ELSEVIER

Journal of Chromatography B, 782 (2002) 317–329

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

# Low-mass proteome analysis based on liquid chromatography fractionation, nanoliter protein concentration/digestion, and microspot matrix-assisted laser desorption ionization mass spectrometry

Bernd O. Keller, Zhengping Wang, Liang Li*

*Department of Chemistry, Faculty of Science, University of Alberta, E3-44 Chemistry Building, Edmonton, Alberta T6G 2G2, Canada*

**Abstract**

HPLC fractionation combined with mass spectrometry can become a powerful tool for analyzing the proteome in the mass range below 15 kDa where efficient protein separation by gel electrophoresis can be difficult. For sensitive and high-resolution separation of the low-mass proteome, the use of analytical rather than preparative HPLC columns is preferred. However, individual fractions collected by a conventional HPLC separation usually contain a small amount of proteins whose concentrations may not be sufficiently high for subsequent enzyme digestion and protein identification by mass spectrometry. In this work, we present a high sensitivity nanoliter sample handling technique to analyze proteins fractionated by HPLC. In this technique, an individual HPLC fraction in hundreds of microliter volume is pre-concentrated to several microliters. About 700 pl of the pre-concentrated fraction is then drawn into a 20-μm I.D. capillary and dried in a small region near the capillary's entrance. This process can be repeated many times to concentrate a sufficient amount of protein to the small region of the capillary. After protein concentration, protein digestion is achieved by drawing 1 nl of chemical or enzymatic reagent into the capillary and placing it in the same region where the dried protein sits. The resulting peptides are then deposited onto a microspot in a MALDI probe for mass analysis. The performance of this technique is demonstrated with the use of a standard protein solution. This technique is applied to the identification of low-mass proteins separated by HPLC from a complex mixture of an *E. coli* extract.


*Keywords:* Proteomics; Low-mass proteomes

## 1. Introduction

Proteome display by gel electrophoresis combined with protein identification by mass spectrometry (MS) and database searching has become a standard technique for many applications in proteomics [1]. A disadvantage often encountered with polyacrylamide gels is the inability to extract whole proteins from gel in their native form or in sufficient amounts for further studies [2]. This problem is especially true for gel spots containing low abundance proteins, where often the only option is in-gel digestion and mapping of the extracted peptides by mass spectrometry. Also, resolution and detection of the low molecular mass (LMW) proteome (<15 kDa) by gel electrophoresis is difficult compared to a medium to

*Corresponding author. Tel.: +1-780-492-3250; fax: +1-780-492-8231.

*E-mail address:* liang.li@ualberta.ca (L. Li).

1570-0232/02/$ – see front matter 
PII: S1570-0232(02)00565-2

high MW proteome (20–200 kDa). The use of Tris–Trycine gels helps the separation of small proteins and peptides, but this technique is still not adequate for displaying the whole LMW proteome [3]. Display and identification of the LMW proteome is thus considered to be a challenging task for post-genomic research. Taking *E. coli* as an example, the genome of the K12 strain is known and searching the *E. coli* protein database reveals that there are 986 probable proteins varying in size from 51 to 150 amino acids in SwissProt and 2077 in TrEMBL and SwissProt combined (http://www.expasy.org/srs5, as of February 2002). In one study, 42 protein spots in the molecular mass region of 6–15 kDa were resolved by 2D-gel electrophoresis, which is impressive considering the limitation of the 2D-gel resolving power for LMW proteins [4]. Nevertheless, to provide a comprehensive LMW proteome analysis, an alternative separation and detection technique is clearly needed.

While new techniques such as the use of modified surfaces to selectively capture proteins [5–7] can be useful in some applications, they do not provide adequate separation power for comprehensive proteome display. Chromatographic separation can potentially become a very powerful technique for protein purification and especially for LMW proteome display. High-performance liquid chromatography (HPLC) is particularly suitable for separating peptides and small proteins with good resolution and in a relatively short separation time. Multidimensional LC techniques employing, for instance, ion-exchange combined with reversed-phase chromatography, have improved LC separation power considerably. Another important attribute of HPLC is that it does not introduce modifications to proteins, as often is the case when using acrylamide gels in gel electrophoresis [2]. Accurate protein mass analysis can be readily performed on HPLC-separated proteins using LC–MS in either on-line or off-line modes [8,9]. As an example, 422 proteins with molecular masses ranging from 2000 to 15,000 Da from an *E. coli* extract can be detected by using HPLC combined with off-line matrix-assisted laser desorption ionization (MALDI) MS [10]. However, an important issue in developing a chromatography-based LMW proteome display system is related to protein identification. While protein mass analysis requires only low femtomoles of protein, protein identification by peptide mapping generally requires sub- to low-picomoles of sample. Moreover, to fully characterize a protein including protein identification with high confidence and investigation of post-translational modifications, multiple experiments on a protein sample are needed, which require even a greater amount of protein for analysis.

A conventional approach to identifying a protein separated by HPLC is to use a preparative column to isolate a sufficient amount of protein for enzyme digestion and subsequent analysis of the digest by MS. Unfortunately, for analyzing complex protein mixtures, preparative HPLC does not always provide adequate resolving power and it requires a large amount of starting material. The latter can be a major limitation for proteomics projects involving a limited supply of cells such as those from a patient's tumor tissue. Analytical or small-bore column separation provides much better resolution and consumes significantly less starting material. However, the use of an analytical column for separation results in a small amount of protein in individual LC fractions. To circumvent this problem, one can carry out multiple runs and then pool the corresponding fractions together for subsequent protein identification. Multiple HPLC runs are of course a time- and solvent-consuming process. Recently, Lubman et al. reported an elegant technique that combines the benefit of high sample loading of preparative HPLC with the high resolving power of analytical HPLC for separation and MS identification of proteins [11]. High sample loading and high resolving power are achieved by using two columns of different lengths and operated at different temperatures linked in series. The column length and temperature of the first column are optimized to provide high sample loading with some resolution of the protein mixture. The conditions for the second column are optimized to provide high resolution of the proteins. While this technique provides an alternative to running multiple analytical-column HPLC, the starting material required for protein identification is still equivalent to that used in a preparative HPLC experiment.

We present a method that allows multiple experiments to be carried out from individual fractions separated using analytical-column HPLC. It involves the use of a small capillary tube to draw a nanoliter

volume of solution from a HPLC fraction and then concentrate the protein inside the capillary. This is followed by chemical and enzymatic reactions of protein residing inside the capillary tube. The resulting peptides are analyzed by microspot MALDI. The performance of the technique is demonstrated with cytochrome *c* as a standard. The application of this method is then illustrated for the characterization of low-mass proteins fractionated by analytical HPLC column from an *E. coli* extract.

## 2. Experimental procedures

### *2.1. Chemicals and materials*

*E. coli* bacteria samples were from Edgewood RDE Center, Aberdeen Proving Ground, MD, USA. Dithiothreitol (DTT), iodoacetamide, α-cyano-4-hydroxycinnamic acid (HCCA) and trypsin (98%, L-1-tosylamide-2-phenylethyl chloromethyl ketone (TPCK) treated for reduction of chymotrypsin activity), horse cytochrome *c*, leucine aminopeptidase (LAP), and trifluoroacetic acid (TFA) were from Sigma–Aldrich–Fluka (Oakville, Ontario, Canada). HCCA was recrystallized from ethanol (95%) at 50 °C before use. Water was obtained from a Milli-Q Plus purification system (Millipore Corporation, Bedford, MA, USA).

### *2.2. Extraction of bacteria samples*

The *E. coli* 9637 extract was prepared by solvent suspension methods [12]. Briefly, about 5 mg lyophilized *E. coli* was suspended in 500 μl 0.1% TFA, the suspension was vortexed for 3–5 min, centrifuged at ~11,750 *g*, and the supernatant was transferred into a fresh vial. This extraction procedure was repeated three times per sample. The supernatants were pooled, filtered with a Microcon-3 (Amicon, Beverly, MA, USA), and then concentrated to 50 μl by a high-speed vacuum centrifuge.

### *2.3. HPLC-fractionation*

Separation of *E. coli* 9637 extract was performed on a HP1100 HPLC (Agilent, Palo Alto, USA) using a 4.6×250 mm $C_8$ column (Vydac, Hesperia, CA). The mobile phases were nanopure water (A) and acetonitrile (B) with 0.05% TFA in both phases. The solvent gradient was 2–20% B in 10 min, 20–40% B in 40 min, and then 40–55% B in 10 min. The flow-rate was 0.5 ml/min. For injection, the total volume of the concentrated extract was used (~50 μl). Fractions were collected in 1-min intervals during the run. The fractions were concentrated to about 10 μl by a high-speed vacuum centrifuge.

### *2.4. In-capillary sample concentration, reaction and microspot MALDI sample preparation*

All experiments were done using a nanoliter chemistry station, which has been described in more detail elsewhere [13]. Twenty-μm-I.D. fused-silica capillaries were purchased from Polymicro Technologies (Phoenix, AZ, USA). The capillaries were treated with the siliconizing agent Glassclad-18 (United Chemical Technologies, Bristol, PA, USA) before use to deactivate the surface silanols and thus minimize irreversible protein adsorption. The capillary tube was connected to a syringe; sub-nanoliters of protein sample were drawn from a horizontally mounted pipette tip. The volume of sample plug was determined using a calibrated recticle that was positioned in the eyepiece of the microscope. As illustrated in Fig. 1, for in-capillary sample concentration, an ~700 pl sample plug was dried inside

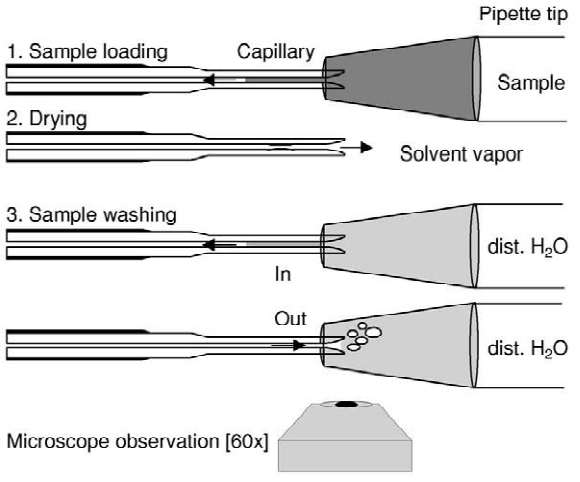


Fig. 1. Schematic of in-capillary sample concentration and washing.

the capillary close to the capillary entrance. This step was repeated up to 20 times to achieve sufficient sample concentration inside the capillary. During the repeated concentration process, after one or two concentration steps, a plug of ~1 nl of triply-distilled water was drawn into the capillary and pushed out after ~20 s. This washing step is very critical since the small capillary is easily plugged by the accumulation of excessive salts, if present. To accelerate the drying process, an orthogonal $N_2$ gas flow was applied at the open capillary end, as is illustrated in Fig. 2.

In-capillary enzymatic or chemical reaction was achieved by drawing a plug of buffered enzyme or other chemical solutions into the capillary (Fig. 3). The capillary was then pushed against a piece of Parafilm to close the entrance and thus prevent any

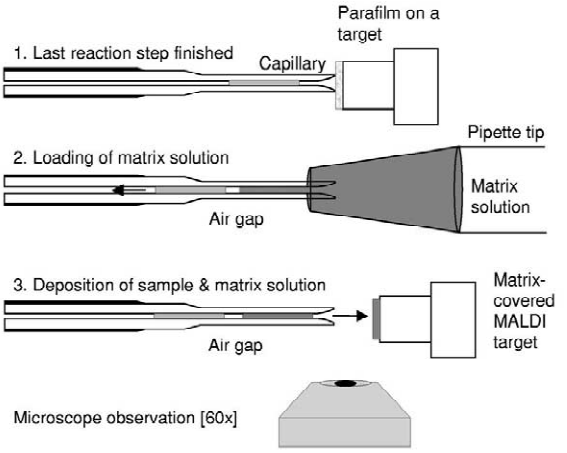


Fig. 3. Schematic of in-capillary reaction and microspot MALDI sample deposition.

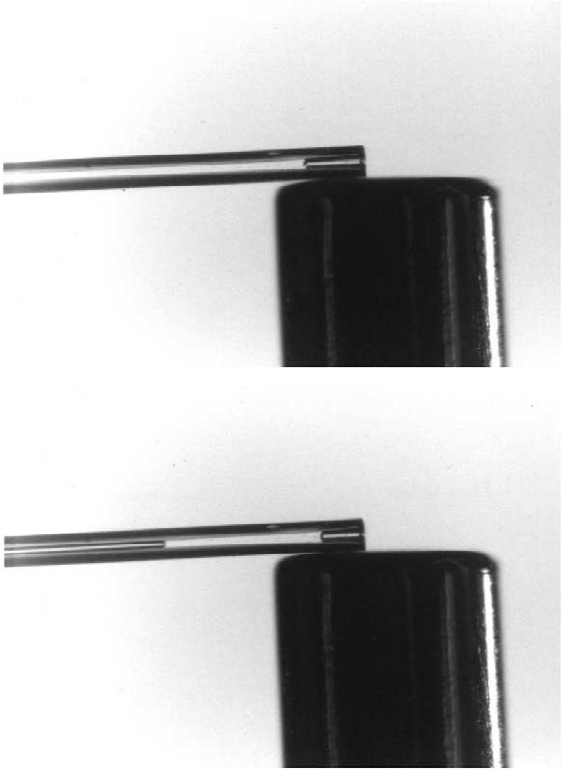

Fig. 2. Microscope images (40×) of capillary with a shrinking sample plug. A needle tip with airflow placed in orthogonal to the capillary is used to speed up the drying of the sample plug inside the capillary.

evaporation. After sufficient reaction time the sample/enzyme or sample/chemical mixture was dried up again inside the capillary. Further chemical or enzymatic reactions can be done by introducing different chemical/enzyme solutions. After the final reaction step the reaction mixture plug was left intact and not dried up. When all desired reactions had been performed, an ~500 pl plug of saturated matrix solution was drawn into the capillary. The sample and matrix solution were separated by a small air gap. Both plugs were then simultaneously deposited from an approximate 0.1 mm distance onto a matrix-covered MALDI target. Typical sample spots deposited by the microspot sample deposition are about 80 to 200 μm in diameter. Enzyme and reactant solutions were composed as follows: trypsin, 2 μ*M* in 50 m*M* $NH_4HCO_3$; DTT, 45 m*M* in dist. $H_2O$; iodoacetamide, 90 m*M* in dist. $H_2O$; leucine aminopeptidase (LAP), 7 μ*M* in 20 m*M* $NH_4HCO_3$. In-capillary tryptic digestion, reduction and alkylation steps were usually carried out for 20–30 min each.

To prepare the matrix-covered MALDI target a two-layer method was employed [14]. Briefly, about 1 μl of a 5 mg/ml solution of HCCA in 80% acetone–methanol (v/v) was first deposited onto the clean target to form a thin first layer. A second layer of 0.4 μl of HCCA saturated in 40% methanol–water (v/v) was deposited and allowed to dry. The

matrix layers were then washed twice with Milli-Q water.

### 2.5. MALDI analysis

Mass spectra of proteins and their digests were collected on a home-built linear time-lag focusing MALDI-TOF mass spectrometer, equipped with a 337 nm laser having a 3 ns pulse width (model VSL 337ND, Laser Sciences, Newton, MA, USA). This home-built instrument has been described in detail elsewhere [15]. In general, 150–200 laser shots (3–5 μJ pulse energy) were averaged to produce a mass spectrum. Spectra were acquired and processed with Hewlett-Packard supporting software and reprocessed with the Igor Pro software package (Wavemetrics, Lake Oswego, OR). Each spectrum was normalized using the most intense signal.

## 3. Results and discussion

### 3.1. Nanoliter sample concentration

To illustrate how effective we can concentrate proteins inside the capillary using the set-up shown in Fig. 1, a plug (~700 pl) of dilute cytochrome *c* solution (30 or 50 n*M*) was drawn into the 20-μm capillary and dried close to the end of the capillary entrance. This step was repeated 20 times so that a total amount of about 400 to 700 amol of cytochrome *c* was deposited inside the capillary. Fig. 4A

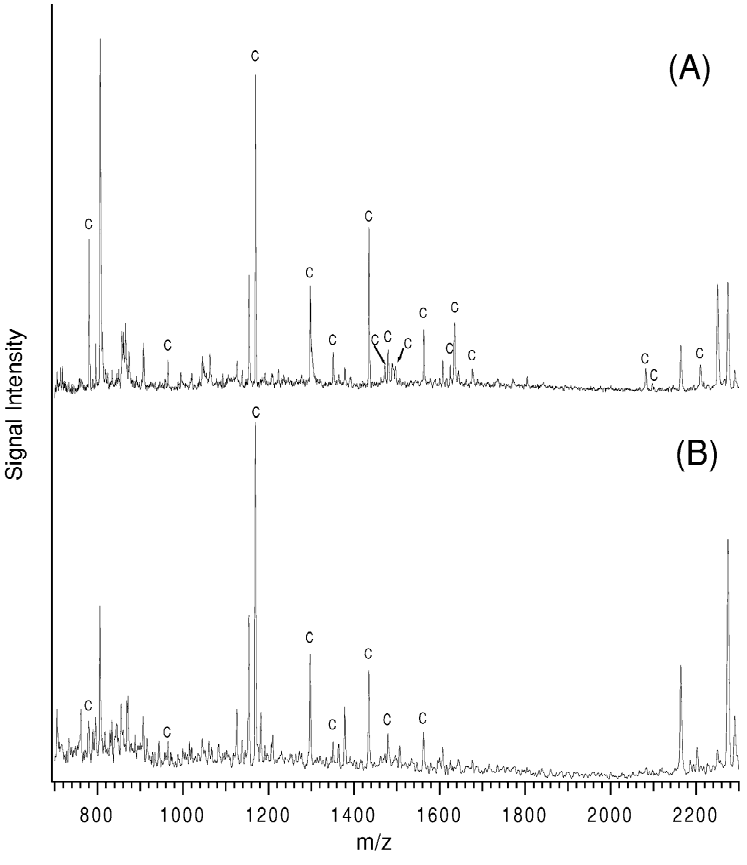


Fig. 4. MALDI mass spectra of peptides generated from in-capillary concentration and digestion of dilute cytochrome *c*: (A) 50 n*M* and (B) 30 n*M*. In both cases about 20 portions of ~700 pl protein solution in dist. $H_2O$ were dried inside the capillary.

shows the spectrum obtained from the digest of the in-capillary concentrated protein from the 50 n*M* cytochrome *c* solution. A total of 16 peaks were identified as tryptic peptides from cytochrome *c*, corresponding to sequence coverage of 71%. For the 30 n*M* cytochrome *c* solution, eight specific peaks could be identified as tryptic peptides from cytochrome *c*, still covering 32% of its sequence (see Fig. 4B). This example demonstrates that the in-capillary nanoliter sample handling technique is very effective for dilute protein sample concentration and subsequent digestion. Combined with microspot MALDI sample deposition, very low amounts of protein sample are required for peptide mass mapping. We note that in-capillary digestion is much more sensitive than either in-tube or on-plate digestion. For in-tube digestion, a solution concentration of 100 n*M* is required for cytochrome *c*. If we assume a minimum volume of 1 μl for in-tube digestion, we need ~100 fmol of protein sample. For on-plate digestion, the lowest amount we were able to do digestion with was about 20 fmol of cytochrome *c* [16].

While peptide mass mapping such as that shown in Fig. 4A is adequate for identifying the protein in question, many instances exist where an insufficient number of peptides are detectable in the tryptic digest, which makes it difficult for unambiguous protein identification by peptide mass mapping alone. This can either be due to the lack of a sufficient amount of the starting material to generate an adequate number of peptides or due to the contamination of digested peptides from other proteins in a protein mixture (e.g. co-eluted proteins in a HPLC fraction). It has been shown by a number of groups that additional sequence information of only one or two tryptic peptides (i.e. sequence tag) is often enough for confident protein identification [17–19]. A common technique for obtaining sequence information is to carry out MS–MS fragmentation of isolated peptides. MS–MS spectra can be obtained by using collision-induced dissociation (CID) in tandem mass spectrometry [20,21]. Alternatively, post-source decay (PSD) fragment ion spectra can be obtained using a reflectron MALDI TOF instrument [22].

A different way of obtaining additional sequence information is the application of exoproteolytic enzymes directly to tryptic digest mixtures [23–30]. For example, leucine aminopeptidase M (LAP) cleaves subsequently single amino acids from the N-terminus of peptides [25]. Several research groups have successfully used this enzyme to create N-terminal peptide ladders [26–29]. We have recently shown that trypsin digestion followed by LAP digestion can be very useful to provide peptide sequence information for protein identification [30]. All of the work reported thus far was performed using microliters of solution and required picomoles or subpicomoles of protein. With the nanoliter sample handling technique, we can now carry out the sequential digestion experiment using a very small amount of protein sample. This is illustrated in Fig. 5. Fig. 5A shows a spectrum of a tryptic digest of cytochrome *c*. The total sample used for digestion and MS analysis was 1.2 fmol or 15 pg of protein. Fig. 5B shows the mass spectrum from another tryptic digest after exposure to LAP for 2 min. In this case, total sample loading was 1.4 fmol or 17 pg. Two of the tryptic peptides originating from cytochrome *c* undergo N-terminal exoproteolytic digestion by LAP. The peptide with average mass 1351.5 Da and sequence TEREDLIAYLK as well as the peptide with average mass 1169.3 Da and sequence TGPNLHGLFGR each lose threonine at their N-terminus, yielding peptides with masses at 1250.4 and 1068.2 Da, respectively. Since the in-capillary technique usually employs equal or excess amounts of trypsin (compared to protein) to speed up the digestion, it is clear that trypsin autolytic peptides can be present in high enough amounts to undergo exoproteolytic digestion as well. Some of the peaks in Fig. 5 are the result of trypsin autolytic peptides being digested further by LAP. The autolytic peptide with average mass 1154.3 Da and sequence SSGTSYPDVLK loses two serines at the N-terminus, yielding a peptide with mass 979.5 Da. In summary, the results shown in Fig. 5 demonstrate that additional sequence information is obtainable at the low femtomole or picogram level using the nanoliter sample handling technique, which can improve the confidence of protein identification, as illustrated below. Since the drying process for a 500-pl plug was in the range of 1–2 min, there could remain significant enzymatic or chemical activity during the drying step. However, it should be noted

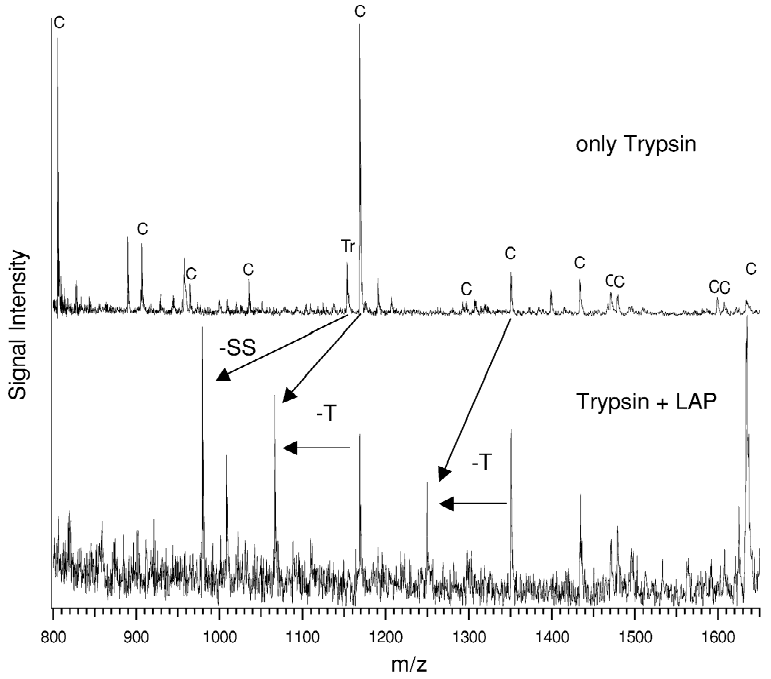


Fig. 5. MALDI mass spectra of sequential enzymatic digestion of cytochrome *c*: (A) peptides generated from trypsin digestion and (B) peptides generated from trypsin digestion, followed by LAP digestion.

that, for tryptic digestion or chemical reactions, further reaction during the drying process is of no concern since the goal of these reactions is towards completion. In case of the time study employing LAP for sequential N-terminal cleavage of tryptic peptides, the LAP step is always the final step and no drying process will follow. Thus accurate control of time is possible, since the LAP reaction stops immediately after sample deposition.

### *3.2. Applications to identify HPLC-fractionated proteins*

The main advantage of the in-capillary protein concentration and digestion technique is that it can handle very small sample volumes and each experiment consumes a very small amount of sample. Thus, many experiments, including optimization of digestion conditions and digestion with different enzymes, can be performed from a microliter volume of individual HPLC fractions. To apply the nanoliter sample handling technique to the identification of bacterial proteins fractionated by HPLC, each fraction was pre-concentrated by SpeedVac from the original volume of 500 μl to about 10 μl shortly after the fractionation experiment. Further concentration to a lower volume in a microvial would result in an excessive loss of the protein sample due to extensive and irreversible protein adsorption to the microvial wall. We note that, for a high concentration of protein solution, sample loss to the container wall does not significantly affect the final solution concentration after drying. However, for a dilute protein solution, which is often the case in HPLC fractions resulting from separation of complex protein mixtures such as cell lysates, any loss to the container can significantly affect the final protein concentration of the dried solution. A detailed study on the issue of protein adsorption and sample loss during vacuum drying of a dilute protein solution will be reported elsewhere.

In the in-capillary concentration experiment, we take advantage of the protein adsorption property by concentrating the protein inside a capillary through repeated drying of a small amount of sample in a fixed area. The area is small so that after con-

centration a nanoliter volume of reagent can be drawn into the capillary for digestion and the digested sample can be deposited onto a MALDI probe. For a pre-concentrated HPLC fraction, a few nanoliters of sample were first taken for molecular mass analysis, followed by drawing another few nanoliters of sample for trypsin digestion and subsequent peptide mass mapping. If peptide mass mapping, along with the protein molecular mass information, could not unambiguously identify the protein in the database, several nanoliters from the remaining fraction were taken for further experiments such as the trypsin/LAP sequential digestion. Techniques were exhausted until we could confidently identify the protein or determine that identification is not possible with the currently available techniques and database. Several examples of identifying low-mass proteins from bacterial cell extracts after HPLC fractionation are given below to illustrate the application and performance of the nanoliter technique. The UV chromatogram of the extract is shown in Fig. 6.

Fig. 7 is an example in which *E. coli* 30S ribosomal protein S20 (SwissProt Accession# P02378, 9554 Da) was positively identified by peptide mass mapping in combination with accurate protein molecular mass determination. A very clean peptide mass map was obtained from this sample (fraction #42). Almost all the major peaks matched 30S ribosomal protein S20 with sequence coverage of ~60%.

Fig. 8 demonstrates the benefit of using multiple experiments to unambiguously identify proteins. The major protein in HPLC fraction #52 was tentatively identified as DNA binding protein HU alpha (SwissProt Accession# P02342, 9535 Da) by peptide mass mapping. Although the molecular mass matched very well with this protein and the matching tryptic peptides covered more than 80% of the sequence, many unmatched peaks were observed which led to a poor probability score when doing a database search. To confirm the protein's identity, a sequential trypsin/LAP digestion was performed on this fraction. Panels B and C in Fig. 8 show sections of mass spectra from in-capillary sequential digestion of this fraction. In the displayed mass range, two

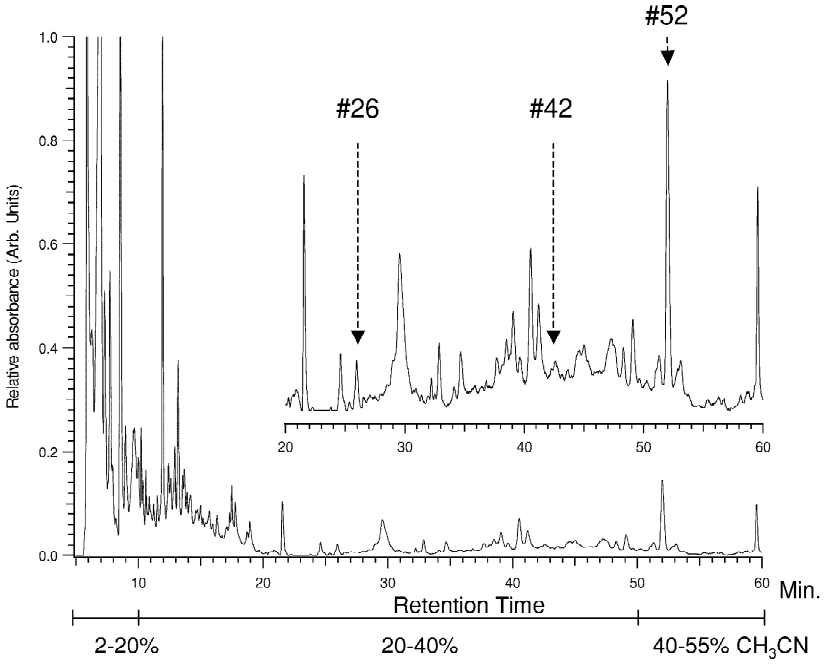


Fig. 6. UV chromatogram of HPLC separation of *E coli* extract. Fractions #26, 42 and 52 used in this work are each indicated with an arrow in the inset. The employed solvent gradient is shown below the chromatogram.

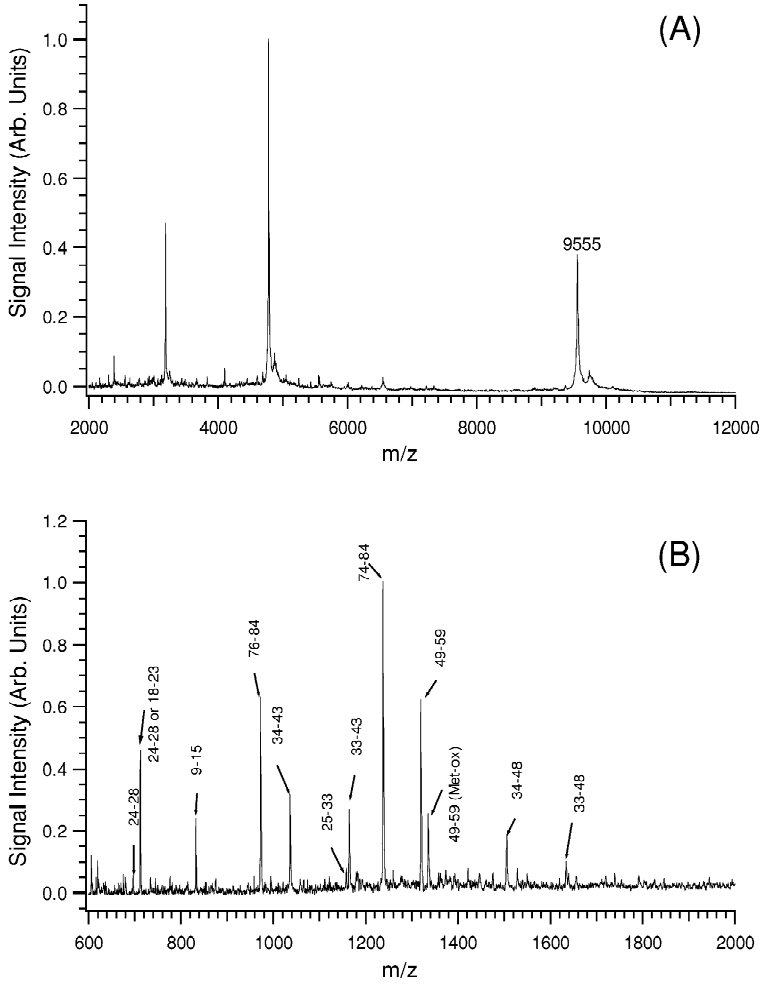


Fig. 7. MALDI analysis of HPLC fraction #42 of *E. coli* extract. (A) Molecular weight determination, (B) MALDI peptide mass mapping.

peptides underwent exoproteolytic digestion by LAP. If the peptide with monoisotopic mass 958.5 Da is from DNA binding protein HU alpha, it should have a sequence of TGRNPQTGK. A new peptide peak at 857.5 Da due to the loss of the N-terminal threonine was detected and it did not undergo further digestion by LAP. If we again assume the protein is DNA binding protein HU alpha, the peptide with mass 1244.7 Da should have a sequence of IAAAN-VPAFVSGK. Indeed, several N-terminal amino acid losses after LAP treatment corresponding to this sequence were observed. The assigned sequence tags are shown in Fig. 8B. Note that the exoproteolytic digestion stopped at the V–P bond, since LAP is not capable of cleaving V–P bonds [21]. It should also be noted that the signal intensity from the peptide VPAFVSGK ($MH^+$=804.47 Da) increased with the process of N-terminal digestion, whereas the intensities of all the intermediate N-terminal ladder peptides decreased, as shown in Fig. 8B and C. The combination of peptide mass map and the short sequence tags obtained by sequential enzymatic digestion confirms the identity of the major protein in fraction #52 to be DNA binding protein HU alpha.

The ability to perform multiple reactions in the nanoliter chemistry station is also very valuable for identifying proteins containing multi-cysteines. Fig.

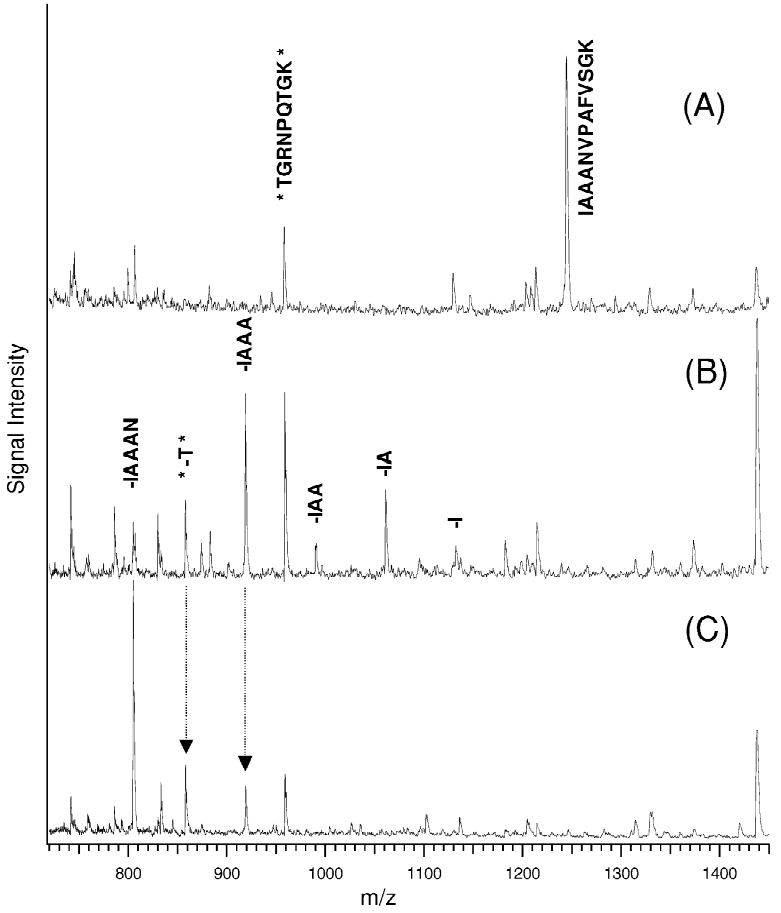


Fig. 8. Expanded MALDI mass spectra of in-capillary digestion of fraction #52 containing DNA-binding protein HU-alpha. (A) Peptides generated from trypsin digestion. (B) Peptides generated from trypsin digestion followed by LAP digestion for 5 min. (C) Peptides generated from trypsin digestion followed by LAP digestion for 15 min. For each experiment a total volume of ~2 nl was concentrated in ~500 pl portions inside the capillary.

9 shows the mass spectra of trypsin digests of *E. coli* HPLC fraction #26 containing a protein with mass 7867 Da (determined by internal mass calibration with ubiquitin and its doubly-charged species, mass spectrum not shown). Peptide mapping using the data from Fig. 9A identified a top candidate of 50S ribosomal protein L31 (SwissProt. Accession# P02432). However the sequence coverage was only 29%, and the detected molecular mass is 4 Da less than the one predicted in the proteome database. Moreover, a number of peptide peaks did not match this protein or any trypsin autolysis peaks. To increase the confidence in protein identification, we examined the mass spectral data carefully with the assistance of protein structure information contained in the SwissProt database. 50S ribosomal protein L31 consists of 70 amino acid residues:

MKKDIHP**K**YEEITAS**C**S**C**GNVM**K**$^{23}$I**R**$^{25}$STVGHDLNLDV**C**S$^{39}$**K****C**H-PFFTG**K**QRDVATGGRVDRFNKRFNIPGSK

Note that there are four cysteine residues (positions 16, 18, 37 and 40), which may potentially form disulfide bonds that would block the access of trypsin to the possible cleavage sites at positions 23 (K), 25 (R), and 39 (K). This assumption was confirmed by the observation of a peptide with $m/z$ 4288 (see inset of Fig. 9A), which likely arises from

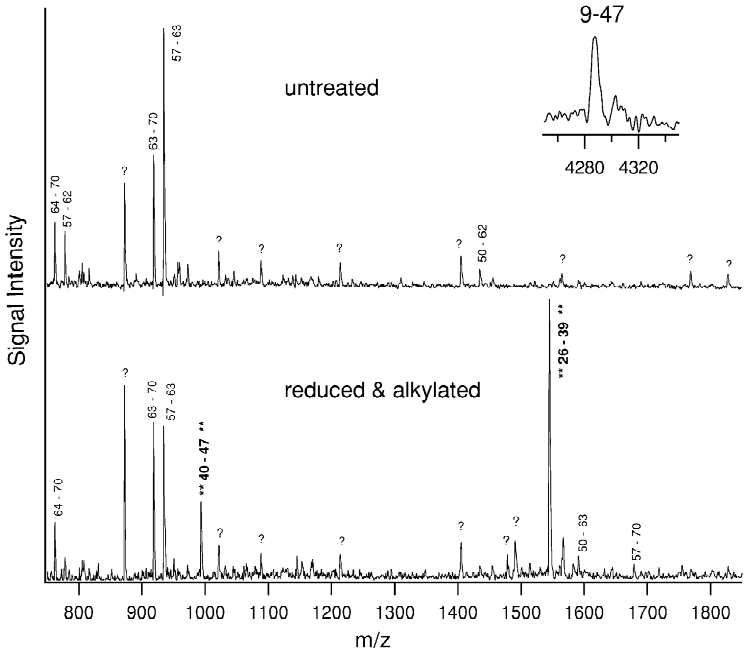


Fig. 9. MALDI mass spectra of in-capillary digestion of fraction #26 containing 50S Ribosomal Protein L31. (A) Peptides generated from trypsin digestion of protein. (B) Peptides generated from trypsin digestion of protein after reduction.

the tryptic peptide from residue 9–47 containing two disulfide bonds. After reduction and alkylation, two additional peptides were detected with $m/z$ corresponding to residues 26–39 and 40–47 which had been carbamidomethylated, while the peak at $m/z$ 4288 disappeared (Fig. 9B). The results from Figs. 9A and B provide the information to identify this protein as 50S ribosomal protein L31 with two disulfide bonds. Disulfide linkages have not been reported for this protein, neither in the SwissProt database nor in the original mass spectrometry work [31]. The molecular mass reported [31] was 7871.1 Da, which is 4 Da higher than our value. The molecular mass discrepancy very likely arises from the difference in sample workup. In our work, no reduction reagent was used in the cell extraction, HPLC fractionation, and protein digestion. Thus the disulfide bonds of the protein were intact for our mass spectrometric measurement for the first experiment (Fig. 9A). For the second experiment (Fig. 9B) reduction and alkylation was performed in-capillary so that only a few nanoliters of the original sample were modified. From the results shown in Fig. 9, we are located, although the experimental results indicate that the disulfide linkages are between residues #16–37 or 16–40 and between #18–37 or 18–40 rather than between 16–18 and 37–40 (MS–MS of selected peptides should provide the information required to determine the exact disulfide linkages, see below). Nevertheless, this example clearly demonstrates that the method of combining HPLC fractionation with MS peptide mass mapping can be very useful to provide information on protein post-translational modifications.

### *3.3. Prospective*

The examples shown here illustrate that the nanoliter sample handling technique can be used to concentrate dilute protein solutions from individual HPLC fractions and, by the combination of peptide mass mapping and multiple digestion experiments, we can use the technique to identify proteins separated by conventional HPLC. The application of this method for identifying low-mass proteome of microorganisms is an on-going research eflus5.1rchtc.2()-39a(tc.2(o3

differences in post-translational modifications of low-mass proteins under different cell growth conditions will be reported in the future.

In the area of technology development, we are currently focusing on research in combining the nanoliter sample preparation technique with MALDI MS–MS. The sequential digestion protocol, as discussed above, was found to be only applicable to simple protein mixtures. However, a number of HPLC fractions contained multiple protein components (>5 components per fraction). In these cases, it became difficult to deduce unambiguous peptide sequence information from the trypsin/LAP experiment. One solution to this problem that we are currently working on is to use multi-dimensional HPLC to separate proteins into individual components. This line of thinking is in analogy to the currently widely used gel electrophoresis–MS proteomics approach where if a 1D-gel fails to resolve many proteins for MS identification, a 2D-gel is then used for proteome display. In the gel electrophoresis experiment, the use of MS–MS to obtain sequencing information on individual peptides after in-gel digestion can greatly facilitate protein identification. For example, ESI MS–MS has been used widely to identify multiple proteins in 1D gel spots in proteomics [32]. Likewise, the development of microspot MALDI MS–MS should enable us to generate sequence information on the digested peptides from in-capillary concentrated proteins. MALDI MS–MS using a quadrupole/time-of-flight mass spectrometer or a time-of-flight/time-of-flight instrument has been demonstrated to be a very powerful technique for producing peptide sequence information [33–35]. We envision that a technique based on the use of multi-dimensional HPLC fractionation, nanoliter sample handling, sequential enzyme digestion, MALDI peptide mass mapping and/or MS–MS will be a valuable tool in proteomics, particularly for studying low-mass proteomes.

## Acknowledgements

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). BOK thanks the University of Alberta for a Ph.D. Dissertation Fellowship.

## References

[1] G. Corthals, S. Gygi, R. Aebersold, S.D. Patterson, in: T. Rabilloud (Ed.), Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods, Springer, Berlin, 2000, p. 197.
[2] M.A. Jeannot, J. Zheng, L. Li, J. Am. Soc. Mass Spectrom. 10 (1999) 512, and references therein.
[3] M. Fountoulakis, J.F. Juranville, D. Roder, S. Evers, P. Berndt, H. Langen, Electrophoresis 19 (1998) 1819.
[4] V.C. Wasinger, I. Humphery-Smith, FEMS Microbiol. Lett. 169 (1998) 375.
[5] E.T. Fung, V. Thulasiraman, S.R. Weinberger, E.A. Dalmasso, Curr. Opin. Biotechnol. 12 (2001) 65.
[6] M. Merchant, S.R. Weinberger, Electrophoresis 21 (2000) 1164.
[7] S.R. Weinberger, T.S. Morris, M. Pawlak, Pharmacogenomics 1 (2000) 395.
[8] K.Y. Dunlop, L. Li, J. Chromatogr. A 925 (2001) 123.
[9] Y. Dai, L. Li, D.C. Roser, S.R. Long, Rapid Commun. Mass Spectrom. 13 (1999) 73.
[10] Z. Wang, K.Y. Dunlop, S.R. Long, L. Li, Anal. Chem. (2002) in press.
[11] D.B. Wall, M.T. Kachman, S. Gong, R. Hinderer, S. Parus, D.E. Misek, S.M. Hanash, D.M. Lubman, Anal. Chem. 72 (2000) 1099.
[12] Z. Wang, L. Russon, L. Li, D.C. Roser, S.R. Long, Rapid Commun. Mass Spectrom. 12 (1998) 456.
[13] R.M. Whittal, B.O. Keller, L. Li, Anal. Chem. 70 (1998) 5344.
[14] Y. Dai, R.M. Whittal, L. Li, Anal. Chem. 71 (1999) 1087.
[15] R.M. Whittal, L. Li, Anal. Chem. 67 (1995) 1950.
[16] A.G. Kiceniuk, Masters Dissertation, University of Alberta, 2001.
[17] For example, see R. Kellner, T. Houthaeve, in: R. Kellner, F. Lottspeich, H.E. Meyer (Eds.), Microcharacterization of Proteins, Wiley–VCH, Weinheim, 1999, p. 97.
[18] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[19] M.R. Wilkins, E. Gasteiger, L. Tonella, K. Ou, M. Tyler, J.-C. Sanchez, A.A. Gooley, B.J. Walsh, A. Bairoch, R.D. Appel, K.L. Williams, D.F. Hochstrasser, J. Mol. Biol. 278 (1998) 599.
[20] For a historical review on CID, see R.G. Cooks, J. Mass Spectrom. 30 (1995) 1215.
[21] For example, see K. Biemann, in: J.A. McCloskey (Ed.), Methods in Enzymology: Mass Spectrometry, Academic Press, San Diego, CA, 1990, p. 455.
[22] B. Spengler, J. Mass Spectrom. 32 (1997) 1019.
[23] W. Staudenmann, P.D. Hatt, S. Hoving, A. Lehmann, M. Kertesz, P. James, Electrophoresis 19 (1998) 901.
[24] C. Korostensky, W. Staudenmann, P. Dainese, G. Gonnet, P. James, Electrophoresis 19 (1998) 1933.
[25] P.J. Sweeney, J.M. Walker, in: M. Burrell (Ed.), Methods in Molecular Biology: Enzymes of Molecular Biology, Humana Press, Totowa, NJ, 1993, p. 319.

[26] B. Thiede, B. Liebold-Wittmann, M. Bienert, E. Krause, FEBS Lett. 357 (1995) 65.
[27] A.S. Woods, A.Y.C. Huang, R.J. Cotter, G.R. Pasternack, D.M. Pardoll, E.M. Jaffee, Anal. Biochem. 226 (1995) 15.
[28] D.C. Schriemer, T. Yalcin, L. Li, Anal. Chem. 70 (1998) 1569.
[29] S. Hoving, M. Münchbach, H. Schmid, L. Signor, A. Lehmann, W. Staudenmann, M. Quadroni, P. James, Anal. Chem. 72 (2000) 1006.
[30] A. Doucette, L. Li, Proteomics 1 (2001) 987.
[31] R.J. Arnold, J.P. Reilly, Anal. Biochem. 269 (1999) 105.
[32] D. Arnott, W.J. Henzel, J.T. Stults, Electrophoresis 19 (1998) 968.
[33] A.V. Loboda, A.N. Krutchinsky, M. Bromirski, W. Ens, K.G. Standing, Rapid Commun. Mass Spectrom. 14 (2000) 1047.
[34] A. Shevchenko, A. Loboda, A. Shevchenko, W. Ens, K.G. Standing, Anal. Chem. 72 (2000) 2132.
[35] K.F. Medzihradszky, J.M. Campbell, M.A. Baldwin, A.M. Falick, P. Juhasz, M.L. Vestal, A.L. Burlingame, Anal. Chem. 72 (2000) 552.