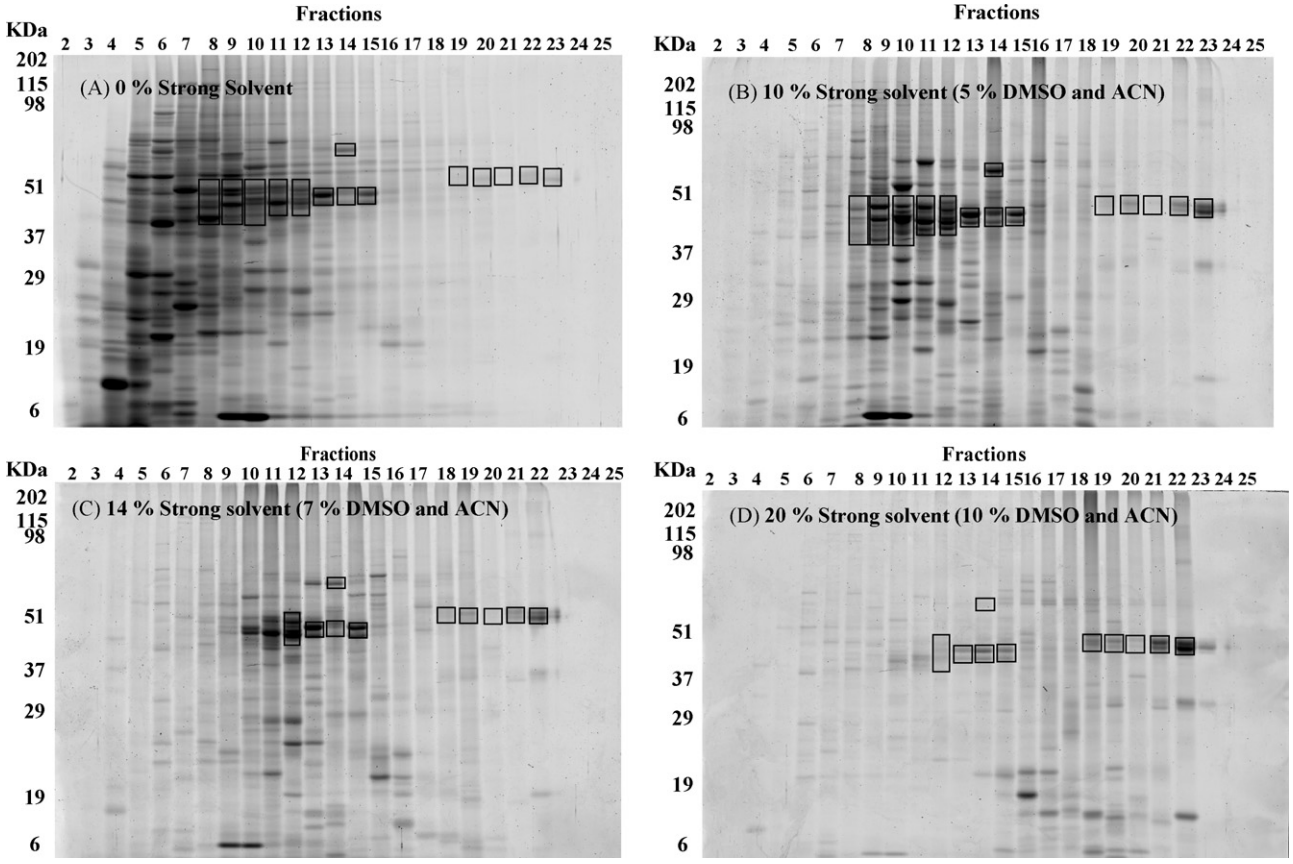

Fig. 4. 1D-GE analysis of the individual LC fractions (Section 2.9). 0.2 ml of the SDS-PAGE buffer was added to every lyophilized LC fraction and a quarter of that volume was loaded per gel lane. The marked gel bands were selected for protein digestion and LC–MS/MS analysis. Panel A: 0% SS LC run (Sections 2.4.1 and 2.7, 1.0 mg sample size). Panel B: 10% SS LC run (Sections 2.4.1 and 2.7, 1.0 mg sample size). Panel C: 14% SS LC run (Sections 2.4.1 and 2.7, 1.0 mg sample size). Panel D: 20% SS LC run (Sections 2.4.1 and 2.7, 1.0 mg sample size).

at the retention time of 55 min. The demarcation line between the proteins depleted and retained on the LC column moved to higher RT in concurrence with the increase of the concentration of strong organic solvent.

When equal sample volumes from the concentrated LC fractions were loaded onto the 1D-GE lanes the comparison of the LC elution patterns shown in Fig. 4 was achieved. For example, the 1D gel image for the LC run using 10% SS (trace B, Fig. 4) confirmed significant protein depletion prior to fraction 9 which corresponded to the LC RT of ca. 41 min. The experiment involving 20% SS (trace D, Fig. 4) resulted in the depletion of the proteins eluting prior to 60 min RT and exhibited an unusual enhancement effect for the late eluting (RT > 60 min) proteins, for which the LC (UV) response was approximately 2-fold the response shown in the traces B and C (Fig. 3). This enhancement might be due to the offset of the loss of late eluting hydrophobic proteins to solid surfaces of vials, syringes and transfer lines by the presence of the organic solvents.

A comparison of the 1D-GE gel images in Fig. 4 panel C (14% SS [7% DMSO–7% acetonitrile] and 80% formic acid) and Fig. 5 panel A (the flow-through from the above LC run), showed distinct protein bands for the LC fractions 1–10 which complemented the corresponding 1D-GE gel image of the proteins retained on the LC column showing strong protein bands for the LC fractions above 10. The flow-through sample was collected after a sample injection, diluted with protein LC solvent "A" and formic acid and a small portion of this volume was re-fractionated under the milder injection conditions using 50% formic acid concentration that ensured full protein retention on the LC column. A similar comparison was achieved between the 1D-GE images in Fig. 4 panel D (20% SS [10% DMSO–10% acetonitrile] and 80% formic acid) and those in Fig. 5 panel B (the flow-through from the above LC run) showing a similar complementary relationship of the strong vs weak intensity of the gel bands.

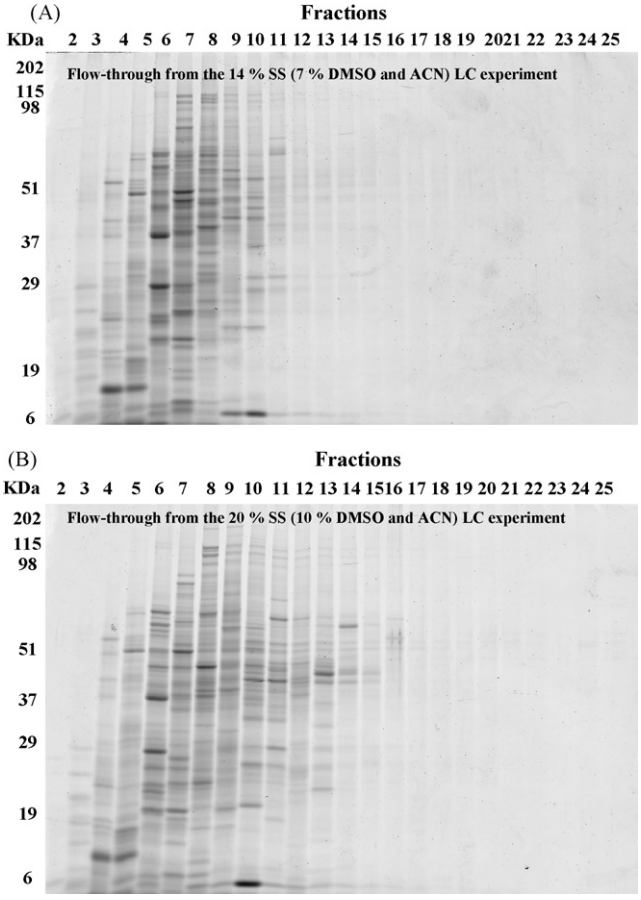

Fig. 5. 1D-GE analysis of the individual LC fractions of the re-fractionated LC flow-through (Sections 2.8, 2.9). Panel A: the fractionated flow-through from the 14% SS LC run. Panel B: the fractionated flow-through from the 20% SS LC run.

3.2.1. LC–MS/MS analysis of the re-fractionated LC flow-through

LC–MS/MS analysis of the protein digest of the re-fractionated flow-through obtained after the sample injection using 10% SS in 80% formic acid in the injection volume resulted in the identification of 297 proteins (Table S-1a, supplementary material) possessing average retention time of 38.4 min, calculated with the standard deviation of 7.4 min. This result supports the finding also evidenced by LC (UV) and 1D SDS gel electrophoresis demonstrating that a partial replacement of water with a strong RP organic solvent in the sample solution prepared with 80% formic acid results in the selective protein transfer to the flow-through. In general, the proteins remaining in the flow-through are those which would appear in the earlier portion of the LC chromatogram had the sample been loaded in the weaker solvent allowing all proteins to be deposited on the LC column.

3.3. LC–MS/MS analysis of the LC fractions separated on 1D-GE

The identification of the fractionated proteins by LC–MS/MS was preceded by in-gel trypsin digestion of the matching gel bands shown in the four panels in Fig. 4. The 1D gel separation step adopted prior to the LC–MS/MS analysis of the protein digests facilitated comparative analysis. The representative gel bands that were excised for the LC–MS/MS analysis spanned the molecular weight range of approximately 40–65 kDa. This selection was based on the fact that there was high protein concentration in these bands and there was interest in the LC separation of P450 proteins that appear in that gel region, and which are known to be difficult to separate by other methods (e.g. 2D-GE; [24,25]).

Direct comparison of the protein elution patterns within the 0% SS (Fig. 4A) and 10% SS (Fig. 4B) was achieved by selecting the representative gel bands of the same shape and size in the 0% SS and 10% SS experiments. Similarly, the 14% SS (Fig. 4C) and 20% SS (Fig. 4D) were compared. That pairing was performed to assure meaningful comparison because of the differences in the amount of protein present in the bands corresponding to the early and late LC fractions of these LC runs. The total number of the proteins identified from the selected gel bands in the 0, 10, 14 and 20% SS experiments was 129 (supplementary Table S-2a). Of these proteins, 108 were found in the LC – 1D-GE – LC–MS/MS experiments involving 10% SS out of which 83 were also identified in the 0% SS experiments thus leaving 25 proteins identified in the 10% SS experiment but unidentified in the 0% SS experiment. By comparison, there were 9 proteins identified in the 0% SS experiment (in the LC fractions 8 and 9) which remained unidentified in the 10% SS counterpart (Table S-2a).

A similar comparison could be made from the results obtained from the 14% and 20% SS LC fractionation experiments. There was a total of 57 proteins identified in the gel bands selected from the 14% SS LC fractions. Within this group there were 9 protein identifications unique to the 14% SS experiment. These unique proteins were identified in the gel bands representing LC fractions 12–19. There were a total of 32 proteins identified in the gel bands selected from the 20% SS LC fractions. Within this group there were 3 protein identifications unique to that experiment.

These results indicated that repeated protein fractionation using different sample injection conditions improved the LC–MS/MS identification scores of some proteins and also produced new protein identifications (Table S-2a). This was not surprising because of

the broad dynamic range of proteins in biological tissues and an automatic (data dependent) peak selection and exclusion used in the MS/MS analysis. However, eventually a plateau in the number of protein identifications would be expected for multiple replicate protein fractionation experiments carried out prior to the mass spectral analysis [26].

Out of the 129 total identified proteins, 10 proteins demonstrated calculated molecular weights that were significantly higher than the molecular weight estimated from their position in the 1D gel band. This finding might be explained by the existence of truncated protein forms. Protein distribution in the 1D-GE bands, and thus in the corresponding LC fractions, can be conveniently represented in a 3D graph (a digital heat map) as a percentage of the sum of all peptide abundances identified for each protein, calculated from the corresponding nano LC–MS reconstructed-ion chromatogram peak heights (supplementary Table S-2b). The top portion of this table compares the LC elution pattern for the 0 and 10% SS LC runs. The data are separated for convenience into three groups of proteins: cytochromes P450 (CYPs), UDP-glucuronosyltransferases (UGTs), and other. Within each group the protein entries are sorted in columns according to the decreasing protein Bioworks™ score. Relative protein amounts sum up within each row to a 100% for each LC experiment. The protein LC elution pattern established from the selected gel bands for the 0% SS LC run approximately matches the elution pattern established from the selected gel bands for the 10% SS LC run demonstrating reproducibility of LC chromatography under these sample injection conditions.

All of the 27 CYP proteins were identified using the excised 1D-GE gel bands in the LC fractions 9–13 (Table S-2b). Structural similarities within the families and subfamilies of the CYP proteins permitted them to elute in the adjacent LC fractions. In contrast, UGT (UDP) proteins exhibited a more scattered chromatographic pattern. The same UGT proteins were often identified in more than one LC fraction (Table S-2b). For example, UGT 2b1 was detected by LC–MS/MS in fractions 10 and also in 21–23. Although the reason for this UGT protein multiplicity remains yet unknown there are several possible explanations including an existence of multiple proteins isoforms, modifications and protein truncation.

3.4. *P-MOD™ search for artifacts of the fractionation method*

A subset of MS/MS data was subjected to P-MOD™ [27] software analysis using protein sequences retrieved from the Bioworks™ investigation to identify any possible artifacts of the LC chromatography with sample injection in formic acid solutions such as lysine and N-terminus formylation, however no such formylation pattern were discovered. Despite the elevated temperature of the LC column, formylation might have been avoided due to a brief contact of a protein sample with concentrated formic acid during sample injection. Similarly, no formylation was previously reported for the mRP-C18 LC protein separations performed at elevated temperature for protein samples prepared in formic acid [13]. Oxidation of the sulfur-containing amino acids, and a sporadic tryptophan oxidation was observed as evidenced by the increase in the tryptophan residue mass by 4, 16, 32 and 44 Da (data not shown). These findings are in agreement with previous research on tryptophan oxidation [28].

4. Conclusion

The presence of equal volumes of formic acid and water in the sample injection volume did not result in protein elution into the LC flow-through during sample injection. Using this method of injection, the samples which are difficult to solubilize directly in dilute formic acid might be initially dissolved in 80–90% formic acid, then diluted with water and immediately injected on the LC column. However, increasing the concentration of formic acid to 80% during sample injection resulted in the separation of smaller-size basic proteins, which eluted in the flow-through, from the bulk of other proteins, which were retained on the LC column. When microsomal protein samples were introduced onto the LC column in 80% formic acid, including increasing amounts (10, 14 and 20%) of acetonitrile or acetonitrile–DMSO mixture, more proteins eluted in the flow-through. The proteome was then divided into the proteins that were retained on-column which eluted at higher retention times (RTs), vs the proteins that collected in the injection flow-through, which normally eluted at lower RTs. This finding underscores the need for a careful choice of sample injection solutions for protein chromatography. The significance of such choice is that it can allow researchers to selectively introduce groups of proteins of interest in higher than normal quantity on a RP column, or collect and enrich target proteins in the flow-through. Because the RPLC fractionation pattern was reproducible (different samples were analyzed on different days) this technique could be potentially used for protein depletion and enrichment to improve recovery of target proteins from a complex protein matrix. In addition, its utilization of volatile solvents makes it compatible with the protein analysis by electrospray MS.

Acknowledgments

The authors acknowledge Dr. Daniel C. Liebler who developed and graciously provided the P-MOD vs. 2.0 shareware software. The authors are indebted to Dr. Jeffrey Ross, Dr. Kirk Kitchin and Dr. Stephen Nesnow from U.S. EPA for their excellent in-house review, and Dr. Amal Abu-Shakra from North Carolina Central University for her excellent review contribution of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2008.09.029.

References

[1] E.I. Chen, J. Hewel, B. Felding-Habermann, J.R. Yates III, Mol. Cell. Proteomics 5 (2006) 53.
[2] A. Bodzon-Kulakowska, A. Bierczynska-Krzysik, T. Dylag, A. Drabik, P. Suder, M. Noga, J. Jarzebinska, J. Silberring, J. Chromatogr. B 849 (2007) 1.
[3] D. Josic, J.G. Clifton, J. Chromatogr. A 1144 (2007) 2.
[4] J. Carroll, I.M. Fearnley, J.E. Walker, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 16170.
[5] J. Adachi, C. Kumar, Y. Zhang, J.V. Olsen, M. Mann, Genome Biol. 7 (2006) R80.
[6] J. Martosella, N. Zolotarjova, H. Liu, G. Nicol, B.E. Boyes, J. Proteome Res. 4 (2005) 1522.
[7] R.P. Lee, S.W. Doughty, K. Ashman, J.E. Walker, J. Chromatogr. A 737 (1996) 273.
[8] A. Premstaller, H. Oberacher, W. Walcher, A.M. Timperio, L. Zolla, P. Chervet, N. Cavusoglu, A. van Dorsselaer, C.G. Huber, Anal. Chem. 73 (2001) 2390.
[9] C.G. Huber, W. Walcher, A.M. Timperio, S. Troiani, A. Porceddu, L. Zolla, Proteomics 4 (2004) 3909.
[10] M. Pepaj, S.R. Wilson, K. Novotna, E. Lundanes, T. Greibrokk, J. Chromatogr. A 1120 (2006) 132.
[11] W.M. Winnik, Anal. Chem. 77 (2005) 4991.
[12] K. Zhu, K. Jeongkwon, C. Yoo, F.R. Miller, D.M. Lubman, Anal. Chem. 75 (2003) 6209.
[13] J. Martosella, N. Zolotarjova, H. Liu, S.C. Moyer, P.D. Perkins, B.E. Boyes, J. Proteome Res. 5 (2006) 1301.
[14] E. Gross, B. Witkop, J. Am. Chem. Soc. 83 (1961) 1510.
[15] N. Zolotarjova, P. Mrozinski, H. Chen, J. Martosella, J. Chromatogr. A 1189 (2008) 332.
[16] N.I. Govorukhina, P. Horvatovich, R. Bischoff, in: J.D. Thompson, C. Schaeffer-Reiss, M. Ueffing (Eds.), Functional proteomics methods and protocols. Methods in Molecular Biology, vol. 484, The Humana Press, Totowa, NJ, 2008, p. p. 49.
[17] D. Wessel, U.I. Flugge, Anal. Biochem. 138 (1984) 141.
[18] M. Kinter, N.E. Sherman, Protein Sequencing and Identification using Tandem Mass Spectrometry, Wiley-Interscience, 2000.

[19] J.E. Coligan, Current Protocols in Protein Science, John Wiley & Sons, Inc., Brooklyn, NY, 1998, Section 11.1.
[20] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome Res. 2 (2003) 43.
[21] W. Dzwolak, A. Loksztejn, V. Smirnovas, Biochemistry 45 (2006) 8143.
[22] S.S.-S. Wang, S.A. Tobler, T.A. Good, E.J. Fernandez, Biochemistry 42 (2003) 9507.
[23] M. Jackson, H.H. Mantsch, Biochim. Biophys. Acta 1078 (1991) 231.
[24] I.P. Kanaeva, N.A. Petushkova, A.V. Lisitsa, P.G. Lokhov, V.G. Zgoda, I.I. Karuzina, A.I. Archakov, Toxicol. In Vitro 19 (2005) 805.
[25] N.A. Petushkova, I.P. Kanaeva, A.V. Lisitsa, G.F. Sheremetyeva, V.G. Zgoda, N.F. Samenkova, I.I. Karuzina, A.I. Archakov, Toxicol. In Vitro 20 (2006) 966.
[26] E. Barnea, R. Sorkin, T. Ziv, I. Beer, A. Admon, Proteomics 5 (2005) 3367.
[27] B.T. Hansen, S.W. Davey, A.-J.L. Ham, D.C. Liebler, J. Proteome Res. 4 (2005) 358.
[28] S.W. Taylor, E. Fahy, J. Murray, R.A. Capaldi, S.S. Ghosh, J. Biol. Chem. 278 (2003) 19587.