Serially coupled microcolumn reversed phase liquid chromatography for shotgun proteomic analysis

Dingyin Tao1, 2, Guijie Zhu1, 2, Liangliang Sun1, 2, Junfeng Ma1, 2, Zhen Liang1, Weibing Zhang1, Lihua Zhang1 and Yukui Zhang1

1 Key Laboratory of Separation Sciences for Analytical Chemistry, National Chromatographic R&A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China
2 Graduate School of Chinese Academy of Sciences, Beijing, China

Microcolumn RPLC (µRPLC) is one of the optimum separation modes for shotgun proteomic analysis. To identify as many proteins as possible by MS/MS, the improvement on separation efficiency and peak capacity of µRPLC is indispensable. Although the increase in column length is one of the effective solutions, the preparation of a long microcolumn is rather difficult due to the high backpressure generated during the packing procedure. In our recent work, through connecting microcolumns of 5, 10, and 15 cm length via unions with minimal dead volume, long microcolumns with length up to 30 cm were obtained, with which 318 proteins were identified from proteins extracted from Escherichia coli by µRPLC-ESI MS/MS, and similar distributions of Mw and pI were found with single and various coupled microcolumns. Furthermore, by using MS/MS with improved sensitivity, with such a serially coupled 30 cm long microcolumn, 1692 proteins were identified within 7 h from rat brain tissue, with false positive rate (FPR) <1%. All these results demonstrated that serially couple microcolumns might be of great promising to improve the separation capacity of µRPLC in shotgun proteomic analysis.

Keywords:
µRPLC MS/MS / Serially coupled microcolumn / Shotgun proteome analysis

1 Introduction

In the postgenomic era, due to the complexity of proteomic samples, the development of novel separation techniques for large scale proteome analysis has been paid more and more attention [1]. Although 2D-PAGE, proposed by O’Farrell and Klose in 1975 [2, 3], has been playing an important role in proteome separation because of the extremely high resolution [4], the unavoidable disadvantages, including the challenges to identify low-abundance proteins [5–7], membrane proteins [8] and proteins with extreme pI and molecular weight (Mw) [9, 10], promote researchers to develop alternative approaches. Therefore, recently HPLC, no matter by 1-D [11] or multidimensional mode [12, 13], has been paid more and more attention. Compared to multidimensional platforms, although the peak capacity of 1-D system is limited, it is easy to construct and manipulate [14–17]. Therefore, to simplify the system and operation, it is imperative to improve the separation capacity of 1-D HPLC columns.

According to chromatographic theories, the longer the column length, the higher the effective column efficiency, the larger the peak capacity [18]. However, due to the pressure limitation during packing procedure, it is difficult to pack long HPLC columns. Therefore, serially coupled columns have been regarded as a good solution [19]. Berger and Wilson coupled ten conventional HPLC columns (4.6 mm id × 200 mm length) for supercritical fluid chromatography (SFC) analysis, and the column efficiency as high as 200 000 theoretic plates/m was obtained [20]. Due to the low back pressure of monolithic col-
Columns in proteome research. Which demonstrated the promising of coupled microcolumns, by which 1692 proteins were identified from high abundance proteins. With elevated analysis temperature, a peak capacity of 900 was obtained using long coupled HPLC columns. Herrero et al. coupled two C30 columns (4.6 mm × 250 mm length) for the separation of carotenoids, and the peak capacity was increased by 30% compared to that obtained using a single column [25]. However, in all above mentioned cases, the relatively great amount injection, low separation efficiency and large dead volume at the connectors for coupling conventional columns prevent them from further application.

For shotgun proteomic study, μRPLC is preferred due to the high resolution, good compatibility with MS/MS, and high orthogonality with other separation modes. Therefore, to avoid the above-mentioned problems caused by conventional columns, in our recent work, several short micro-columns were coupled serially by unions with the minimal dead volume, of which 1692 proteins were identified from the digest of proteins extracted from rat brain by 1-D μRPLC, which demonstrated the promising of coupled micro-columns in proteome research.

2 Materials and methods

2.1 Chemicals and materials

Peeksil and PEEK tubes were purchased from Upchurch Scientific (Oak Harbor, WA, USA). Frits were donated by Bona (Tianjin, China). PMSF was from Shanghai Sangon Biological Engineering Technology and Services (Shanghai, China). Urea was purchased from Invitrogen (Carosbad, CA, USA). DTT and iodoacetamide were from Acros (Morris plains, NJ, USA). Trypsin TPCK treated (bovine pancreas), Tris and cofeine were from Sigma (St. Louis, MO, USA). Methionyl-arginyl-phenylalanyl-alanine and Ultramark 1621 were donated by Thermo (San Jose, CA, USA). C18 silica particles (5 μm, 300 Å pore) were bought from Bona (Tianjin, China). The rat brain was ordered from Dalian Medical University (Dalian, China). Water was purified by a Milli-Q system (Millipore, Bedford, MA, USA). All other chemicals and solvents were analytical-grade.

2.2 Sample preparation

Protein mixture (1 mg/mL), dissolved in 8 M urea solution, was composed of BSA, β-lactoglobulin, α-casein, and cytochrome c with the same amount. The extraction of Escherichia coli whole cell lysate proteins was performed according to the previously described method [26]. In brief, E. coli (Strain BLT 5403) grown on LB culture medium was cultured at 37°C for 14 h, then at 4°C the mixture was centrifuged with the speed of 4300 × g for 10 min to precipitate cells. After washed with PBS for three times, 9 M urea together with 1 mM PMSF were added into the precipitates with the ratio of 2:1 v/w, followed by ultrasonication (Cole-Parmer, IL, USA) for 180 s. The resulting mixture was centrifuged at 20000 × g for 20 min, and the supernatant was collected as the soluble fraction and stocked at −80°C. The protein concentration of the soluble fraction was determined by Bradford method [27] using BSA as a standard protein. The whole brain was dissected and then sliced into small tissue pieces about 1 mm³ each. After washing three times with PBS, tissue pieces were transferred into individual Eppendorf tubes containing 8 M urea together with 1 mM PMSF and 1 mM Protease Inhibitor Cocktail Set I from Merck with the ratio of 2:1 v/w, followed by homogenate by Tissue Tearer from Biospec Products (Bartlesville, OK, USA) at 20 000 rpm for 3 min, then followed by ultrasonication for 50 s. All of the protein sample was reduced by DTT and alkylated by iodoacetamide. Then, the solution was diluted with water till the urea concentration was decreased to 1 M, and adjusted to pH 8.1 with Tris/HCl. Finally, trypsin was added with trypsin/protein w/w ratio of 1:50, and incubated at 37°C for 40 h. Before usage, the tryptic digest was desalted with a C18 solid-phase cartridge.

2.3 Column packing

Peeksil tubes with 300 μm id and 5–15 cm length were used to prepare μRPLC columns. One end of the tube was connected to a two-port union with a frit inside. The other end was connected to another union without a frit, which was further connected to a tank filled with C18 slurry. Then under the pressure of 5000–7000 psi, the column was packed overnight with a high-pressure pump (Lab Alliance, Scientific Systems, State College, PA, USA). Finally, the end without frit was disconnected from the union, and ready for further usage.

2.4 μRPLC separation

Experiments were performed on μRPLC-ESI-MS/MS, either by the hyphenation of MAGIC MS4 dual solvent delivery system (Michrom, Auburn, CA, USA) and a LCQ detector (Thermo) controlled by software of Paradigm Home Version 2.0.0.5, or by the coupling of a quaternary Surveyor pump with an LTQ linear IT MS equipped with an ESI probe Ion Max Source with microspray kit, controlled by Xcalibur software version 2.0 (Thermo). Two kinds of buffer solution were H2O with 2% ACN and 0.1% formic acid (A), and ACN with 2% H2O and 0.1% formic acid (B). The final flow rate after splitting was 5 μL/min.

For standard proteins mixture digests analysis with a 15 cm long serially coupled microcolumn (5 ± 10 cm), the
mobile phase was 0–10% buffer B for 10 min, 10–40% B for 50 min, 40–80% for 0.1 min, and maintained at 80% B for 10 min. For *E. coli* lysate digests analysis with 5 cm long microcolumn, the binary gradient for RP separation was 0–40% buffer B for 80 min, 40–80% B for 0.1 min, and maintained at 80% B for 10 min; with serially coupled microcolumn (5 + 10 cm) microcolumn, mobile phase was 0–40% buffer B for 180 min, 40–80% B for 0.1 min, and maintained at 80% B for 30 min; with a 30 cm long serially coupled microcolumn (5 + 10 + 15 cm), the mobile phase was 0–10% buffer B for 30 min, 10–40% B for 320 min, 40–80% for 0.1 min, and finally maintained at 80% B for 60 min. For rat brain tissue protein digest analysis, the mobile phase condition of 30 cm long serially coupled microcolumn (5 + 10 + 15 cm) was the same as those for *E. coli* analysis with the same column length.

2.5 Mass spectrometric analysis

Finnigan LCQDUO and LTQ XL IT mass spectrometers were used for ESI-MS/MS detection. When hyphenated with an LCQDUO detector, the effluents from the coupled microcolumns were sprayed directly into the electrospray source using a home-made interface without sheath or auxiliary gas. The ESI voltage was set up to 1.8 kV for LCQ and the capillary was heated to 150°C. The LTQ linear IT MS equipped with an ESI probe Ion Max Source with microspray kit. The temperature of the ion transfer capillary was set at 250°C, and the spray voltage was set at 3.2 kV. The sheath gas was highly purified nitrogen with flow rate of 3 L/min, and the normalized collision energy was set at 35.0%.

Total ion chromatograms and mass spectra were recorded on a PC with Xcalibur software version 1.4 for LCQDUO and version 2.0.7 for LTQ XL. Mass calibration and tuning were performed in the positive ion mode by direct infusion of a solution of caffeine, methionyl-arginyl-phenylalanyl-alanine and Ulramark 1621. For both LCQ and LTQ, an automated gain control function was used to manage the number of ions injected into the IT. Three microscans were set for each MS and MS/MS scan. All MS and MS/MS spectra were acquired in the data dependent mode. The MS was set as one full MS scan followed by three MS/MS scans on the three most intense ions for LCQ and one full MS scan followed by ten MS/MS scans on the ten most intense ions for LTQ. The dynamic exclusion function was set as follows: repeat count 2, repeat duration 30 s, and exclusion duration 180 s. The tandem mass spectra analysis and protein database searching were operated with the Bioworks software version 3.1 for LCQDUO and version 3.3 for LTQ.

2.6 Data analysis

Protein identification was performed using BioWorks Software 3.1 for LCQ and BioWorks Software 3.3.1 for LTQ with SEQUEST search program. The *E. coli* database was downloaded from a website (http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein&cmd=search&term=E.coli). The rat database was ipi.RAT.v3.26. fasta, and reversed sequences were appended to the database for the evaluation of false positive rate (FPR). Cysteine residues were searched as static modification of 57.0215 Da. Peptides searched using fully tryptic cleavage constraints and up to two internal cleavages sites were allowed for tryptic digestion. The mass tolerances were 2 Da for parent masses and 1 Da for fragment masses. The peptides were considered as positive identification if Xcorr was higher than 1.9 for singly charged peptide, 2.2 for doubly charged peptide, and 3.75 for triply charged peptides. Besides this, data from LTQ, ΔCn cutoff values were 0.38 to control the FPR less than 1%, determined by the calculation based on the reversed database. FPRs were calculated by using the following equation, $FPR = \frac{2n(\text{rev})}{n(\text{forw})}$, where $n(\text{forw})$ and $n(\text{rev})$ are the number of peptides identified in proteins with forward (normal) and reversed sequence, respectively.

3 Results and discussion

Since μRPLC has been widely applied into shotgun proteomic analysis, it is important to improve the effective separation efficiency and peak capacity. The serially coupled microcolumns with the minimal dead volume might be a good solution.

3.1 Microcolumns coupling and performance evaluation

In our recent work, two-port unions with minimal dead volume were applied for the connection of various microcolumns. As shown in Fig. 1, two frits were first put into both sides of the inner bore of the union to prevent the leakage of stationary phase out of the microcolumn. Then the microcolumns were coupled to each other by nuts. Thus, the dead volume at the connection point was only generated from the hole of the union, less than 13 nL, which could hardly affect the separation efficiency. By such means, long microcolumns could be successfully coupled to improve the separation capacity without challenging in the packing procedure. For example, a microcolumn (0.3 mm id) with the total length of 30 cm could be easily prepared by coupling three microcolumns with the length, respectively of 5, 10, and 15 cm. Furthermore, such long microcolumns could be easily disassembled and reassembled according to the complexity of samples.

To evaluate the reproducibility of μRPLC-MS/MS with serially coupled microcolumns, the digests of the mixture of BSA, β-lactoglobulin, α-casein, and cytochrome c digestion with equal amount were analyzed in three consecutive runs. After database searching, all proteins were identified with the varieties of sequence coverage and identified peptide number, respectively less than 4.5% and 2 (as shown in Table 1), which demonstrated that the reproducibility of such a system was quite good.
Figure 1. Schematic diagram of microcolumns coupling by two-port unions.

Table 1. Reproducibility of protein digests analysis with a serially coupled microcolumn (15 cm) by μRPLC-MS/MS (n = 3)

<table>
<thead>
<tr>
<th>Sample, 800 ng tryptic digest of the mixture of BSA, β-lactoglobin, α-Casein, and cytochrome c.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of unique peptides</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>No. of unique peptides</td>
</tr>
<tr>
<td>Sequence coverage %</td>
</tr>
</tbody>
</table>

Table 2. Comparison of protein identification with long single and coupled microcolumns (15 cm)

<table>
<thead>
<tr>
<th>Coupled column (5 + 10 cm)</th>
<th>Single column (15 cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backpressure</td>
<td>1150</td>
</tr>
<tr>
<td>Unique peptide number</td>
<td>307</td>
</tr>
<tr>
<td>Unique protein number</td>
<td>154</td>
</tr>
<tr>
<td>Protein number identified by two or more peptides</td>
<td>61</td>
</tr>
</tbody>
</table>

According to Eq. (1), it could be expected that, for a given \( \mu \), \( \eta \), \( \psi \), and \( d_p \), with the increasing of column length, the back pressure of the column should be increased as well. Therefore, to avoid too high back pressure generated with extremely long microcolumns, the effect of microcolumn length on pressure drop should be studied.

From Fig. 2 it could be seen that, at the flow rate of 5 μL/min, as expected from Eq. (1), a good linear relationship is obtained, which further demonstrates that the connecting of short microcolumns has no effect on the pressure property of a long coupled microcolumn. In addition, the increase in back pressure per cm of coupled microcolumn was ca. 78.8 psi, and with the microcolumn length of 45 cm, it reached 3500 psi with 5 μm C18 as the packing material. Although microcolumns with the length of 45 cm, or even longer, could be easily coupled, the back pressure was beyond the stable and safe operation range. Therefore, in our following experiments, the coupled microcolumn length was limited to 30 cm.

3.2 Effect of microcolumn length on pressure drop

The pressure drop across HPLC columns can be expressed by the following equation [28],

\[
\Delta P = \frac{\mu \times L \times \eta \psi}{d_p^2}
\]

(1)

where \( \mu \) is the linear velocity of the mobile phase, \( \eta \) the viscosity, \( d_p \) the particle diameter, and \( \psi \) is given by 180(1 + \( \omega \))(1 - \( \varepsilon \))^2/\( \varepsilon^2 \), where \( \varepsilon \) is the interstitial porosity and \( \omega \) is the volume ratio of the intraparticulate and interstitial void spaces.

3.3 Effect of sample injection amount

The amount of injected samples has great effect on the results of proteome analysis by μRPLC. In general, without the overloading of samples, the larger the injection amount, the higher the detection sensitivity, and the more identified proteins. Therefore, for a coupled microcolumn with the length of 30 cm (5 + 10 + 15 cm), the optimal sample injection amount was studied. From Table 3, it could be seen that, with the increase of injection amount from 2 to 50 μg, the identified unique peptide and protein number was
Figure 2. Effect of microcolumn length on pressure drop. Experimental conditions: mobile phase, ACN/H2O v/v 2:98; flow rate, 5.0 μL/min after splitting; room temperature; microcolumns, 300 μm id × 5 cm (single), 10 cm (single), 15 (5 + 10 cm), 30 (5 + 10 + 15 cm), and 45 cm (5 + 10 + 15 + 15 cm) length.

Table 3. Effect of injection amount on protein identification with a coupled 30 cm long microcolumn

<table>
<thead>
<tr>
<th>Injection amount (μg)</th>
<th>Unique peptide number</th>
<th>Unique protein number</th>
<th>Protein ratio identified with two or more peptides (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>61</td>
<td>47</td>
<td>19.14</td>
</tr>
<tr>
<td>6</td>
<td>153</td>
<td>92</td>
<td>35.87</td>
</tr>
<tr>
<td>10</td>
<td>302</td>
<td>167</td>
<td>33.53</td>
</tr>
<tr>
<td>25</td>
<td>449</td>
<td>241</td>
<td>34.43</td>
</tr>
<tr>
<td>50</td>
<td>604</td>
<td>318</td>
<td>32.70</td>
</tr>
</tbody>
</table>

increased from 61 to 604, and from 47 to 318, respectively. In addition, the ratio of proteins identified by two or more unique peptides was also increased from 19.14% to over 30%. All these results show that by coupling several short microcolumns, the sample injection amount of long microcolumns could be increased, thus the identification results of proteome samples could be obviously improved.

3.4 Analysis of proteins extracted from E. coli by various microcolumns

In order to evaluate the performance of the coupled microcolumns, microcolumns with the length of 5, 15 (5 + 10 cm), and 30 cm (5 + 10 + 15 cm) were applied to analyze the digest of proteins extracted from E. coli.

Under the optimal separation conditions, for the 5 cm long column, with 2 μg injection amount, 126 unique peptides, corresponding to 70 unique proteins, were identified within 90 min, as shown in Fig. 3A. For the coupled 15 cm microcolumn, with 5 μg injection amount, a total of 307 unique peptides, corresponding to 154 unique proteins, were identified within 210 min, as shown in Fig. 3B. With the further increase in microcolumn length to 30 cm, with 50 μg injection amount, 604 unique peptides, corresponding to 318 unique proteins were identified within 410 min, as shown in Fig. 3C.

The identified proteins by 5, 15, and 30 cm microcolumns were compared. From Fig. 4, it could be seen that most proteins were overlapped with such three columns, and
with the microcolumn length increased from 5 to 30 cm, although the total analysis time was increased from 90 to 410 min, the identified protein number could be increased by a factor of 4.54 (318/70). The overlapped protein number of a single 5 cm microcolumn and a coupled 30 cm column was 64. Only six unique proteins were identified by a 5 cm microcolumn, while 254 unique proteins were identified by a coupled 30 cm microcolumn, occupying 80% of all identified proteins (MS/MS results are listed in Supporting Information). The results reveal that the increase in microcolumn length by serially coupling could effectively enhance the separation capacity of μRPLC.

To prove the universality of μRPLC with serially coupled microcolumns, the physicochemical properties of the identified proteins from E. coli by such a technique were compared with those obtained by other commonly used methods [29], in which 670 proteins were identified within 38 h from E. coli by the combination of protein separation by Agilent 3100 OFFGEL Fractionator and peptides separation by microchip RPLC-MS/MS, which might be one of the best proteomic analysis results of E. coli.

Although less proteins were identified from E. coli by 1-D μRPLC with coupled microcolumns than those obtained by the 2-D separation [29], from Fig. 5 it could be seen that, similar physicochemical properties of the identified proteins were obtained. As shown in Fig. 5A, the ratios of acid proteins (pI less than 7) were respectively 82.84% according to [29] and 80.73% analyzed with a 30 cm long coupled column, indicating more acidic proteins were identified than basic proteins. So, it was in all the other cases. In addition, for 5, 10, and 15 cm long microcolumns, similar pI distribution of identified proteins were obtained. Especially in the range from pI 5 to 6, the protein ratios were of 52.09, 48.57, and 51.95%, respectively.

In addition, from the molecular weight distribution, as shown in Fig. 5B, it could be seen that the molecular weights of over 90% proteins were less than 80 kDa, focused in the range from 10 to 60 kDa, similar with that reported in [29]. For coupled microcolumns with different lengths, similar molecular weight distribution was obtained, especially in the range of less than 10 kDa and between 20 and 30 kDa.

### 3.5 Application of serially coupled microcolumn for real sample analysis

To further evaluate the applicability of serially coupled microcolumns in real sample analysis, a 30 cm long microcolumn was employed in the analysis of proteins extracted from rat brain by μRPLC-ESI-MS/MS with improved sensitivity, by changing LCQ MS (for above-mentioned data) to LTQ MS. After database searching, totally 1692 proteins were identified with FPR <1% determined by the calculation based on the reversed database.

To evaluate the separation power of coupled microcolumn in shotgun proteomics, some representative shotgun methods were compared, with details shown in Table 4. It could be seen that, generally in a single nano-LC-MS/MS
Table 4. Comparison of time-based efficiency with commonly used methods in shotgun proteomics analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample</th>
<th>Sample inject amount (µg)</th>
<th>Identified protein number</th>
<th>Total analysis time (h)</th>
<th>FPR</th>
<th>Time-based efficiency (protein number/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Nano-LC-MS/MS [18]</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>20</td>
<td>103</td>
<td>1.7</td>
<td>none</td>
<td>61</td>
</tr>
<tr>
<td>MudPIT [13]</td>
<td>Yeast</td>
<td>1350</td>
<td>1484</td>
<td>82.5</td>
<td>none</td>
<td>18</td>
</tr>
<tr>
<td>Dual trap-based 2-D</td>
<td>Human Jurkat</td>
<td>15</td>
<td>681</td>
<td>18</td>
<td>&lt;5%</td>
<td>38</td>
</tr>
<tr>
<td>nano-LC-MS/MS [30]</td>
<td>T-cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coupled microcolumn RPLC</td>
<td>Rat brain</td>
<td>50</td>
<td>1692</td>
<td>6.8</td>
<td>&lt;1%</td>
<td>249</td>
</tr>
</tbody>
</table>

analysis, hundreds of protein can be identified in 1.5–3 h, with the time-based efficiency of 60–120 proteins per hour. In 2-D LC-MS/MS system, more than 1000 proteins can be identified, however, it takes decades of hours and needs large sample inject amount. In MudPIT, 1484 proteins were identified in 80 h analysis, cost 1350 µg sample, and the time-based efficiency was 18 proteins per hour. However, by our method, in 1-D separation with a coupled microcolumn, 1692 proteins with FDR<1% were identified within 7 h, corresponding to 249 proteins per hour. Therefore, it could be concluded that no matter judged by total identified protein number or time-based efficiency, µRPLC with serially coupled microcolumns is prominent in shotgun proteomics research.

4 Concluding remarks

A simple method to prepare long microcolumns was developed by coupling short ones via two-port union with minimal dead volume. Through the analysis of the digests of proteins extracted from *E. coli*, it was demonstrated that the separation efficiency and peak capacity of µRPLC could be obviously improved by increasing the microcolumn length, while no extra back pressure and dead volume could be introduced. In addition, by comparison with proteins identified by other techniques, it was found that similar physicochemical properties were obtained. Furthermore, with high sensitive MS/MS as the detector, compared to other published methods, the time-based efficiency of our proposed techniques reached 249 proteins per hour in analyzing proteins extracted from rat brain. All these results demonstrate that the serially couple microcolumns for µRPLC might be of great promising in shotgun proteomic analysis, even by one multidimensional separation.

The authors are grateful for the financial support from National Nature Science Foundation (20435020), National Basic Research Programs of China (2007CB914100), and Knowledge Innovation Project of Chinese Academy of Sciences (K2006A16, KJCX2.YW.H09).

The authors have declared no conflict of interest.

5 References


