ORIGINAL PAPER

Improvement of the liquid-chromatographic analysis of protein tryptic digests by the use of long-capillary monolithic columns with UV and MS detection

M. H. M. van de Meent · G. J. de Jong

Abstract Optimisation of peak capacity is an important strategy in gradient liquid chromatography (LC). This can be achieved by using either long columns or columns packed with small particles. Monolithic columns allow the use of long columns at relatively low back-pressure. The gain in peak capacity using long columns was evaluated by the separation of a tryptic bovine serum albumin digest with an LC–UV–mass spectrometry (MS) system and monolithic columns of different length (150 and 750 mm). Peak capacities were determined from UV chromatograms and MS/MS data were used for Mascot database searching. Analyses with a similar gradient slope for the two columns produced ratios of the peak capacities that were close to the expected value of the square root of the column length ratio. Peak capacities of the short column were 12.6 and 25.0 with 3 and 15 min gradients, respectively, and 29.7 and 41.0 for the long column with 15 and 75 min gradients, respectively. Protein identification scores were also higher for the long column, 641 and 750 for the 3- and 15-min gradients with the short column and 1,376 and 993 for the 15- and 75-min gradients with the long column. Thus, the use of long monolithic columns provides improved peptide separation and increased reliability of protein identification.

Keywords Peak capacity · Monoliths · Long columns · Gradient liquid chromatography · Protein identification

M. H. M. van de Meent (✉) · G. J. de Jong
Division of Biomedical Analysis,
Department of Pharmaceutical Sciences, Utrecht University,
P.O. Box 80082, 3508 TB Utrecht, The Netherlands
e-mail: m.h.m.vandemeent@pharm.uu.nl

Introduction

Identification of unknown proteins is a key step in proteome analysis. The standard method of protein identification consists of enzymatic digestion of the protein(s), usually by using trypsin, followed by mass spectrometry (MS) analysis of the resulting digest. When analysing a single protein, e.g. from an excised 2D gel spot, a peptide separation is usually unnecessary. When analysing a digest of a mixture of proteins, the resolution of only a mass spectrometer is usually not sufficient. In those cases, a peptide separation is applied before MS detection. In shotgun proteomics, where a whole proteome is digested without prior protein fractionation, 2D liquid chromatography (LC) is the method of choice owing to the high complexity of the peptide mixtures. A drawback of multidimensional methods is the time frame, as typical analysis times for 2D-LC analysis of peptide mixtures range from several hours to more than 1 day [1–4]. 1D-LC methods can be used for separation of less complex samples (digests of fewer than 100 proteins), but this requires very efficient separations in combination with the additional separation power of MS detection. Such high-efficiency separations have been reported for analyses with long, packed columns, using (ultra-) high pressure (about 70 MPa) systems [5, 6].

Peak capacity is the primary parameter for evaluation of efficiency in gradient chromatography [7]. Peak capacity was first defined by Giddings [8] as the maximum number of peaks that can be separated by a phase system. This theory was later adapted for gradient chromatography by Horváth and Lipsky [9]. There are generally two approaches for optimisation of the peak capacity. The first is increase of the gradient length; gradients of up to 10 h have been reported on single columns [10]. However,

according to theory, peak capacity increases to a maximum and then decreases as the gradients become longer [11, 12]. The second approach is the use of longer columns, as the peak capacity increases linearly with the square root of the plate number and thus with the square root of the column length. Wang et al. [13] illustrated this by connecting several columns packed with a pellicular stationary phase. They showed that the ratio of the peak capacity and the square root of the column length was constant for columns ranging in length from 7.5 to 60 cm. The main limitation to simply increasing column length is the higher back-pressure. A possible solution for this problem is the use of monolithic columns. Monolithic columns have a higher permeability compared with packed columns, facilitating fast separations or the use of long columns [10, 14–16]. Wang et al. [13] reported a back-pressure of 28.5 MPa for a 60 cm×2.1 mm packed column at a linear flow of 0.94 mm/s. In contrast, other groups have obtained 3 times higher linear flow rates for monolithic columns of similar length [10, 17]. The higher permeability of monolithic columns makes it possible to operate long columns at higher flow rates while using conventional LC equipment. Luo et al. [10] separated a bacterial protein digest at a linear flow rate of 2.4 mm/s with a 70-cm monolithic column at 34.5 MPa (5,000 psi). Tolstikov et al. [17] analysed plant metabolomic extracts with a flow rate of 2.6 mm/s for a 60-cm column.

The quality of the analysis of a protein digest can be expressed in different ways. One way is to describe chromatographic efficiency in terms of parameters like peak width, peak capacity and resolution. Another way is to use the reliability of protein identification, like SEQUEST [18] or Mascot (http://www.matrixscience.com) [19] scores. Such an approach is also useful, as the mass spectrometer adds a second dimension to the separation, which is overlooked when only chromatographic efficiency is measured.

In this paper the evaluation of capillary monolithic silica columns of different lengths for the LC-UV-MS analysis of a bovine serum albumin (BSA) tryptic digest is described. Columns of 150- and 750-mm length were investigated using gradient times varying from 3 to 75 min. Chromatographic peak capacities, based on UV detection, and protein identification, based on Mascot scoring data of the MS detection, were determined as a measure for the efficiency of peptide separation.

Theoretical aspects

The peak capacity is defined as the maximum number of bands that will fit within a chromatogram with a resolution of $R_s=1.0$ [20]. In gradient LC, where peak width is about

constant throughout the separation, the theoretical peak capacity is given by Eq. 1:

$$PC = 1 + \frac{t_g}{w_{av}}, \qquad (1)$$

where t_g is the gradient time and w_{av} is the average baseline peak width. For large values of PC, this approaches t_g/w_{av}. Because sample peaks often do not occupy the entire length of the gradient, the sample peak capacity can then be defined as

$$PC^{**} = \frac{(t_z - t_a)}{w_{av}} = \frac{\Delta t}{w_{av}}, \qquad (2)$$

where t_a and t_z are the retention times of the first and last eluting peaks, respectively.

Wang et al. [13] investigated the effect of column length on peak capacity in packed columns and found that the peak capacity is proportional to $\sqrt{L}$ if the gradient slope is proportional to L. The gradient slope can be defined as

$$\frac{\Delta\phi t_0}{t_g}, \qquad (3)$$

where $\Delta\phi$ is the change in organic modifier fraction during the gradient ($0 \leq \phi \leq 1$) and t_0 is the column dead time. If the linear flow rate is constant, t_0 is only dependent on column length; therefore, t_g should be varied proportionally to column length in order to keep the gradient slope constant.

Experimental

Materials and reagents

BSA, trypsin (porcine, type IX-S, EC 3.4.21.4) and 1,4-dithiothreitol (DTT) were purchased from Sigma (St. Louis, MO, USA), and iodoacetamide (IAA) and NH_4HCO_3 from Fluka (Buchs, Switzerland). Acetonitrile was high-performance LC gradient grade (Biosolve, Valkenswaard, the Netherlands), and spectroscopy-grade trifluoroacetic acid (TFA) was obtained from Merck (Darmstadt, Germany). All solutions were prepared using water from a Milli-Q water-purifying system (Millipore, Bedford, MA, USA).

All reagents for the digestion of BSA were prepared in 200 mM NH_4HCO_3 buffer, pH 8. The tryptic digest was prepared as follows. A 100-µl aliquot of BSA stock solution (3.5 µg/µl in water) was set to pH 8 by adding 25 µl of 1 M NH_4HCO_3 buffer (pH 8). After addition of 25 µl of 10 mM DTT solution, the sample was incubated at 50 °C for 30 min to reduce disulfide bonds. After cooling to room temperature, 25 µl of 30 mM IAA solution was added and the sample was incubated in the dark for 60 min to alkylate the free thiols. Trypsin was dissolved in 10 µl of buffer to reach a trypsin-to-protein mass ratio of 1:50 in the final solution; the trypsin solution was added to the sample,

which was incubated at 37 °C overnight for 15 h. The digestion was stopped by addition of 15 μl of 10% TFA. The sample was diluted to 200 ng/μl (3 μM) with LC mobile phase A (water plus 0.05% TFA) and injected without further purification.

Apparatus and LC columns

All analyses were performed with an Agilent 1100 nanoLC system (Agilent Technologies, Waldbronn, Germany), consisting of a vacuum degasser, a binary Nano-Pump, a μ-well plate sampler and a column switching module with a trapping column in the 1–4 position of the six-port column-switching valve. The trapping pump was a Gynkotek model 480 (Gynkotek, Germering, Germany). Detection was performed by UV and MS detection, with the detectors connected in series. The UV detector was an MU 701 UV–VIS detector (ATAS GL International, Veldhoven, the Netherlands), equipped with an external optical-fibre flow

cell (6 nl, 3-mm light path); peptides were detected at 215 nm. The mass spectrometer was an Agilent LC/MSD Trap XCT (Agilent Technologies, Waldbronn, Germany) ion-trap mass spectrometer, equipped with an orthogonal electrospray ionisation (ESI) interface. The external flow cell of the UV detector allows minimal time delay and band-broadening between UV and MS detection.

The monolithic columns were provided by GL Sciences (Tokyo, Japan). The columns were a 150 mm×0.1 mm MonoCap for nano-flow C18-silica monolithic column and a 750 mm×0.2 mm MonoCap high resolution C18-silica monolithic column. For trapping of the digest a 5 mm× 0.3 mm column packed with 5 μm Zorbax 300 SB-C18 (Agilent Technologies, Waldbronn, Germany) was used.

Method and data analysis

LC solvent A was water plus 0.05% TFA (v/v); solvent B was acetonitrile plus 0.04% TFA (v/v). The trapping
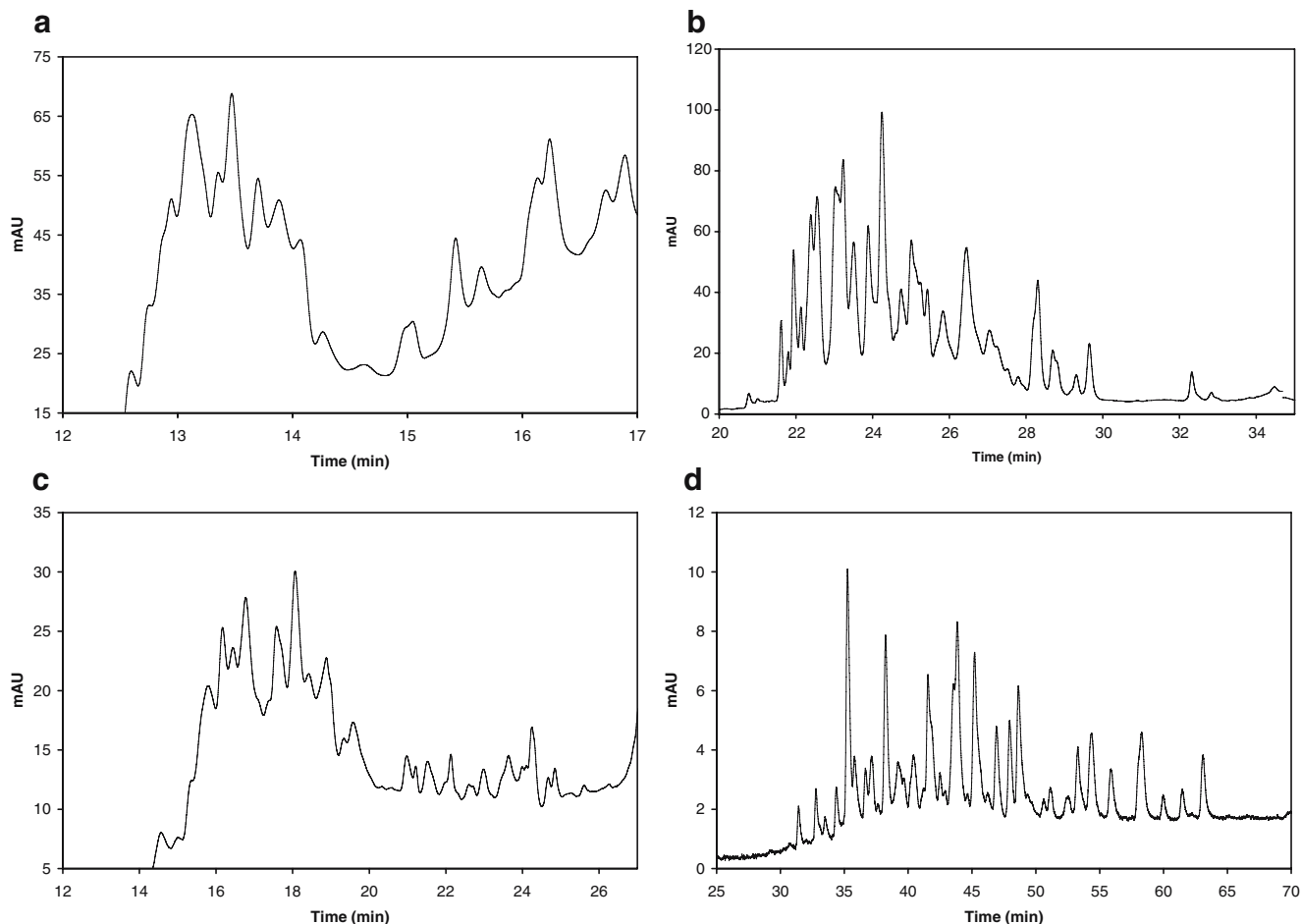

Fig. 1 Liquid chromatography (LC)–UV chromatograms of a tryptic bovine serum albumin (BSA) digest, separated by monolithic silica capillary columns of 150 mm×0.1 mm (**a**, **c**) and 750 mm×0.2 mm (**b**, **d**) using a gradient of 5–50% acetonitrile (0.04% v/v trifluoroacetic acid, TFA) in water (0.05% v/v TFA). **a** A 3-min gradient (15%/min); **b**, **c** 15-min gradient (3%/min); **d** 75-min gradient (0.6%/min). The run times include 5 min of trapping and 2.5 and 12.5 min of gradient delay time for the 150-mm and 750-mm columns, respectively

Table 1 Chromatographic parameters from liquid chromatography–UV analysis

t_g (min)	150 mm×0.1 mm			750 mm×0.2 mm		
	w_{av}	Δt	PC**	w_{av}	Δt	PC
3	0.24	3.05	12.6	0.28	3.16	11.2
15	0.44	11.0	25.0	0.39	11.6	29.7
75	1.27	37.6	29.6	0.77	31.7	41.0

t_g gradient time, w_{av} average width of seven to 12 peaks which were found to contain only a single peptide (based on mass spectrometry data), Δt elution window between the first and last eluting peptides, PC** sample peak capacity ($\Delta t/w_{av}$)

solvent was a mixture of 5% (v/v) solvent B in solvent A. After injection of the digest (0.25 µl for the 0.1-mm inner diameter column and 1.0 µl for the 0.2-mm inner diameter column), the sample was trapped on the trapping column at a flow rate of 5 µl/min. After 5 min, the trapping column was switched on-line with the separation column and the gradient was started. All separations were performed at room temperature using a gradient of 5–50% solvent B with gradient times varying from 3 to 75 min. MS spectra were acquired in the positive ion mode over the 400–2,000 m/z range, after which the two most intense ions (with a preference for doubly charged ions) were selected for fragmentation. MS/MS fragmentation spectra were acquired over the 100–2200 m/z range. An ESI spray voltage of -3 kV was used for all experiments.

The effect of separation efficiency on protein identification was evaluated using the Mascot search engine [19]. LC-MS/MS data were converted to the Mascot generic format (.mgf file) using the data-analysis software, and the .mgf files were searched against the MSDB database using Mascot's MS/MS ion search module. The database was searched for tryptic peptides from all entries in the database, allowing one missed cleavage per peptide and containing carbamidomethyl cysteine as a variable modification. Mass tolerances were set to default values: peptide mass tolerance ±2.0 Da, MS/MS tolerance ±0.8 Da.

Results and discussion

Liquid chromatography–UV analysis

Because of the difference in diameter, the 150 mm×0.1 mm and the 750 mm×0.2 mm columns were used with different flow rates. For the 150- and the 750-mm columns, the flow rates were 0.5 and 2.0 µl/min, respectively, resulting in a linear flow rate of 1.06 mm/s. Injection volumes were also proportional to the square of the column diameter, 0.25 µl of the digest for the 0.1-mm column and 1.0 µl onto the 0.2-mm column. During the gradient, the maximum back-pressure of the 750-mm column was below 20 Mpa, which is well below the manufacturers limit of 30 Mpa.

Figure 1 shows the LC-UV chromatograms of 3- and 15-min gradients run on the 150-mm column and 15- and 75-min gradients run on the 750-mm column. When the chromatograms of the analyses with similar gradient slopes are compared (Fig. 1a,b, and Fig. 1c,d), it is clear that an increase in column length improves the peptide separation. In order to quantify the efficiency of the separation, the sample peak capacity was calculated for all analyses. Because of the incomplete resolution of the digest, the peak capacity was estimated by using the average peak width of a selected number of peaks that appeared to contain only a single peptide. Using this method, we calculated peak capacities for all analyses and the results are summarised in Table 1. The peak capacities found for the short column are comparable to those found in the literature for similar columns [21, 22]. As expected, the peak capacities of the long column are higher than those of the short column, but they are relatively low compared with the values reported in [10]. However, when gradient time is taken into consideration, the difference is significantly less: PC**/t_g is 0.55 peaks per minute with the present system (75-min gradient on the 750-mm column) and 1.62 peaks per minute for the 260-min gradient reported in [10]. A possible negative effect on the resolution of our system is the use of a trapping column, filled with a different
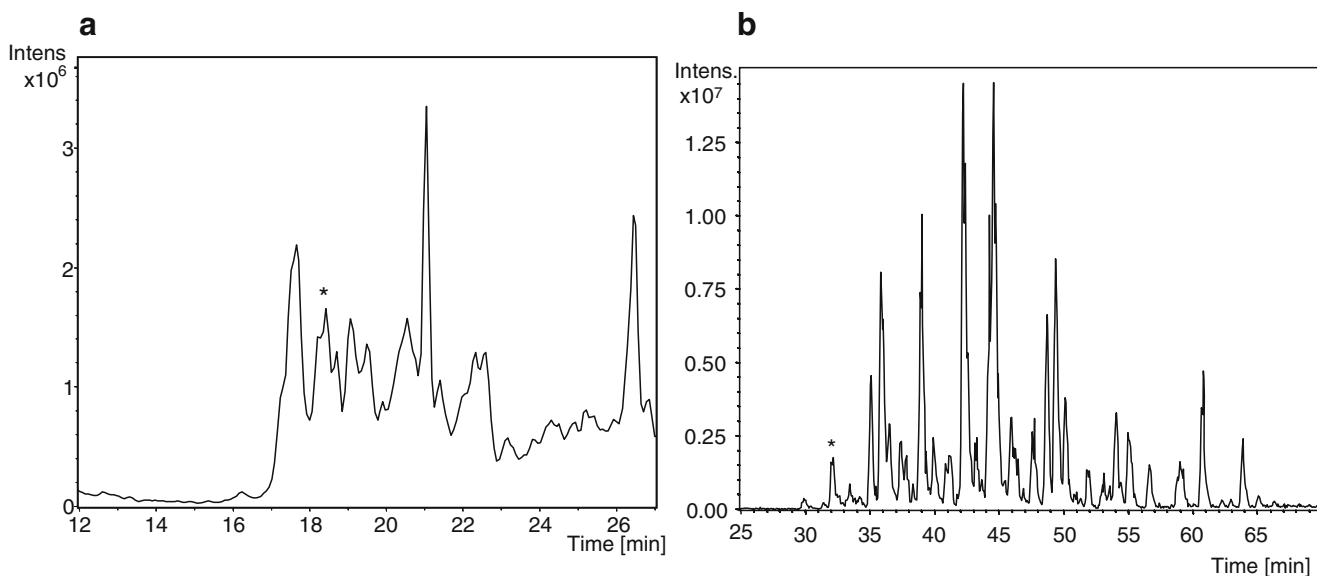
Table 2 Mascot® database search results

t_g	150 mm×0.1 mm			750 mm×0.2 mm		
	Mascot score	Peptides[a]	Coverage[b] (%)	Mascot score	Peptides[a]	Coverage[b] (%)
Infusion[c]	317	9	17	376	10	18
3 min	641	14	28	787	13	25
15 min	750	16	29	1,376	24	38
75 min	693	16	29	993	19	29

[a] Number of unique peptides identified by database search.
[b] Percentage of amino acid sequence covered by identified peptides.
[c] Direct infusion of the digest into the mass spectrometer.

a

b

Fig. 2 LC–mass spectrometry (MS) base peak chromatograms and mass spectra of tryptic BSA digest, separated by monolithic silica columns of 150 mm×0.1 mm (**a**) and 750 mm×0.2 mm (**b**), using a gradient of 5–50% acetonitrile (0.04% v/v TFA) in water (0.05% v/v TFA). **a** A 15-min gradient (3%/min); **b** 75-min gradient (0.6%/min). The run times include 5 min of trapping and 2.5 and 12.5 min of gradient delay time for the 150-mm and 750-mm columns, respectively

stationary phase, which probably has a selectivity that differs from that of the analysis column.

If chromatograms with the same gradient slope are compared (3- and 15-min gradients for the 150-mm column and 15- and 75-min gradients for the 750-mm column, respectively), the ratio of the peak capacities should be close to the square root of the column length ratio (2.24). For the 3-/15-min gradient pair this ratio is 29.7/12.6=2.37; for the 15-/75-min gradient pair the ratio is 41.0/25.0=1.64. For the short column, no increase in peak capacity is observed with a gradient of more than 15 min, but for the long column the peak capacity increases up to a 75-min gradient. A possible explanation can be found in the study of Stadalius et al. [11, 12], who demonstrated that peak capacity will increase with gradient time, until it reaches a maximum, after which it will even decrease. The gradient time at which this maximum is obtained is greater for longer columns.

Liquid chromatography–mass spectrometry analysis

To assess the influence of peptide separation on protein identification, MS/MS data were investigated by database-
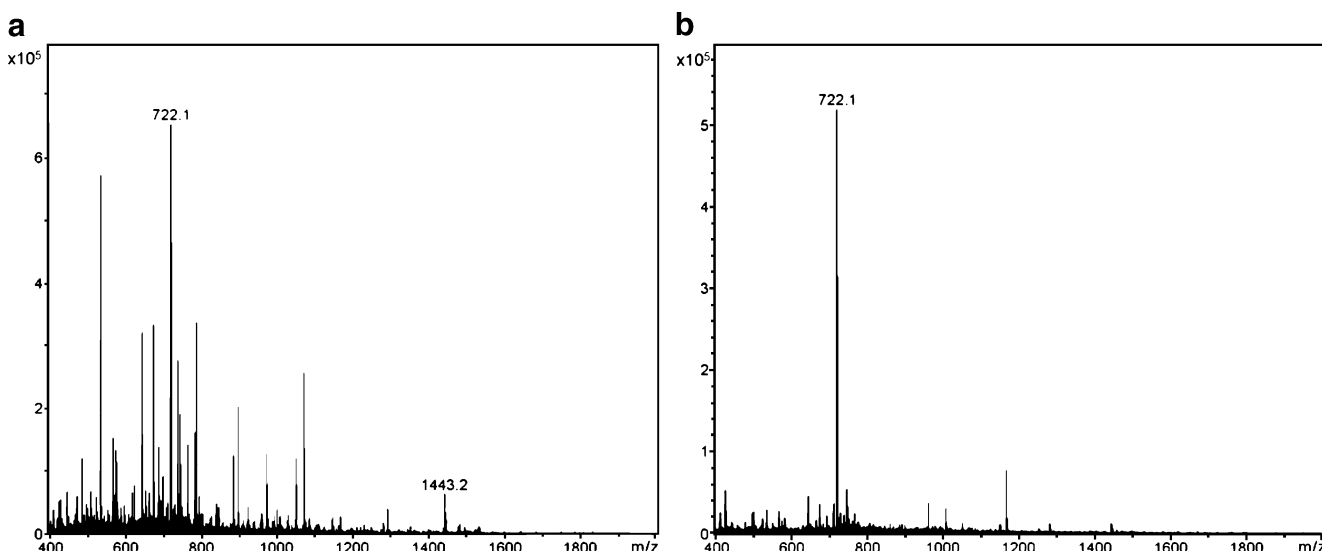
a

b

Fig. 3 Averaged mass spectra of peptide YICDNQDTISSK (*m/z* 722.32, M2H^{2+}) as identified from extracted ion chromatograms in the LC-MS analysis of a tryptic BSA digest. **a** A 150-mm×0.1-mm silica monolithic column, 15-min gradient of 5–50% acetonitrile (0.04% v/v TFA) in water (0.05% v/v TFA); **b** 750-mm×0.2-mm column, 75-min gradient

searching using the MS/MS ion search module from the Mascot search engine. The results are expressed as a Mascot score, the number of unique identified peptides and the percentage of the BSA amino acid sequence covered by these peptides (Table 2). These results were compared with the scores obtained for direct infusion of the BSA digest at the same flow rates as for the LC separations. The protein identification parameters for the infusion experiments were about similar for both flow rates. All database searches gave bovine albumin as the top protein match and the only other significant matches were albumins from other species.

Figure 2 shows base peak chromatograms of separations with the same gradient slope, a 15-min gradient for the 150-mm column and a 75-min gradient for the 750-mm column. The Mascot scores for the 750-mm column are higher than those for the 150-mm column. The average score per identified peptide is between 52 and 60 for the long column and between 42 and 51 for the short column. Combined with the larger number of identified peptides on the long column, this adds up to a higher Mascot score (Table 2).

Despite providing only a low resolution, even the shortest gradients cause a considerable increase in the reliability of protein identification. Increase in gradient length leads to cleaner mass spectra (Fig. 3) and a higher number of identified peptides and sequence coverage, compared with direct infusion. These numbers, however, decrease beyond a certain gradient time. This could possibly be attributed to the lower peak heights for longer gradient times. This result indicates that there is a gradient slope where an increase in chromatographic separation no longer improves protein identification.

Conclusions

The use of long silica-based capillary monolithic columns provides a clear advantage over use of shorter columns, i.e. an increase of chromatographic efficiency and reliability of protein identification. As expected from chromatography theory, a factor 5 longer column gives a 1.6–2.4 times increase in peak capacity for separations with similar gradient slope. The use of longer gradients also leads to an initial improvement of the protein identification score, but the score seems to have a maximum at longer gradient times.

While the use of longer columns for the separation of peptides has a clear advantage because of the gain in chromatographic efficiency, this also gives a longer analysis time. As maximum protein identification scores for rather simple digests are reached at relatively short gradient times, it is important to find a compromise between chromato-

graphic efficiency and analysis time. However, if the sample is more complex, the use of longer columns is more attractive as longer gradients are necessary to achieve sufficient separation.

In the near future, short and long columns of the same diameter (0.1-mm inner diameter) will be compared. Further improvement of the separation might be obtained by optimisation of the combination of the trap column and the analysis column. Moreover, the potential of longer monolithic capillary columns will be demonstrated by the analysis of more complex and real samples.

References

1. Delahunty C, Yates JR III (2005) Methods 35:248–255
2. Dai J, Shieh CH, Sheng Q, Zhou H, Zeng R (2005) Anal Chem 77:5793–5799
3. Washburn MP, Wolters D, Yates JR III (2001) Nat Biotechnol 19:242–247
4. Opiteck GJ, Jorgenson JW, Anderegg RJ (1997) Anal Chem 69:2283–2291
5. Shen Y, Smith RD (2002) Electrophoresis 23:3106–3124
6. Shen Y, Zhao R, Belov ME, Conrads TP, Anderson GA, Tang K, Pasa-Tolic L, Veenstra TD, Lipton MS, Udseth HR, Smith RD (2001) Anal Chem 73:1766–1775
7. Neue UD (2005) J Chromatogr A 1079:153–161
8. Giddings JC (1967) Anal Chem 39:1027–1028
9. Horváth CG, Lipsky SR (1967) Anal Chem 39:1893
10. Luo Q, Shen Y, Hixson KK, Zhao R, Yang F, Moore RJ, Mottaz HM, Smith RD (2005) Anal Chem 77:5028–5035
11. Stadalius MA, Quarry MA, Snyder LR (1985) J Chromatogr A 327:93–113
12. Snyder LR, Stadalius MA (1986) In: Horvath C (ed) High-performance liquid chromatography, advances and perspectives. Academic, New York, pp 195–312
13. Wang X, Barber WE, Carr PW (2006) J Chromatogr A 1107:139–151
14. Xiong L, Zhang R, Regnier F (2004) J Chromatogr A 1030:187–194
15. Hennessy TP, Boysen RI, Huber MI, Unger KK, Hearn MTW (2003) J Chromatogr A 1009:15
16. Leinweber FC, Schmid DG, Lubda D, Wiesmüller K, Jung G, Tallarek U (2003) Rapid Commun Mass Spectrom 17:1180–1188
17. Tolstikov VV, Lommen A, Nakanishi K, Tanaka N, Fiehn O (2003) Anal Chem 75:6737–6740
18. Eng JK, McCormack AL, Yates JR III (1994) J Am Soc Mass Spectrom 5:976–989
19. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Electrophoresis 20:3551–3567
20. Dolan JW, Snyder LR, Djordjevic NM, Hill DW, Waeghe TJ (1999) J Chromatogr A 857:1–20
21. Barroso B, Lubda D, Bischoff R (2003) J Protein Res 2:633–642
22. Rozenbrand J, van Bennekom WP, Unger KK, de Jong GJ (2006) Anal Bioanal Chem 385:1055–1061