ELSEVIER

Journal of Chromatography A, 913 (2001) 429–436

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Proteolysis of whole cell extracts with immobilized enzyme columns as part of multidimensional chromatography

Shihong Wang, Fred E. Regnier*

Department of Chemistry, Purdue University, Lafayette, IN 47907, USA

Abstract

This paper describes the efficacy of immobilized trypsin columns in the digestion of cellular extracts that contained thousands of proteins. Effectiveness of proteolysis was evaluated with extracts of *Escherichia coli* by size-exclusion chromatography. Immobilized trypsin columns were operated in either the continuous-flow or stopped-flow mode at temperatures ranging from ambient to 37°C with incubation times of 0–2 h. The results of these studies indicate that reduced and alkylated extracts of proteins from *E. coli* can be digested in 20 min when the enzyme column is operated at elevated temperature. The advantage of immobilized enzyme columns is that they can be easily incorporated into multidimensional separation systems for automated proteomics. © 2001 Published by Elsevier Science B.V.

Keywords: *Escherichia coli*; Immobilized trypsin; Immobilized proteins; Cells; Proteins; Peptides

1. Introduction

Cellular extracts and biological fluids contain thousands of proteins. The objective of proteomics is to identify large numbers, if not all these proteins simultaneously [1–3]. Recent advances in protein identification have led to the emergence of direct strategies to isolate and identify protein complexes involved in diverse biological processes ranging from transcription and cell cycle regulation to protein turn over [4–6]. With the increasing availability of genomic databases, this is now being done by comparing the mass of a few tryptic peptides derived from a protein with the molecular mass of peptides predicted from databases [7]. This means that proteolysis plays a central role in proteomics.

The general strategy in proteomics is to fractionate proteins by two-dimensional polyacrylamide gel (2D gel) electrophoresis [8] and visualize protein 'spots' by staining. Visualized proteins are then excised from the gel and digested by trypsin individually [9]. The established protocol for trypsin digestion is to add the enzyme at a concentration 50 times lower than that of the protein being digested and to incubate for 24 h at neutral pH. Proteolysis is known to be enhanced by reducing disulfide bridges with dithiothreitol (DTT) to open the protein structure to trypsin. Subsequent derivatization of sulfhydryl groups with an alkylating agent, such as iodoacetic acid, is used to prevent refolding [10]. A fingerprint of the protein is derived from the masses of peptide cleavage fragments and compared to those predicted from databases [7]. Excising and tryptic digesting the thousands of individual spots from a 2D gel is very labor intensive. Although this process can be automated through robotics, peptides are delivered to the mass spectrometer at a sub-optimum rate.

An alternative strategy is to digest all proteins in

*Corresponding author. Tel.: +1-317-494-1648; fax: +1-317-494-0359.

E-mail address: fregnier@purdue.edu (F.E. Regnier).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)00604-5

the sample before separation and peptide analysis [11–13]. The problem with this approach is that very large numbers of peptides are generated, too many to be separated in any single dimension separation system. This problem has been addressed by using affinity chromatography to either select a small number of peptides from each protein in the mixture or target-specific proteins. Affinity selected peptides are then transferred to a reversed-phase chromatography column where they are further resolved and delivered to either a matrix-assisted laser desorption ionization (MALDI) or electrospray ionization (ESI) mass spectrometer. Protein identification in this case is based on much smaller numbers of peptides and often requires more information about peptides, such as some amino acid composition and sequence. This approach depends on the ability to digest complex mixtures of proteins, such as cell lysates. One problem with samples of this origin and complexity is that some proteins can be very difficult to digest. In this case, one would not get a representative sample of the proteome by proteolysis. Another problem is that biological fluids and cellular extracts often contain proteolytic inhibitors [14–16].

This paper examines the question of how to digest whole cell extracts.

2. Experimental

2.1. Materials

DL-Dithiothreitol (DTT), iodoacetic acid, L-cysteine, TPCK-treated trypsin, urea, *N*-tosyl-L-lysyl chloromethyl ketone (TLCK), isopropyl β-thiogalactopyranoside (IPTG), phenylmethylsulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane (Tris base) and tris(hydroxymethyl)aminomethane hydrochloride (Tris acid) were purchased from Sigma (St. Louis, MO, USA). BactoAgar, tryptone peptone and yeast extract were from Difco (Sparks, MD, USA). Sodium phosphate, sodium chloride, and calcium chloride were from Mallinckrodt Baker (Phillipsburg, NJ, USA).

2.2. Escherichia coli cell culture

E. coli strain (BL21 BE3) was plated on an Luria–Bertani (LB) plate and incubated at 37°C overnight. A colony was selected and incubated in 5 ml LB medium for overnight and then added to LB medium at a 1:100 ratio and cultured at 37°C. IPTG was used to induce the culture. Cells were harvested 2 h after induction.

2.3. Preparation of E. coli lysate

E. coli cells were washed once with 0.85% NaCl and then centrifuged at 6000 *g* for 20 min. The pellet of cells was suspended in ~25 ml phosphate-buffered saline (PBS) with 50 m*M* DTT and 100 m*M* PMSF. Cell membranes were disrupted in a French Press at 8.16 MPa and cellular debris removed by centrifugation at 25 000 *g* for 15 min. The pellet was again suspended in ~10 ml PBS and centrifuged at 25 000 *g* for 15 min. In addition to the removal of particulates that would block columns, centrifugation serves the useful purpose of removing nucleic acids contained in cell nuclei. Supernatant from these two centrifugation steps was combined and will be subsequently referred to as the *E. coli* lysate. The concentration of lysate was estimated using the BCA Protein Assay method with a reagent kit from Pierce (Rockford, IL, USA). The total protein concentration of lysates was generally in the range of 10 mg/ml.

2.4. Free solution tryptic digestion

A total of 8 *M* urea was added to *E. coli* lysates to denature proteins. DTT was used at a 1:40 mole ratio for 1 h, then iodoacetic acid was added at 1:80 mole ratio for another hour and followed with the addition of cysteine at a 1:40 mole ratio for 30 min. After dilution with 20 m*M* Tris buffer to a final urea concentration of 2 *M*, TPCK-treated trypsin was added at a 1:100 (w/w) ratio. Two hours later, trypsin at a 1:100 (w/w) ratio was added again and the sample incubated at 37°C for 24 h. Digestion was terminated by adding TLCK in a slight molar excess over that of trypsin.

An additional 1.5-h incubation of samples with trypsin was found to be necessary to achieve more complete proteolysis. After digestion *E. coli* lysate was fractionated by size-exclusion chromatography and three fractions were collected. Trypsin was added to each of these three fractions separately at a

1:50 (w/w) ratio. The samples were incubated at 37°C for additional 1.5 h.

2.5. Trypsin column digestion

All chromatographic steps were performed using a BioCAD Micro-Analytical Workstation from PE biosystems (Framingham, MA, USA). Porosyzme immobilized trypsin cartridges (250×4.6 mm I.D.) were provided by Dr. Tim Nadler of PE Biosystems. Digestion buffer containing 50 m*M* Tris and 10 m*M* calcium chloride at pH 8.0 was used as the mobile phase at a flow-rate of 0.2 ml/min in immobilized enzyme columns. Digestion was achieved either under continuous flow or the flow was interrupted and the sample incubated for a fixed period of time at zero flow. The static incubation period ranged from 10 min up to 2 h. Sample incubation was executed either at room temperature or 37°C.

2.6. Size-exclusion chromatography

TSK-Gel G3000 SW columns (10 μm, 300×7.5 mm I.D., Tosohaas Montgomeryville, PA, USA) were used in the size-exclusion chromatography analysis of proteolysis. A mobile phase of 50 m*M* phosphate and 0.2 *M* NaCl at pH 7.5 was used at a flow-rate of 0.3 ml/min. Detection was by absorbance at 220 nm.

3. Results and discussion

3.1. Experimental design

The bacterium *Escherichia coli* has been intensively studied over the last 50 years. The genome of this organism suggests it is capable of expressing more than 4000 proteins [17]. This makes it an ideal model for these studies focusing on proteolysis of cell lysates. Two approaches to proteolysis of *E. coli* lysates were examined. One was the more traditional solution digestion method with soluble trypsin and the other an immobilized enzyme method more recently described for trypsin digestion of pure proteins [18,19]. Immobilized trypsin columns of 100–250×4.6 mm I.D. were used in these studies. Columns of this size can contain in excess of 10 mg of immobilized enzyme. The rationale for using immobilized trypsin in such large columns is as follows. First, immobilization minimizes trypsin auto-digestion. Second, sample handling and contamination can be reduced by coupling the immobilized enzyme column directly to other devices in the analytical train. Third, immobilization allows the use of a huge excess of enzyme. The amount of enzyme in the column exceeded that of the protein(s) being digested in many cases. It was reasoned that this could reduce reaction times an order of magnitude. And finally, the impact of inhibitors on proteolysis would be reduced. With such a large excess of enzyme there would still be sufficient enzyme remaining to achieve proteolysis with 75% inhibition.

A general characteristic of endopeptidases, such as trypsin, is that they rapidly reduce proteins in size by cleaving them internally. For this reason, efficacy of the various proteolysis protocols was accessed by measuring this size reduction with size-exclusion chromatography (SEC). A size-exclusion column with an exclusion limit greater than 100 000 was chosen so that proteins of 50 000 would be easily seen in the middle of the chromatogram. This assay strategy has both advantages and disadvantages. An advantage is that it is fast and very sensitive to the initial stages of proteolysis. The disadvantage is that it does not sense the end-point of proteolysis, particularly when using a column with an exclusion limit exceeding 100 000. Once the molecular mass of proteolytic fragments has dropped below 5000, peptides coelute. This means the size-exclusion assay used in this study cannot determine whether peptide fragments remain that could be further cleaved with trypsin.

3.2. Size-exclusion chromatography of E. coli lysate

SEC of the *E. coli* lysate shows a distribution of proteins ranging from less than 10 000 to greater than 100 000 (Fig. 1). Dividing the chromatogram into segments of >100 000, <10 000, and proteins from 10 000–100 000, the amount of material in these three fractions, based on absorbance at 220 nm, was 24, 18, and 58%, respectively (Table 1). It is likely that most of this absorbance comes from polypeptides in the >100 000 and 10 000–100 000

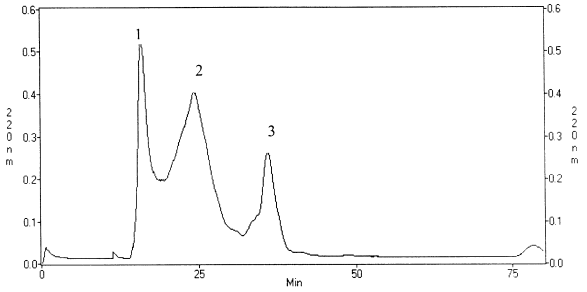

Fig. 1. Size-exclusion chromatogram of an *E. coli* lysate. The separation was achieved with a TSK 3000 SW column operated at 0.3 ml/min.

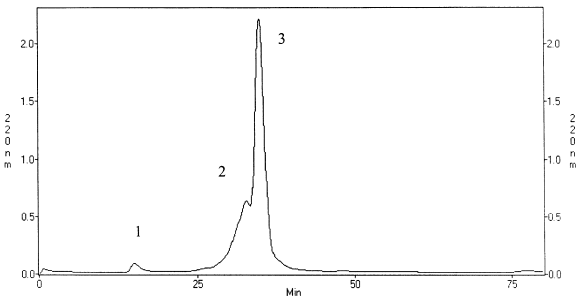

Fig. 2. Size-exclusion chromatogram of a proteolysis product of *E. coli* lysate produced by trypsin in a 24-h solution digestion at 37°C. Column and conditions as in Fig. 1.

fractions, but not in the fraction of <10 000. Because the lysate was chromatographed directly, it is likely to contain a substantial number of low-molecular mass, non-polypeptide substances that absorb at 220 nm. In light of the fact that the analytes of interest in this fraction are peptides of greater than 500 molecular mass, low-molecular-mass substances do not interfere in the final mass spectral analysis used for peptide identification and do not have to be eliminated.

According to the DNA database for *E. coli*, 10–20% of the proteins in the genome are of a size that they should be excluded from a TSK-Gel 3000 SW size-exclusion column (unpublished results). Of course this may not represent the distribution of expressed proteins in this case, but it is within the range of the observed value. Differences in the amount of protein in these molecular mass ranges can be seen between organisms and strains, but the distribution pattern is similar (unpublished results). The literature consistently reports the general distribution seen in Fig. 1 [20].

3.3. Solution tryptic digest of E. coli lysate

The most widely prescribed protocol for protein digestion with trypsin is a 24-h solution digestion at 37°C with a 50:1 (w/w) sample-to-enzyme ratio [10]. *E. coli* lysate was digested by this protocol and the digest examined by SEC. Again analyte absorbance was quantified by dividing the chromatogram (Fig. 2) into segments of >100 000, <10 000, and proteins from 10 000–100 000. It is seen in Table 1b that a substantial shift in molecular mass distribution occurred after proteolysis. Absorbance in the >100 000 segment dropped from 24 to 2%, while in the 10 000–100 000 segment it dropped from 57 to

Table 1
Distribution of proteins from *E. coli* lysate by size-exclusion chromatography (a), molecular mass distribution of digestion products from *E. coli* lysate after either a 24 h solution digestion or 24 h plus another 1.5 h solution digestion (b) and molecular mass distribution of digestion product from *E. coli* lysate after digestion on trypsin column at either room temperature or 37°C (c)

	Molecular mass range		
	M_r>100 000	10 000<M_r<100 000	M_r<10 000
Percentage within total product	24	58	18
24 h solution digestion (%)	2	36	62
24 h+1.5 h solution digestion (%)	0	2	98
Trypsin column digestion: at room temperature (%)	0	8	92
at 37°C (%)	0	0	100

36%. Production of lower molecular mass proteolysis products caused absorbance in the <10 000 fraction to increase from 18 to 62%. But these numbers are slightly misleading. The difference between the chromatograms in Figs. 1 and 2 clearly indicate substantial proteolysis in the 10 000–100 000 fraction to smaller fragments, albeit of >10 000. Analysis of the *E. coli* genome indicates that few fragments should be >10 000 in size. Obviously many of these fragments of >10 000 are incompletely digested.

These results are interpreted to mean that there are groups of proteins in the high-molecular-mass range along with lower-molecular-mass cleavage fragments that resist proteolysis. Incomplete proteolysis is unacceptable when the objective is to characterize all the proteins in a sample through identification of their tryptic fragments. The question is whether conditions can be found that will digest these polypeptides.

This problem of whether resistant polypeptides could be digested was examined by adding 1:50 (w/w) trypsin to fractions collected from the SEC column (Fig. 2) and digesting the mixture at 37°C for an additional 1.5 h (data not shown). At least 98% of the proteolysis products were in the <10 000 fraction (Table 1b) after this treatment. However, a small number of peptides shifted to an elution time beyond the total included volume.

Retention of polypeptides beyond the totally included volume in a size-exclusion column is due to adsorption at the surface of the chromatographic matrix. When this happens there is no relationship between analyte size and retention volume in SEC. Although the adsorption mechanism may either be due to electrostatic or hydrophobic interactions with the surface, hydrophobic effects are the most likely cause in this case. The mobile phase contained 0.2 *M* NaCl and would negate any ion-exchange effects. If true, this would mean that some of the polypeptides resulting from this final stage of proteolysis in solution are relatively hydrophobic. Unfortunately it is not possible to determine the molecular size of these partially adsorbed analytes.

3.4. Immobilized trypsin column digestion

E. coli lysate was pumped with a flow-rate of 0.2 ml/min through an immobilized trypsin column (250×4.6 mm I.D.) held at room temperature. It is seen in Fig. 3 that the effluent was separated into two peaks. This is attributed to the fact that the support on which the trypsin is immobilized is a porous matrix and some degree of size-exclusion chromatography was obtained. The second peak, eluting at roughly 27 min, is thought to be substances of <5000 that are eluting in the total inclusion volume of the column. These two peaks were collected and evaluated individually by SEC (Fig. 4). Fraction 1 from Fig. 3 produced tracing A in Fig. 4 and is seen to contain a small amount of product with a molecular mass of >10 000, some just under 10 000, and the rest being small molecules eluting in the totally included volume. The amount of material exceeding 10 000 was around 8% according to absorbance at 220 nm (Table 1c). Fraction 2 in Fig. 3 produced tracing B in Fig. 4 and in contrast contains low-

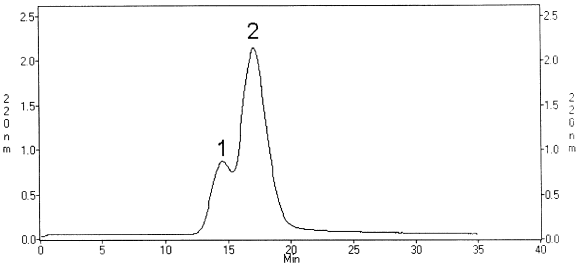

Fig. 3. Chromatogram of a proteolysis product of *E. coli* lysate after passage through an immobilized trypsin column at room temperature. Contact time in the immobilized enzyme column was 20 min.

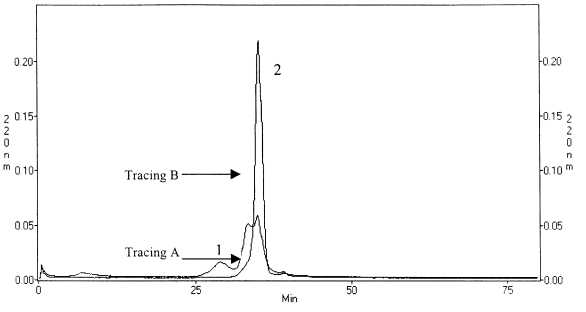

Fig. 4. Size-exclusion chromatogram of a proteolysis product of *E. coli* lysate subsequent to passage through an immobilized trypsin column at room temperature. Column and conditions as in Fig. 1.

molecular-mass product exclusively. From this we may conclude that the immobilized enzyme column serves both as an immobilized enzyme reactor and a size-exclusion chromatography column. Comparing data from the trypsin column and solution digestion protocols, it is seen that a 27-min passage through the immobilized trypsin column converted 92% of the lysate to products of less than 10 000, while only 62% of the proteins was converted to small peptides in the case of the 24 h solution digestion. Clearly the immobilized trypsin column is more efficient. However, Figs. 3 and 4 show that even the immobilized enzyme column failed to totally digest all the polypeptides in the lysate.

Efforts to increase the degree of proteolysis in the immobilized trypsin column included increasing the reaction time and increasing the temperature. Reaction time was varied by interrupting flow through the column for times ranging from a few minutes to 2 h. It was observed that increasing reaction time from 15 to 120 min in the immobilized enzyme column did not increase the degree of proteolysis (Fig. 5). There was still a small amount of undigested material as judged by the size-exclusion separation provided by the immobilized enzyme column. Because there was no observable difference in proteolysis between the constant and stopped flow modes, the continuous flow mode was used throughout the rest of this study because it was faster and easier to incorporate into automated processes.

Elevating the reaction temperature to 37°C increased the degree of proteolysis in the immobilized enzyme column as judged by the elimination of the peak doublet from the effluent (data not shown). It should also be noted that the peak is broader. This is probably the result of the continuous flow mode of incubation. Recalling that the immobilized enzyme column has size-exclusion properties, the mobility of faster moving proteins will decrease continuously as they are converted to lower molecular mass, slower moving peptides during passage through the column. This would have the net effect of smearing or band spreading the sample. Although effluent from the immobilized enzyme column is a single peak, the higher resolution SEC column still showed two peaks (data not shown) with retention times of 34.5 and 35.7 min. According to Figs. 2 and 4, the totally included volume of this column is 35.1 min. Again, the exact molecular mass of these peptides is not clear. The two peaks could be due as much to differences in hydrophobicity as in molecular mass.

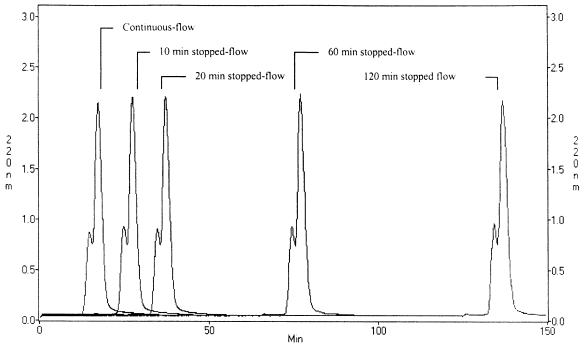

Fig. 5. Chromatogram of proteolysis products of *E. coli* lysate after passage through an immobilized trypsin column at room temperature. Contact time in the immobilized enzyme column varied from 20 min up to 2 h.

Data from all the above experiments are summarized in Fig. 6. It is seen by comparing the chromatogram from immobilized enzyme digestion at 37°C with those from solution digests (Fig. 2) that the efficiency of the immobilized enzyme column is much higher than obtained in solution digestion. One hundred percent of the protein was converted to peptides in less than 20 min in the immobilized enzyme column. Although it is not known that all these peptides are completely digested, it is apparently the best that is possible. The greater efficiency of the immobilized enzyme column is attributed to the high enzyme to protein ratio in the column. The manufacturer indicates that a 250×4.6 mm I.D. column would contain at least 20 mg of trypsin, but it is not known that all the enzyme is active. These results also indicate that proteolytic inhibitors are not a serious problem, even though crude cell extracts are known to contain trypsin inhibitors [14–16]. Again this is probably due to the huge excess of trypsin.

Finally, it is concluded that the immobilized enzyme column and solution digestion protocol give some products that are different, even when very long reactions times were used in solution digestion. The more hydrophobic peptides seen when fresh trypsin was added a second time (data not shown) were never seen in the immobilized enzyme digests. It is likely that these more hydrophobic peptides are

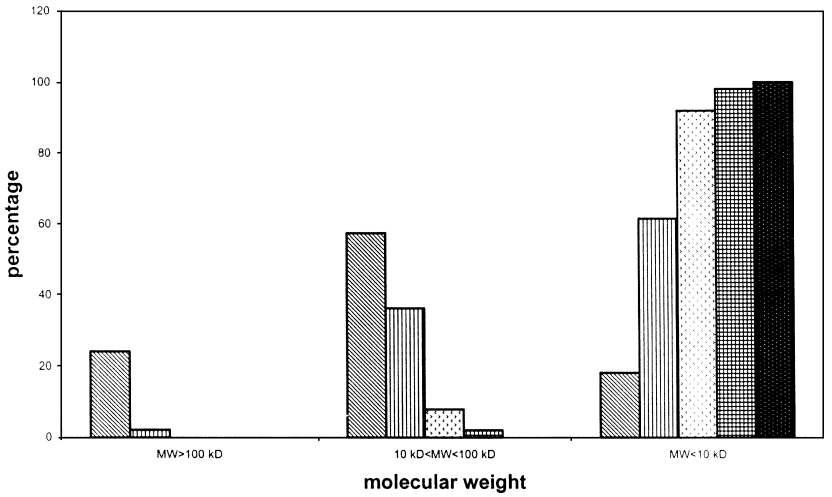

Fig. 6. Distribution of molecular mass in *E. coli* lysate and trypsin digests of the lysate produced either in solution or with an immobilized enzyme column. (▨) Undigested *E. coli* lysate; (▯) products produced after 24 h solution digestion of *E. coli*; (▫) products produced after digestion on trypsin column at room temperature; (⊞) 24 h digestion plus additional 1.5 h solution digestion; (▩) products produced after digestion on trypsin column at 37°C.

incompletely digested products of proteolysis, sometimes referred to as limit peptides [21]. These peptides are probably formed in the immobilized enzyme column too but are further digested before elution from the column.

4. Conclusions

It may be concluded that *E. coli* lysate can be effectively digested in 20 min by an immobilized trypsin column operated at 37°C. Furthermore, beyond being of higher efficiency than solution digestion with trypsin, some of the products are different. More hydrophobic peptides seen in solution digests were not observed in the effluent from the immobilized enzyme column. Finally, it may be concluded that proteolytic inhibitors are not a problem with larger immobilized enzyme columns in the case of *E. coli* lysate.

Acknowledgements

The authors gratefully acknowledge financial support from National Institute of Health grant number GM59996.

References

[1] L. Anderson, in: 2nd International Conference on Proteomics, Coronado, CA, 1998.
[2] P. Roepstorff, Curr. Opin. Biotech. 8 (1997) 6.
[3] K. Kivirikko, R. Myllyla, T. Philajaniemi, in: J. Harding, M. Crabbe (Eds.), Post-translational Modification of Proteins, CRC Press, Boca Raton, FL, 1992, p. 1.
[4] C. Rachez, B.D. Lemon, Z. Suldan, V. Bromleigh, M. Gamble et al., Nature 398 (1999) 824.
[5] G. Neubauer, A. King, J. Rappsilver, C. Calvio et al., Nat. Genet. 20 (1) (1998) 46.
[6] W. Shou, J.H. Seol, A. Shevchenko, C.I. Baskerville et al., Cell 97 (1999) 233.
[7] J.R. Yates, J. Mass. Spectrom. 33 (1998) 1.
[8] P.H. O'Farrel, J. Biol. Chem. 250 (1975) 4007.
[9] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (5) (1996) 850.
[10] K. Stone, M. LoPresti, K. Williams, Tech. Protein Chem. (1989) 377.
[11] M. Geng, J. Ji, F. Regnier, J. Chromatogr. A. 870 (2000) 295.
[12] J. Ji, A. Chakraborty, M. Geng, F. Regnier et al., J. Chromatogr. B 745 (2000) 197.
[13] S.P. Gygi, B. Rist, S. Gerber, F. Turecek, M. Gelb, R. Aebersold, Nat. Biotechnol. 17 (10) (1999) 994.

[14] V.V. Mosolov, Bioorg. Khim. 24 (5) (1998) 332.
[15] M. McGrath, S. Gillmore, R. Fletterick, Protein Sci. 4 (2) (1995) 141.
[16] M. AcGrath, W. Hines, J. Sakanari et al., J. Biol. Chem. 266 (10) (1991) 6620.
[17] http://www.ncbi.nlm.nih.gov/genbank/genomes/bacteria/ecoli
[18] J. Lei, D. Chen, F. Regnier, J. Chromatogr. A 808 (1998) 121.
[19] T. Nadler, C. Blackburn, J. Mark, N. Gordon, F. Regnier, G. Vella, J. Chromatogr. A 743 (1) (1996) 91.
[20] P. Demirev, Y. Ho, V. Ryzhov, C. Fenselau, Anal. Chem. 71 (1999) 2732.
[21] S. Rao, J. Hennessey, G. Scarborough, Anal. Biochem. 173 (2) (1988) 251.