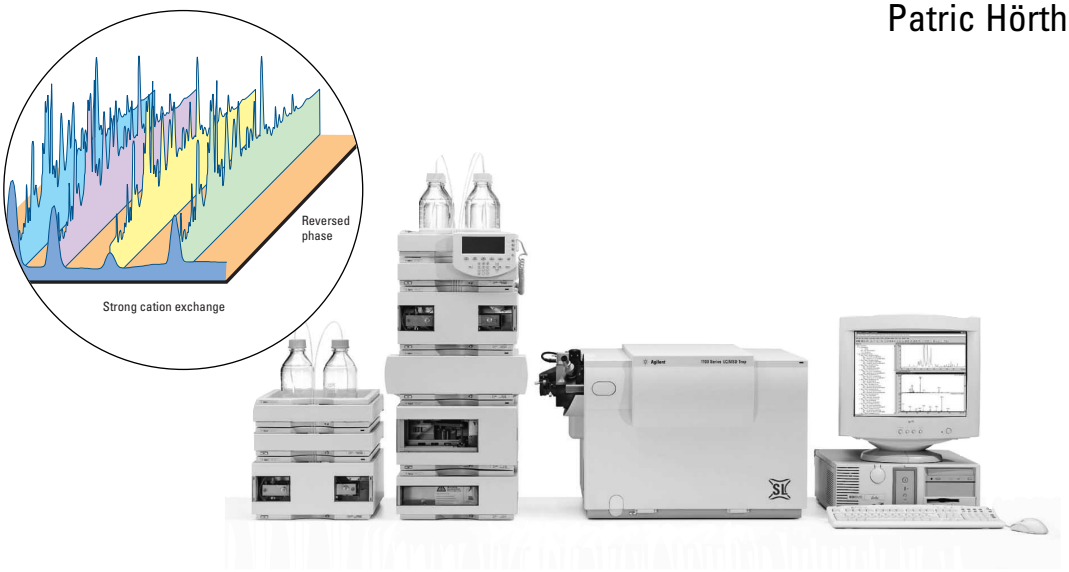
Two-dimensional nano-LC/MS for complex proteomics application

Application

Martin Vollmer
Edgar Nägele
Patric Hörth



Abstract

This Application Note demonstrates the Agilent Nanoflow Proteomics Solution system[1)] as a powerful tool for analysis of proteomics samples. Using the approach of cation exchange, sample enrichment, reversed phase chromatography and nanospray ion trap mass spectrometry with subsequent database search all of ten proteins were identified from a tryptic digestion model mixture and hundreds of proteins from a complex *E. coli* cell lysate.

Agilent Technologies

Introduction

The proteome, which is defined as the totality of all proteins derived from the genome of a cell or an organism is now the focus of many international research groups and pharmaceutical companies. Since the deciphered human genome and expression levels of the mRNA do not sufficently describe the current state of a living cell, scientists discovered the importance of analyzing the actual protein composition of a biological system in order to fully understand it. A new scientific discipline called proteomics has emerged with the goal of rapidly identifying complex protein patterns from a certain cell, tissue or organism[2,3]. The methodology needed to achieve this goal should also be capable of detecting quantitative changes and qualitative alterations of the protein profile in order to identify target proteins and their modified variants. Such comprehensive proteomic characterization will give new insight into cellular responses for disease pathogenesis such as carcinogenesis, development, aging, drug action and environmental damage.

It is obvious that these challenging tasks are not achieved with only one analytical technology, but with a combination of separation and detection techniques. The current method of choice for separating complex proteomic samples is still 2D gel electrophoresis (2D PAGE)[4] due to its high resolving power for biomolecules. In the laborious 2D-procedure, the protein mixture is separated by isoelectric focusing in the first dimension and by SDS-PAGE in the second dimension. Identification of proteins after in-gel digestion is usually performed by MALDI-TOF or ESI-ion trap.

2D PAGE, however, has some major drawbacks. It is time-consuming, difficult to reproduce and automation is hard to achieve. In addition, low abundance and membrane proteins which account for a high percentage of cellular proteins and which are usually the most promising targets for drug action, are marginally resolvable with 2D GE. Therefore, several research groups[5,6] have successfully applied 2D LC/MS in order to overcome this major obstacle. Washburn et al.[5] was able to identify 1484 proteins from yeast by combining strong cation exchange chromatography (SCX) with reversed-phase LC separation and electrospray ion trap MS/MS. This and other impressive examples, as well as the fact that 2D LC is much more flexible in combining different separation techniques and modification chemistry, strongly suggest that 2D LC will gain much more importance in future proteomic solutions.

In this Application Note, a model system of 10 bovine proteins digested with trypsin was used and showed that the Agilent Nanoflow Proteomics Solution is capable of analyzing complex proteomic samples by applying two-dimensional HPLC in combination with nanoelectrospray ion trap MS/MS analysis. The full capability is shown by the analysis of an *E. coli* cell extract.

Equipment

For this application the Agilent Nanoflow Proteomics Solution was used, which includes:

- Agilent 1100 Series nanoflow pump with micro vacuum degasser
- Agilent 1100 Series thermostatted micro well-plate autosampler
- Agilent 1100 Series thermostatted column compartment with 2-position/6-port micro valve or instead the Agilent 2-position/6-port micro switching valve box
- Agilent 1100 Series LC/MSD Trap SL, equipped with a nanoflow electrospray ion source (Bruker Daltonics, Bremen, Germany)
- Agilent ChemStation A09.02 and ion trap software 4.1

Second pump:

- Agilent 1100 Series quaternary pump with micro vaccum degasser

Software for database search:

- Mascot software (Matrix Science Inc.)

The Agilent Nanoflow Proteomics Solution is shown in figure 1. The Agilent 1100 Series micro well-plate autosampler is connected directly to the quaternary pump. From the pump, the flow path goes through the 2-position/6-port micro valve of the micro well-plate sampler and further through sample loop and needle (figure 2A). The SCX column is positioned after the needle seat and the column outlet flow is directed to the 6-port valve of the autosampler.

This valve is connected by a capillary with the second 2-position/ 6-port micro valve in the column compartement. Sample eluting from the SCX column is concentrated on a C8 or C18 enrichment column which is mounted between two ports of this valve and flow is further directed to waste. The nanoflow pump is also connected to the valve of the column compartement and the flow from this pump passes through the analytical column to the nanospray source of the ion trap.

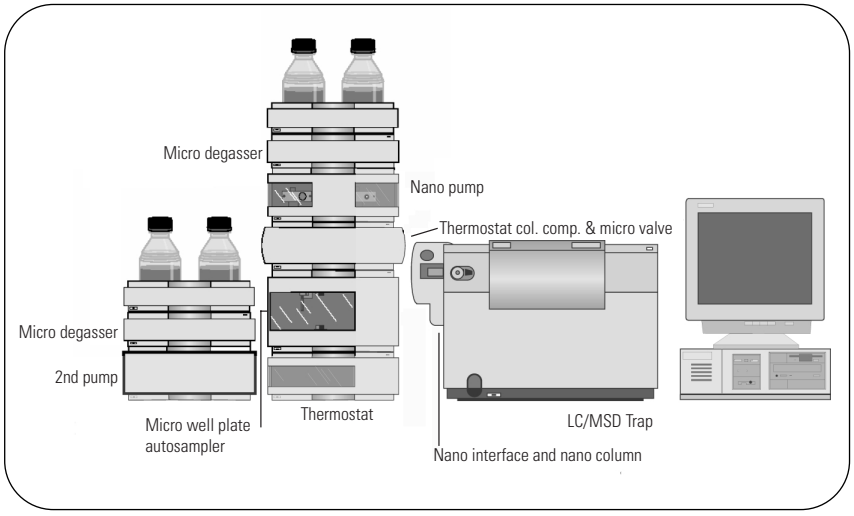


Figure 1
The Agilent Nanoflow Proteomic Solution

Results and discussions

The principle of the 2D LC

In the first step of the separation, the sample is loaded onto a strong cation exchange column (PolySulfoethyl A, PolyLC Inc.). The flowthrough with peptides that do not bind to the SCX column is trapped on top of a C18 enrichment column (ZORBAX 300SB-C18) and washed free from salt (figure 2A). In the second step, this enrichment column is switched into the solvent path of the nanopump and backflushed onto the nano column (figure 2B). An increasing concentration of organic solvent elutes the concentrated sample and further separation is achieved onto the analytical reversed-phase column (for chromatographic details see appendix, page 8). The analytical column effluent sprays into the nanoelectrospray source of the ion trap which provides low fmol sensitivity for peptides. After the first analysis is completed, the enrichment column is switched again into the solvent path of the cation

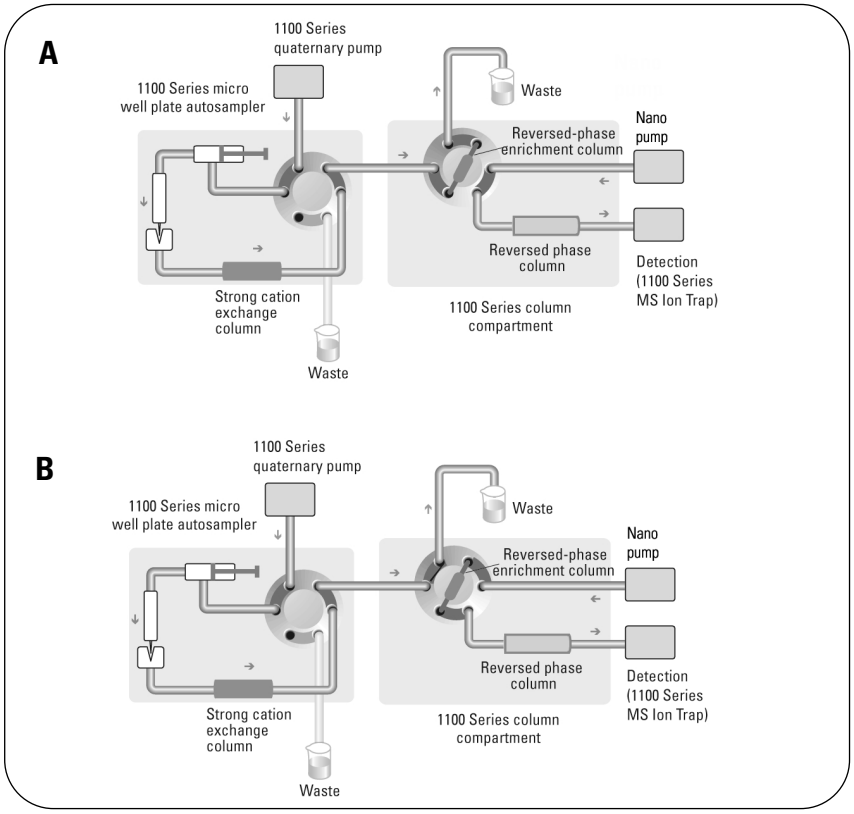


Figure 2
Flow path of the Agilent Nanoflow Proteomics Solution. A: Sample loading, elution from SCX column and enrichment on enrichment column: B: Valve switch in column compartement, elution from enrichment column, separation on RP column and MS-analysis.

exchange column. Retained peptides are eluted step by step from the SCX column by sequential injections of an increasing concentration of salt solution, followed by valve switching and reversed phase chromatography, respectively. This procedure is repeated until all peptide fragments are eluted and analyzed. To identify peptides originating from complex enzymatic protein digests it is necessary to perform MS/MS analysis of separated peptides in order to obtain sequence information from fragmentation in addition to molecular weight information. With intelligent data-dependent automatic scan functions including Auto-MS^n and ActiveExclusion high quality MS and MS^n spectra can be acquired. A fully automated protein database search[7] (Mascot, Matrix Science Inc.) completes the identification of tryptic peptide fragments and the overall protein identification based on MS/MS.

2D LC analysis of BSA

To illustrate the workflow for 2D LC, a tryptic digest from bovine serum albumin (BSA) was subjected to the nano 2D LC analysis. The base peak chromatograms obtained for the different injected salt fractions in combination with RP chromatography in contrast to the RP chromatogram alone clearly revealed a higher peak capacity for the 2D chromatography in addition to lower complexity for the single chromatogram (not shown). Results for MS/MS ion trap analysis with database search are summarized in table 1.

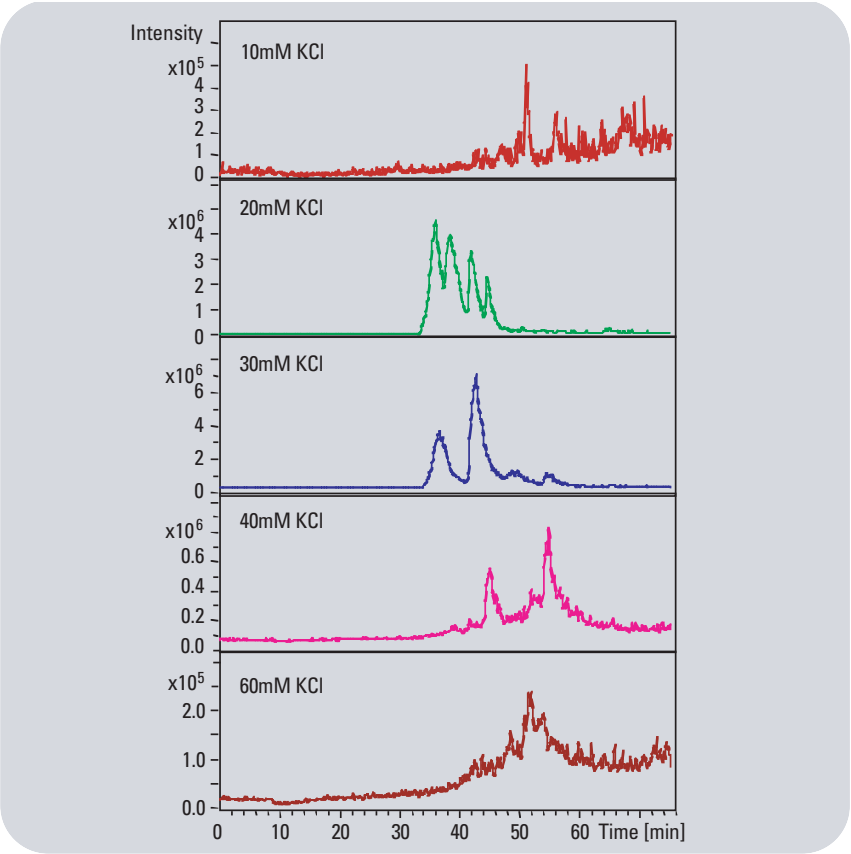


Figure 3
Base peak chromatograms (BPCs) of a ten protein digest obtained from SCX fractions after trapping and reversed phase chromatography.

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

Table 1
Trytic digest of 100 fmol of BSA was used for analysis. SC [%] = sequence coverage.

MOWSE scores, sequence coverage, and peptide sequence obtained from MS fragmentation data and database search (Mascot) indicate the identity of the detected molecules. The investigation was performed with two different concentrations of BSA (100 and 400 fmol). By combining the results from the different fractions, a sequence coverage of 9 % and 6 different peptides for the 100-fmol concentration was obtained. The analysis of a 400-fmol BSA sample resulted in a maximum total score of 508 with a sequence coverage of 20 % and the detection of 13 different tryptic peptides originating from BSA, respectively (data not shown).

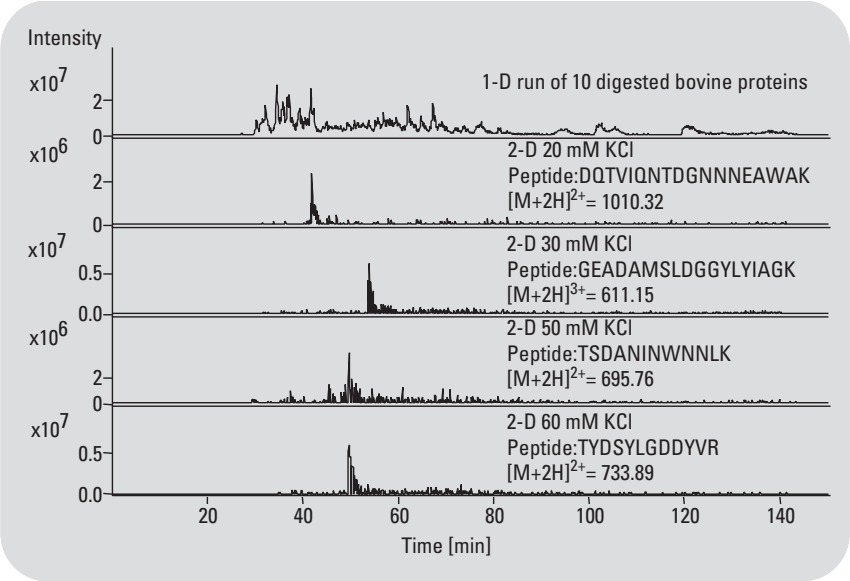


Figure 4
Selected mass spectra for serotransferin obtained from a 10-protein tryptic digest after 2D (SCX and RP) and 1D (RP only) chromatography. The spectra show the occurrence of different peptides originating from a single protein in different fractions. In general, single peptides were detected only in one to two different fractions, indicating good separation.

Analysis of a sample of 10 digested bovine proteins

In order to demonstrate the capability for proteomic applications, the Agilent Nanoflow Proteomics Solution was used for analysis of a model 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/µl for each protein. Elution from the SCX column was performed in injected 10 mM KCl increments, followed by an gradient of acetonitrile on the RP column (for chromatographic details see appendix, page 8). Figure 3 shows the base peak chromatograms of RP chromatography for selected KCl fractions.

Figure 4 represents a selection of mass spectra for serotransferrin. The peak representing a specific peptide originating from serotransferin respectively only occurs in one single 2D fraction. The results of the Mascot database search are summarized in table 2. By combining the data files of different fractions into one single data search file, it becomes obvious that the total score for protein identification increases. All proteins in the mixture were therefore unambigiously identified after database search as indicated by the Mascot scores and sufficient sequence coverage for the proteins. No false positive hits were observed among the 8 proteins with the highest scores.

Elucidation of the *E. coli* proteome

The completion of the DNA-Sequence from *E. coli* has revealed that this model organism encodes for about 4200 genes and hence more than 4000 possible protein products. The protein composition however, is highly variable depending on many extra- and intra-cellular stimuli.
Using the fully automated online nano 2D LC-MS/MS system more than 500 proteins were identified in a single nano 2D LC experiment. This experiment was performed repeatedly and in average 100 – 300 proteins were found in each analysis. Figure 5 shows the base peak chromatogram of a representative 2D LC experiment. All tryptic peptides elute with salt concentrations between 20 mM

and 200 mM ammonium formate. The proteins with the top 20 scores from all performed experiments are presented in table 3. The peptide maps revealed that cytosolic as well as hydrophobic membrane spanning proteins could be detected by coupling liquid chromatography to nano electrospray ion trap MS analysis. In addition we were able to identify very small and very large proteins, acidic and basic proteins.

mM KCl	10	20	30	40	50	60	70	80	10-80	SC[%]	Pept. no.
Hemoglobin			50						**40**	**6**	**1**
ß-Lactoglobulin		55	44						**77**	**19**	**5**
Chymotrypsinogen			55						**102**	**16**	**3**
Carbonic anhydrase				61					**51**	**17**	**10**
Carboxypeptidase A			52	70					**102**	**7**	**4**
Glutamate dehydrogenase		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**

Table 2
Tryptic digest of 10 proteins analyzed by nano 2D LC MS/MS. 100 fmol of each digested protein was used for analysis. Columns in bold show the results by combination of data analysis from all selected fractions. Numbers indicate the scores obtained by Mascot database search for the different SCX fractions. SC [%] = sequence coverage. Pept. no. = numbers of peptide identified in total.

Conclusion

The results obtained from the experiments describe clearly the capability of the 2D Agilent Nanoflow Proteomics Solution for separation and identification of complex peptide samples. The described method using a strong cation exchange and a reversed phase column in series is ready to be used in two-dimensional HPLC applications for global and functional proteomic investigations. While one-dimensional chromatography may be sufficient and preferable for low complexity samples or more focused proteomic questions due to higher sensitivity and shorter analysis time, 2D LC is the method of choice for "real life samples", low abundance and hydrophobic membrane proteins. This method has a high resolution power and is truly orthogonal, highly automateable and flexible in terms of combining different separation techniques, and therefore represents a good alternative for 2D gel electrophoresis.

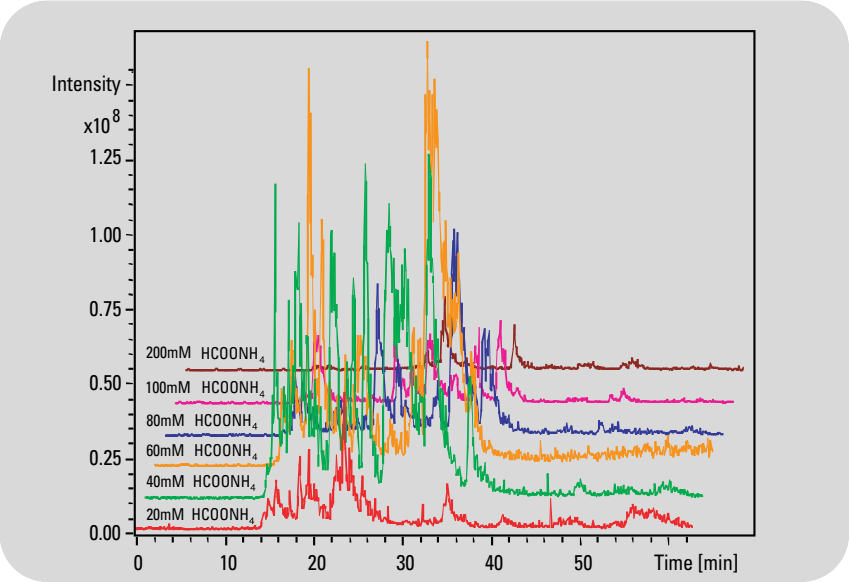


Figure 5
Base peak chromatograms of the nano 2D LC experiment from the *E. Coli* cell lysate

Protein name	MASCOT Score	Number of different peptides identified	Sequence coverage (%)	Isoelectric point (pI)	MW (kDa)
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	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 assoc. protein	201	41	46	5.0	110
30S Ribosomal Protein	200	14	50	10.3	26
Transcription Repair Coupling F.	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 HRPA	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

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

Appendix

Chromatographic details:

Columns:	
SCX column:	PolyLC Inc., PolySulfoethyl Aspartamide, 0.32 x 50 mm, 5 µm particle size.
Enrichment column:	Zorbax 300SB C-18, 0.3 x 5 mm, 5 µm particles.
Reversed phase column:	Zorbax 300SB C-18, 75 µm x 50 mm, 3.5 µm particle size.
Gradient 1:	**Separation of BSA and of the tryptic digest from 10 proteins**
Quaternary pump:	isocratic (water + 3 % acetonitrile + 0.1 % formic acid), flow: 30 µm/min
Nano pump:	A = water + 0.1% formic acid; B = AcN + 0.1 % formic acid. flow: 450 nl/min, 6 min 3 % B, 120 min 60 % B, 125 min 80 % B, 130 min 80 % B, 131 min 3 % B, 140 min 3 % B,
Stop time:	140 min
Post time:	10 min
Enrichment column switch:	5 min
Gradient 2:	**Analysis of the *E. coli* cell lysate**
Quaternary pump:	flow-gradient (water + 3 % AcN + 0.1 % formic acid), 0 min 0.1 ml, 0.5 min 0.05 ml, 0.51 min 0.01 ml, 8 min 0.01 ml, 8.01 min 0.005 ml, 9.09 min 0.005 ml, 10.00 min 0.000 ml, 85 min 0.000 ml, 85.01 min 0.005 ml,
Nano pump:	A = water + 0.1 % formic acid; B = AcN + 0.1 % formic acid.
Flow:	450 nl/min, 0 min 5 % B, 10 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
Enrichment column switch:	10 min in nano flow path, 85 min out of nano flow path.

MS details:

Source: positive nanoelectrospray		Octopole 2:	2.1 V	Target 40.000
Drying gas flow:	3 l/min	Trap drive:	80 V	Max. accu. time: 150 ms
Drying gas temp.:	225 °C	Oct. RF:	150V	Averages: 4
Skim1:	40 V	Lens 1:	-5V	Ion mode: positive
Cap. exit:	135 V	Lens 2:	-60V	MS/MS mode: auto MS2
Octopole 1:	12 V	ICC:	on	and active exclusion

References

1.
"Agilent Nanoflow Proteomics Solution", *Agilent Technologies Brochure*, **2002**, publication number 5988-3621EN.

2.
"Proteomics: from protein sequence to function" *In: Pennington SR, Dunn MJ eds. Oxford: BIOS-Scientific Publishers Limited*, **2001**.

3.
"Proteome Research: New Frontiers in Functional Genomics". *In: Wilkens MR, Williams KL, Appel RD, Hochstrasser DF, eds. Berlin Springer*, **1997**.

4.
O'Farrel PH. "High resolution two-dimensional electrophoresis of proteins". *J. Bio. Chem. 250: 4007–21*, **1975**.

5.
Washburn MP, Wolters D, Yates JR. "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." *Nat. Biotechnol. 19: 242–247*, **2001**.

6.
Davis MT, Beierle J, Bures ET, McGinley et al. "Automated LC-LC—MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses". *J. Chromatogr. B 752: 281–291*, **2001**.

7.
Perkins DN, Pappin DJC, Creasy DM, and Cotrell JS "Probability-based protein identification by searching sequence databases using mass spectrometry data." *Electrophoresis, 20: 3551–3567*, **1999**.

Martin Vollmer is R&D Biochemist, Edgar Nägele is Application Chemist and Patric Hörth is R&D Biochemist, all at Agilent Technologies GmbH, Hewlett Packard Strasse 8, 76337 Waldbronn, Germany.

www.agilent.com/1100

Printed April 1, 2003
Publication Number 5988-9287EN

Agilent Technologies