



Two-dimensional nano-liquid chromatography–mass spectrometry system for applications in proteomics

Edgar Nägele*, Martin Vollmer, Patric Hörth

Agilent Technologies Deutschland GmbH, Hewlett-Packard-Strasse 8, D-76337 Waldbronn, Germany

Abstract

This work demonstrates the development of a method for the analysis of complex proteome samples by two-dimensional nano-liquid chromatography–mass spectrometry. This approach includes strong cation-exchange, sample enrichment, reversed-phase chromatography and nanospray ion trap mass spectroscopy with data dependent tandem mass spectrometry spectra acquisition, and subsequent database search. The new methodology was first evaluated using standard protein digest samples. Finally, data for the analysis of a total *Escherichia coli* proteome are provided.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Proteomics; *Escherichia coli*; Two-dimensional liquid chromatography; Proteins

1. Introduction

The proteome, which can be defined as the complete protein set of a cell, derived from a certain time-point and condition of a cell, a microorganism, or a body fluid is now the focus of many international research groups and pharmaceutical companies. As the sequenced human genome and expression levels of mRNA do not describe the current state of a living cell in sufficient detail due to the broad variety of posttranslational modifications, scientists discovered the importance of analyzing the actual protein composition of a biological system in order to understand complex cellular processes and networks. Proteomics as a scientific discipline emerged with the purpose to identify rapidly complex protein patterns in a comprehensive manner [1–3]. The methodology needed to achieve this ambitious goal

must also be powerful enough to detect subtle quantitative and qualitative differences of a protein profile in order to finally identify target proteins and modified variants [4,5]. Such comprehensive proteome characterization will give new insight in cellular responses for disease pathogenesis such as carcinogenesis, for development, as well as ageing, drug action and environmental damage.

In order to identify proteins from a complex mixture of 5×10^3 to 5×10^4 with a dynamic range of at least 10^5 [6] it will be crucial to develop technologies with extremely good resolving power on one hand, and with extraordinary sensitivity on the other hand. It is obvious that these challenging tasks will not be achieved with a single analytical technology, but with a combination of separation and detection techniques. The current method of choice for separating complex proteomic samples is still two-dimensional gel electrophoresis (2D-GE) [6–8] due to its high resolving power for proteins. During the 2D-GE procedure the protein mixture is separated by isoelectric focusing in the first dimension

*Corresponding author. Tel.: +49-7243-602663; fax: +49-7243-602501.

E-mail address: edgar_naegele@agilent.com (E. Nägele).

and in the second dimension by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). After in gel digestion of target proteins, identification is usually achieved by matrix-assisted laser desorption ionization time-of-flight (MALDI–TOF) MS or LC electrospray ionization (ESI) ion trap MS.

Despite the recent progress of 2D-GE in terms of reproducibility, automation and quantitation [9–11] 2D-GE still suffers from some major drawbacks. This holds true especially for low abundance proteins, membrane proteins, proteins with extreme isoelectric points (*pI*) values and very large and very small proteins, which account for a high percentage of cellular proteins and which are often the most promising targets for drug development or disease diagnostics [12].

In order to overcome these major obstacles in 2D-GE, liquid chromatography based separation techniques directly coupled to MS detection were developed with the goal of obtaining a comparable resolution as in 2D-GE. The orthogonal combination of different HPLC techniques (2D-LC or multidimensional LC) for proteins and especially for peptide mixtures has been described in a variety of reports [13]. In general, peak capacity for a multidimensional separation is calculated by multiplying the peak capacity of the single separation steps [14]. By combining two different LC techniques with individual peak capacities between 50 and 100, it will be theoretically possible to achieve total capacities of 2500–10 000 peaks.

Two different approaches have been addressed in order to separate and resolve complex protein expression patterns: (1) the multidimensional separation of proteins; and (2) the combination of different techniques adapted for separation on the peptide level [13]. The latter approach can be more easily automated since the digestion of the complete protein content of a sample takes place at the beginning of the workflow. In addition, it is also more suitable to handle large proteins and unspecific adsorption to separation devices and tubings is much less pronounced. The initial complexity of the starting sample however is increased significantly by increasing the total amount of different compounds by one to two orders of magnitude.

For the separation of complex peptide mixtures the

combination of several orthogonal techniques has been used: (i) reversed-phase liquid chromatography (RPLC)–capillary electrophoresis [15,16]; (ii) size-exclusion chromatography (SEC)–RP-LC [17]; (iii) strong cation-exchange chromatography (SCX)–RPLC [18,19]; and (iv) affinity chromatography (AC)–RPLC [20,21]. While the latter approach is mainly applied for more specific functional approaches like phosphopeptide analysis [22], SCX–RPLC with MudPIT technology in contrast has successfully been applied for the analysis of comprehensive protein expression profiles from yeast *Saccharomyces cerevisiae* or *Plasmodium falciparum* [23,24].

Here we show the development and evaluation of an alternative SCX–RPLC 2D-LC–MS method. The concept of nanoflow RP HPLC with extremely low flow rates followed by nanospray ionization MS–MS provides excellent sensitivity. This second separation dimension is coupled to SCX chromatography as first dimension by an enrichment column, ensuring that sample fractions eluting from SCX are concentrated and properly channeled in the nanoflow path of the second dimension. This methodology was first tested with a tryptic digest of 10 defined proteins and further evaluated with a complex total lysate from *Escherichia coli*.

2. Experimental

2.1. Chemicals

All HPLC grade solvents and chemicals were purchased from Merck (Darmstadt, Germany).

The tryptic digests of 10 bovine proteins (hemoglobin, beta-lactoglobulin, chymotrypsinogen, carbonic anhydrase, carboxypeptidase A, glutamic dehydrogenase, bovine serum albumine, serotransferrin, lactoperoxidase, catalase) used for this work were purchased from Michrom BioResearch (Auburn, CA, USA).

2.2. Equipment

For this work the Agilent “Nanoflow Proteomics Solution” (Agilent Technologies Deutschland, Wald-

bronn, Germany) consisting of the following components was used:

- (i) 1100 Series nano pump with micro vacuum degasser;
- (ii) 1100 Series quaternary pump with vacuum degasser;
- (iii) 1100 Series thermostatted micro well-plate autosampler;
- (iv) 1100 Series thermostatted column compartment with six-port/two-position valve or Agilent six-port/two-position micro switching valve box;
- (v) Agilent 1100 Series MSD ion trap SL with nano electrospray ion source (Bruker Daltonics, Bremen, Germany);
- (vi) software: Agilent ChemStation A09.03 and ion trap software 4.1; and
- (vii) MASCOT software (Matrix Science, London, UK) for database search.

The “Nanoflow Proteomics Solution” consists of a unique instrument set-up. The micro well-plate autosampler was connected directly to the quaternary pump. From the pump the hydraulic flow passed through the micro six-port/two-position valve of the micro well-plate autosampler and further through a sample loop and needle (Fig. 1). The SCX column was positioned at the location of the seat capillary and the column outlet flow was further directed to the six-port valve of the autosampler. This valve was connected with a capillary to the second micro six-port/two-position valve in the column compartment or to a separate micro valve box. Sample eluting from the SCX column was concentrated on top of a C_{18} enrichment column, which was mounted be-

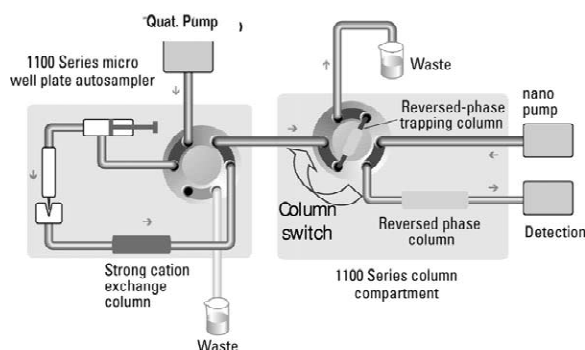


Fig. 1. Arrangement of LC modules and micro valves to perform 2D-LC.

tween two ports of this second valve and the flow was further directed to waste. The nano pump was also connected to this second valve and the flow from this pump passed through the analytical column into the MS ion trap.

2.3. Liquid chromatography

The following separation conditions and columns were used throughout this study:

- (i) SCX column: polysulfoethyl aspartamide, 50 mm×0.32 mm, 5- μ m particles (PolyLC, Columbia, USA);
- (ii) enrichment column: Zorbax 300 SB C_{18} , 5 mm×0.3 mm, 5- μ m particles (Agilent Technologies);
- (iii) RP column: Zorbax 300 SB C_{18} , 150 mm×75 μ m, 3.5- μ m particles (Agilent Technologies);
- (iv) nano pump mobile phase: A=water+0.1% formic acid; B=acetonitrile+0.1% formic acid;
- (v) nano pump gradient: 0 min, 5% B; 8 min, 5% B; 12 min, 15% B; 72 min, 55% B; 74 min, 75% B; 75 min, 75% B; 75.01 min, 5% B; stop time, 90 min; post time, 10 min;
- (vi) nano pump flow-rate: 300 nl/min;
- (vii) quaternary pump mobile phase: water+3% acetonitrile+0.1% formic acid;
- (viii) quaternary pump flow gradient: 0.1 ml/min, 0 min; 0.05 ml/min, 0.5 min; 0.01 ml/min, 0.51 min; 0.01 ml/min, 8 min; 0.005 ml/min, 8.01 min; 0.005 ml/min, 9.9 min; 0 ml/min, 10 min; 0 ml/min, 85 min; 0.005 ml/min, 85.01 min;
- (ix) valve switch: enrichment column into the nano flow path at 8 min, enrichment column out of the nano flow path at 85 min; and
- (x) salt gradient steps: 20- μ l injections of 10–100 mM (10 mM increments); 125 mM, 150 mM, 200 mM, 300 mM, 500 mM, 1000 mM KCl or $HCOONH_4$.

2.4. Mass spectrometry

Mass spectrometry was performed with the following instrumentation parameters and methodology:

- (i) source and inlet system: on-line nano LC with nano electrospray ion source; drying gas (N_2)

- flow, 5 l/min; temperature, 225 °C; capillary, 1700 V; capillary exit, 135 V; skimmer, 40 V;
- (ii) mass analyzer: octopole1, 12 V; octopole2, 2 V; trap drive, 80 V; collision gas, He; target, 40 000; max. accu. time, 150 ms; ion mode positive;
- (iii) detection: dynode, 7 kV; electron multiplier, 1850 V.

2.5. Sample preparation

2.5.1. Standard tryptic digests

The lyophilized commercially obtained tryptic digests of 10 bovine proteins were dissolved in water containing 5% acetonitrile and 0.1% formic acid to the initial stock solution with a concentration of 1 pmol/ μ l. Ten microliters of the stock solution from each digest was mixed to the final test solution used for this work. The concentration of each digested protein was 100 fmol/ μ l.

2.5.2. Tryptic digest of the *E. coli* proteome

E. coli K12 was grown at 30 °C in LB-Medium until mid-log phase. Cells were harvested and spun down in a Stratos Biofuge (Heraeus Instruments, Hanau, Germany) for 15 min at 5000 rpm and 4 °C. Pellets were resuspended in 50 mM ammonium hydrogen carbonate containing 8 M urea and lysed with glass beads for 2.5 min in a Bead Beater (BioSpec Products, Bartlesville, OK, USA). Cell debris and beads were removed by centrifugation. The clear supernatant was subjected to protein concentration determination with the Coomassie Plus Protein Assay Kit (Pierce Biotechnology, Rockford, IL, USA). Protein samples were further reduced with 1 mM dithiothreitol (DTT) (45 min at 37 °C), alkylated with 10 mM iodoacetamide (1 h, room temperature), diluted to 1 M urea with 50 mM ammonium hydrogencarbonate and digested with *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK) trypsin (protein/trypsin 30:1) at 37 °C for 24 h. The quality of the digest was assessed by injecting an aliquot onto a reversed-phase chromatography column. The digest was acidified to pH 3.0 with formic acid, desalted and concentrated by solid-phase extraction using an Accubond C₈ disposable column (Agilent Technologies). Peptides were eluted in 75% acetonitrile, 0.1% formic acid. The eluate was

lyophilized to dryness using a Vacuum Concentrator (Bachofner, Reutlingen, Germany) and frozen until analysis was performed.

3. Results and discussion

3.1. The general concept of the developed 2D nano-HPLC–MS–MS

As an initial step in the separation procedure, the sample was loaded onto a strong cation-exchange column. The flow-through with peptides that do not bind to the SCX column were trapped on top of a small C₁₈ enrichment column and washed free from salt (Fig. 2). In the second step this column was switched into the solvent path of the nano pump and flushed backward onto the analytical column. An increasing concentration of organic solvent eluted the concentrated sample and further separation was achieved onto the analytical reversed-phase column. The column effluent sprayed into the nano electro-spray ion trap-based MS–MS system. After the first part of the analysis was completed, the enrichment column was switched back into the solvent path of the SCX column. Peptides that were still retained on the SCX column were eluted step-by-step: every injection with a plug of increasing salt concentration onto the SCX column was followed by a switch of the second valve and subsequent reversed-phase chromatography with final MS–MS analysis. This procedure was repeated until all peptide fragments were eluted and analyzed.

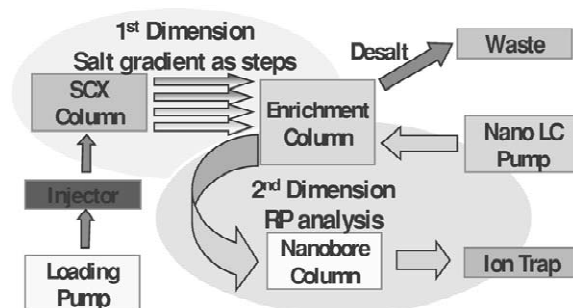


Fig. 2. Concept of the developed 2D-nano-LC methodology.

3.2. 2D-LC analysis of BSA

In order to test the method in general, 100 fmol of a tryptic digest from bovine serum albumin (BSA) was subjected to 2D-LC–MS analysis. The BSA peptides were enriched on the SCX column and fractionated by salt plugs with increasing salt concentrations (KCl) followed by reversed-phase chromatography with an acetonitrile gradient.

Results of a database search from MS–MS ion trap analysis are summarized in Table 1. MASCOT scores [25], sequence coverage, and peptide sequence obtained from MS fragmentation data and database search (MASCOT) indicate the identity of the detected molecules. Among the detected BSA peptides only one peptide occurs in more than two SCX–RPLC fractions, while most of the other identified peptides are detected only in a single fraction. This indicates a good chromatographic separation in the first dimension and an optimized usage of MS scan time for the different peptides.

By combining the results from the different salt/RP fractions, a total score of 198, a sequence

coverage of 9% and six different peptides for the 100 fmol concentration were obtained.

3.3. Analysis of a sample of 10 digested bovine proteins

In order to demonstrate the capability of the method for proteomics applications, the 2D-nano-LC–MS system was used for analysis of a medium complex tryptic digest. Ten tryptically digested bovine proteins, representing more than 250 peptides in total were combined and subjected to analysis. The final concentration of the single proteins in the mixture was 100 fmol/ μ l for each protein. Elution from the SCX column was performed in 10 mM KCl increments, followed by an increasing gradient of acetonitrile on the reversed-phase dimension. Fig. 3 shows base peak chromatograms (BPC) for selected salt fractions with subsequent RPLC. Most of the tryptic peptides were detected in the fractions between 10 and 100 mM salt. Therefore, a narrower step width up to 100 mM seems reasonable while a wider step width for higher salt concentrations is

Table 1
Results for the 2D-LC–MS method evaluation with a trypsin digest from BSA

KCl (mM)	Score	SC* (%)	Peptides	Peptide sequence	Peptide mass	Peptide position
30	66	1	1	EYEATLEECCA	1503.573	375–386
45	46	2	2	RPEEER YLYEIAR	814.393 926.986	461–466 161–167
60	81	2	2	YLYEIAR LVNELTEFAK	926.986 1162.623	161–167 66–75
75	90	3	4	LVNELTEFAK GVFR QEPERNECFLAHK	1162.623 477.269 1673.796	66–75 20–23 118–130
90	63	1	1	LVNELTEFAK	1162.623	66–75
105	48	3	4	ETYGDMADCCEK	1479.484	106–117
30–105	198	9	6	GVFR LVNELTEFAK ETYGDMADCCEK QEPERNECFLAHK YLYEIAR EYEATLEECCA	477.269 1162.623 1479.484 1673.796 926.986 1503.573	20–23 66–75 106–117 118–130 161–167 375–386

Shown are MASCOT scores, sequence coverage (SC) and number of identified peptides.

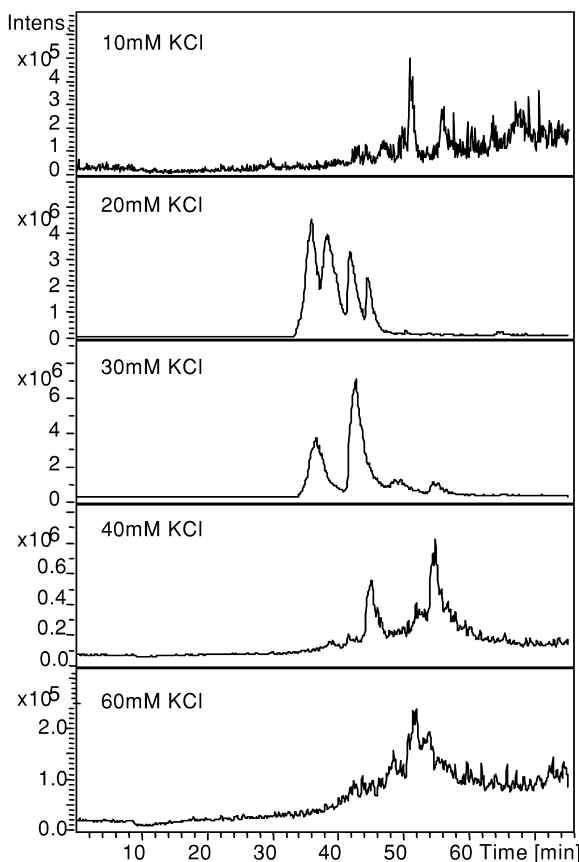


Fig. 3. Base peak MS chromatograms (m/z 400–1800) for different salt step elutions from a SCX column with subsequent RP chromatography obtained from a model mixture of 10 tryptically digested bovine proteins.

sufficient to elute the few remaining strongly retarded peptides (see experimental details).

Fig. 4 represents a selection of base peak chromatograms for peptides from serotransferin compared to a BPC of the entire mixture of 10 proteins obtained from a one-dimensional (1D) experiment, which shows the whole complexity of the mixture. The base peak chromatograms are representing one specific peptide each, which occurs in only one single fraction originating from serotransferin. The full power of the method is demonstrated by the resolution of the serotransferin peptides in the 30 and 50 mM fraction originating from the lower abundant part of the 1D-BPC and especially by the resolution of the two peptides in the 50 and 60 mM

fractions occurring at the same retention time in the 1D-BPC.

The results of the MASCOT database search for the 10 protein mixture are summarized in Table 2. By combining the data files of different fractions into a single data search file it becomes obvious that the total score for protein identification increases. All proteins in the mixture were unambiguously identified after a database search indicated by the obtained database scores and sufficient sequence coverage and numbers of peptides for the single proteins.

3.4. Analysis of the *E. coli* proteome

The completion of the DNA sequence from *E. coli* has shown that this organism encodes about 4200 genes and hence more than 4000 possible protein products. The protein composition, however, is highly variable depending on many extra- and intracellular stimuli.

Using the fully automated online 2D-LC–MS–MS system we were able to identify more than 500 proteins in a single 2D-LC experiment. Each run was done repeatedly and on average 100–300 proteins were found in each analysis. Fig. 5 shows the base peak chromatogram of a representative 2D-LC experiment. Tryptic peptides from *E. coli* eluted with salt concentrations between 20 and 200 mM. The proteins with the top 20 scores from all experiments are presented in Table 3.

The peptide maps obtained from these experiment contained cytosolic as well as hydrophobic membrane spanning proteins. In addition we were able to identify very small and very large proteins, acidic and basic proteins. Particularly for these protein classes, 2D gel electrophoresis is limited and it has been independently shown that 2D-LC–MS is superior in identifying such proteins in a global proteome analysis [23,24].

4. Conclusion

The results obtained from the experiments described above clearly demonstrate the capability of 2D-LC–MS for the separation and identification of complex peptide samples. The described method using a strong cation-exchange and a reversed-phase

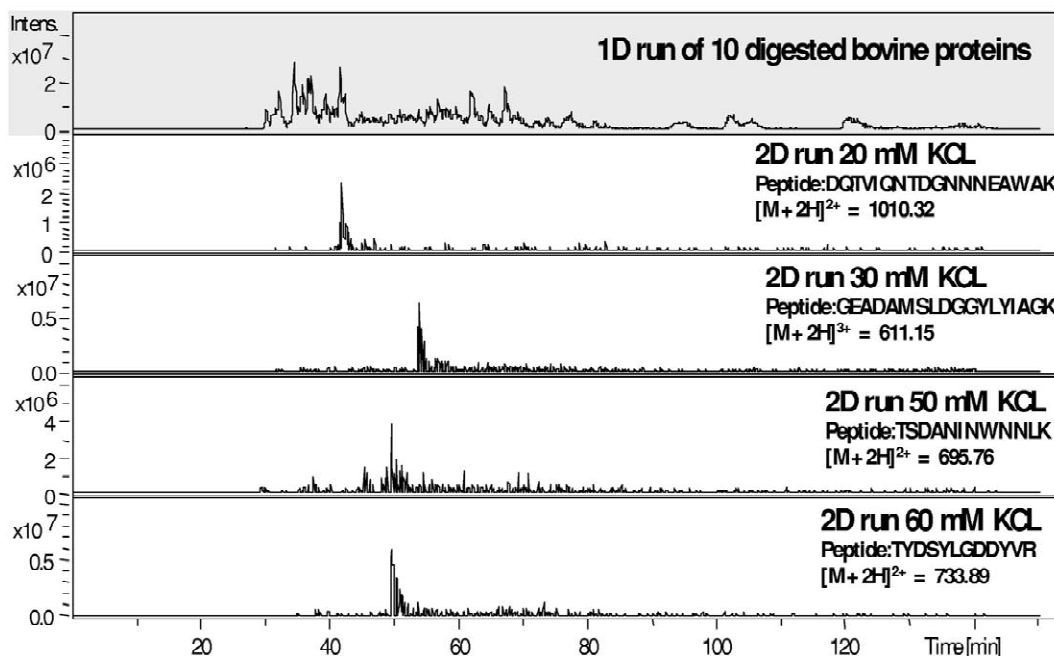


Fig. 4. Representation of a selection of base peak chromatograms for peptides from serotransferrin compared to a BPC of the entire mixture of 10 proteins obtained from a one-dimensional experiment (upper panel).

column in sequence is ready for use in complex proteomics applications for global and functional proteome investigations. While for samples of low complexity 1D chromatography might be sufficient and preferable due to easier handling, higher sensitivity, and shorter analysis time, 2D-HPLC is the

method of choice for “real life samples”, low abundance and hydrophobic membrane proteins.

In general, 2D-LC allows much more flexibility for the combination of different separation techniques than gel based separation systems and it can be easily integrated in workflows where quantitative

Table 2
2D-LC–MS analysis of a mixture of 10 trypsin digested bovine proteins

	KCl (mM)								10–80	SC (%)	Peptide no.
	10	20	30	40	50	60	70	80			
Hemoglobin			50						40	6	1
β -Lactoglobulin		55	44						77	19	5
Chymotrypsinogen			55						102	16	3
Carbonic anhydrase				61					51	17	10
Carboxypeptidase A			52	70					02	7	4
Glutamic DH		171	163	73	62	75			160	8	4
Bovine serum albumin	88	107		212	68		123	131	170	11	6
Serotransferrin		104	215		190	240	217	169	580	19	14
Lactoperoxidase			100						96	10	5
Catalase		158	215						258	13	6

Shown are MASCOT scores for individual fractions 10–80 mM KCl, combined scores for fractions 10–80 mM KCl, total sequence coverage (SC) and total number of identified peptides.

Table 3
Top score proteins identified in the *E. coli* proteome sample

Protein name	MASCOT score	No. of peptides	SC (%)	pI	M_r ($\times 1000$)
Elongation factor TU	262	15	50	5.1	43
RNA polymerase beta chain	239	57	47	6.7	155
DNA polymerase III alpha subunit	227	41	47	5.2	130
Preprotein translocase SECA subunit	224	53	49	5.4	102
Transposase TN21	223	41	48	9.3	111
Isoleucyl-tRNA synthase	221	51	48	5.7	105
DNA helicase I	219	53	42	5.8	192
50S ribosomal protein L2	216	24	66	10.9	30
ATP dependent helicase LHR	214	50	38	6.3	170
Glutamate synthase (large chain)	202	45	41	6.3	168
RNA polymerase beta chain	202	51	36	5.2	151
RNA polymerase associated protein	201	40	46	5.0	110
30S ribosomal protein	200	14	50	10.3	26
Transcription repair coupling factor	199	52	39	5.8	130
Outer membrane usher protein FASD	199	39	45	7.0	92
DNA primase TRAC	198	56	40	5.5	159
Bifunctional puta protein	197	50	32	5.7	144
ATP dependent helicase HSPA	197	61	41	7.9	149
ATP dependent protease LA	197	43	45	6.0	88
Cell division protein MUKB	194	67	42	5.1	170

Shown are total MASCOT scores, total sequence coverage (SC), total number of identified peptides, isoelectric point (pI), and relative molecular mass (M_r) of the respective proteins.

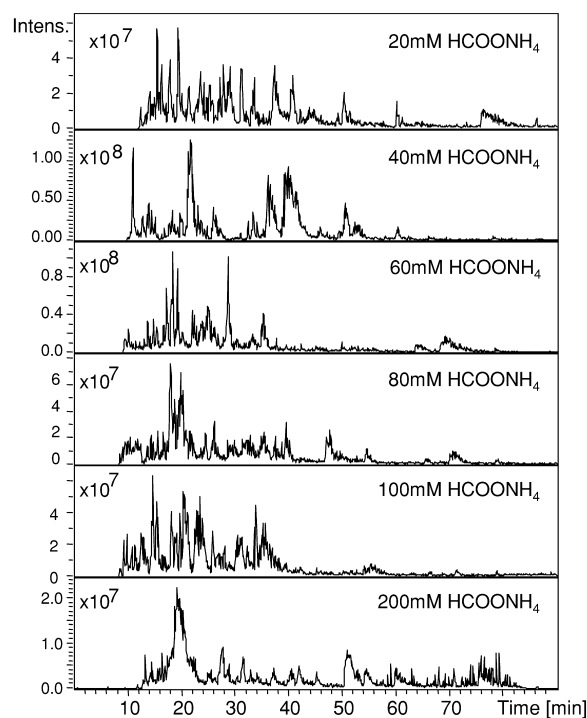


Fig. 5. Selected base peak chromatograms of the 2D-LC experiment from the *E. coli* cell lysate.

expression analysis or functional analysis of post-translational modifications requires chemical modification of protein or peptide molecules. This strongly suggest that the importance of various multi-dimensional LC-MS approaches will increase in the future field of proteomics.

References

- [1] S.R. Pennington, M.J. Dunn, *Proteomics: From Protein Sequence to Function*, BIOS-Scientific Publ, Oxford, 2001.
- [2] M.R. Wilkens, K.L. Williams, R.D. Appel, D.F. Hochstrasser, *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin, 1997.
- [3] E. Phizicky, P.I.H. Bastiaens, H. Zhu, M. Snyder, S. Fields, *Nature* 422 (2003) 208.
- [4] A. Pandey, M. Mann, *Nature* 405 (2000) 837.
- [5] M. Tyers, M. Mann, *Nature* 422 (2003) 193.
- [6] T. Rabilloud, *Proteomics* 2 (2002) 3.
- [7] P.H. O'Farrell, *J. Biol. Chem.* 250 (1975) 4007.
- [8] J. Klose, *Humangenetik* 26 (1975) 231.
- [9] A. Goerg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber et al., *Electrophoresis* 21 (2000) 1037.
- [10] M.F. Lopez, *Electrophoresis* 21 (2000) 1082.
- [11] W.F. Patton, *J. Chromatogr. B* 771 (2002) 3.

- [12] S. Hanash, *Nature* 422 (2003) 226.
- [13] H. Wang, S. Hanash, *J. Chromatogr. B* 787 (2003) 11.
- [14] J.C. Giddings, *J. High Resolut. Chromatogr.* 62 (1990) 978.
- [15] J.P. Larmann, A.V. Lemmo, A.W.J. Moore, J.W. Jorgenson, *Electrophoresis* 14 (1993) 434.
- [16] H.J. Issaq, K.C. Chan, G.M. Janini, G.M. Muschik, *Electrophoresis* 20 (1999) 1533.
- [17] G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, R.J. Anderegg, *Anal. Chem.* 69 (1997) 1518.
- [18] D. Wolters, M.P. Washburn, J.R. Yates III, *Anal. Chem.* 73 (2001) 281.
- [19] M.T. Davis, J. Beierle, E.T. Bures, M.D. McGinley, J. Mort, J.H. Robinson et al., *J. Chromatogr. B* 752 (2001) 281.
- [20] L. Giggs, C. Sioma, F.E. Regnier, *J. Chromatogr. A* 924 (2001) 359.
- [21] S. Wang, X. Zhang, F.E. Regnier, *J. Chromatogr. A* 949 (2002) 153.
- [22] M. Adamczyk, J.C. Gebler, J. Wu, *Rapid Commun. Mass Spectrom.* 15 (2001) 1481.
- [23] M.P. Washburn, D. Wolters, J.R. Yates III, *Nat. Biotechnol.* 19 (2001) 242.
- [24] L. Florens, M.P. Washburn, J.D. Raine, R.M. Anthony, M. Grainger, J.D. Haynes et al., *Nature* 419 (2002) 520.
- [25] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cotrell, *Electrophoresis* 20 (1999) 3551.