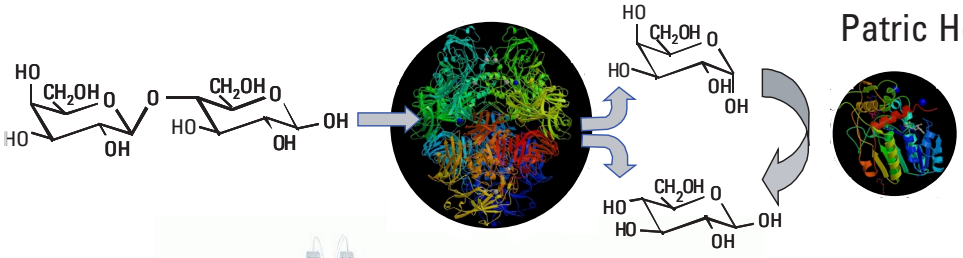
Differential proteome analysis: Two-dimensional nano LC/MS of *E. coli* proteome grown on different carbon sources

Application

Edgar Nägele
Martin Vollmer
Patric Hörth



Abstract

When provided to bacteria as single sources of carbon and energy, different sugars require the induction of different metabolic enzymes, transporters, and uptake systems to support cell growth and survival. Using the Agilent Nanoflow Proteomics Solution comprehensive peptide maps were recorded for *E. coli* grown with either lactose or glucose in minimal medium. This Application Note presents data which shows the ability of the Agilent Nanoflow Proteomics Solution to detect thousands of peptides for the identification of several hundred different proteins, which are expressed significantly under both conditions. Several enzymes and binding proteins related to the lactose metabolism were only identified in the sample grown with this carbon source.

Agilent Technologies

Introduction

Proteomics can be defined as qualitative and quantitative comparison of proteomes under different conditions to unravel biological processes. However, single proteomes can consist of more than 10^6 individual proteins with concentration ranges differing by a factor of 10^5, depending on the nature of the sample[1]. In addition, proteomes are highly dynamic systems where synthesis, degradation and modification are changing constantly in response to internal and external stimuli. To develop suitable techniques for separation, detection and analysis of complete proteomes and their constituents, there are currently two different approaches that promise to fulfill this enormous task: 2D gel electrophoresis (2DGE) and 2D nano LC/MS. 2DGE has been the method of choice for many years[2,3] and, due to its high resolving power, this technique is now applied by many laboratories for research in protein science. Despite significant improvements, 2DGE still suffers from a lack of reproducibility and from time consuming manual interventions. In addition, hydrophobic membrane proteins, very large and very small proteins, and proteins that exhibit extreme pI values are difficult to resolve with 2DGE. In contrast, 2D nano LC/MS is more flexible because it allows the combination of different separation techniques. Samples can be tagged and modified before, in between or after a single separation step and sequences of multiple runs are automated easily. In several recent reports it has been demonstrated that the combination of various orthogonal HPLC separation techniques coupled to tandem mass spectrometric analysis provides a competitive technique to 2DGE[4,5,6,7] in proteomics applications.

Carbohydrate uptake and utilization in *E. coli* and the corresponding induction of enzymes has been under investigation for many years since the lac operon was first described by Jacob and Monod. Different binding proteins, transporters and metabolic enzymes for peripheral pathways are induced during growth on glucose and lactose, respectively, whereas the central metabolism by glycolysis and the tricarbonic acid cycle is common for both sugars. Therefore, the investigation of *E. coli* grown with these carbohydrates as single source of carbon and energy provide a good model system to show subtle differences in the proteomes by changing just one parameter. Furthermore, this study demonstrates the power of 2D nano LC in combination with nano electrospray ion trap MS/MS for comparative proteome studies.

In this Application Note data is presented which shows the ability of the Agilent Nanoflow Proteomics Solution to detect several thousand peptides for the identification of several hundred different proteins significantly expressed under both mentioned conditions, respectively. Several enzymes and binding proteins related to the lactose metabolism were only identified in the sample grown with this carbon source.

Materials and methods

Equipment

For analysis of the complete *E. coli* proteomes the Agilent Nanoflow Proteomics Solution[8] was used. The system included the following components:

- Agilent 1100 Series nanoflow pump with micro vacuum degasser
- Agilent 1100 Series quaternary 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 Agilent 2-position/6-port micro switching valve box
- Agilent 1100 Series LC/MSD Trap SL with orthogonal nanospray source
- Agilent ChemStation A09.03 and Ion Trap software 4.1
- Agilent Spectrum Mill MS Proteomics Workbench[9]

Columns:

1. Reversed phase (RP): ZORBAX 300 SB C18, 75 µm x 50 mm, 3.5 µm particles.
2. Enrichment: ZORBAX 300SB C18, 0.3 mm x 5 mm, 5 µm particles
3. Strong cation exchange (SCX): PolyLC Inc., Polysulfoethyl A, 0.32 mm x 50 mm, 5 µm particles.

Cultivation and sample preparation:

E. coli was grown in parallel at 37 °C in M9 minimal medium until mid-log phase supplied with either 5 g of glucose or lactose. Cells were harvested in a Stratos Biofuge (Heraeus Intruments,Hanau,

Germany) for 15 min with 5000 rpm at 4 °C. Pellets were resuspended in 5 mM ammonium bicarbonate and lysed with glass beads for 2.min (Beat beater, BioSpec Products, Inc., 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, Inc., Rockford, Il, USA). Protein samples were reduced with 1mM DTT (45 min at 37 °C), alkylated with 10 mM iodoacetamide (1h, RT) and digested with TPCK trypsin (protein/trypsin 30/1) at 37 °C for 24 h. Quality of the digest was assessed by injecting an aliquot onto a reversed phase 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, Waldbronn, Germany). Peptides were eluted in 75 % acetonitrile, 0.1 % formic acid. The eluate was lyophilized to dryness using a SpeedVac Concentrater (Bachofer, Reutlingen, Germany) and frozen until analysis was performed.

Chromatography

The principle of the 2D nano HPLC is illustrated in figure 1. For the first dimension, 20 µl of redissolved digest in mobile phase A (about 50 µg total peptide) was injected onto the SCX column. The column was directly connected to the needle seat of the micro well-plate autosampler (figure 2A). Mobile phase (3 % ACN, 0.1 % Formic acid) was pumped from the second pump (quaternary pump) through autosampler and SCX column. The

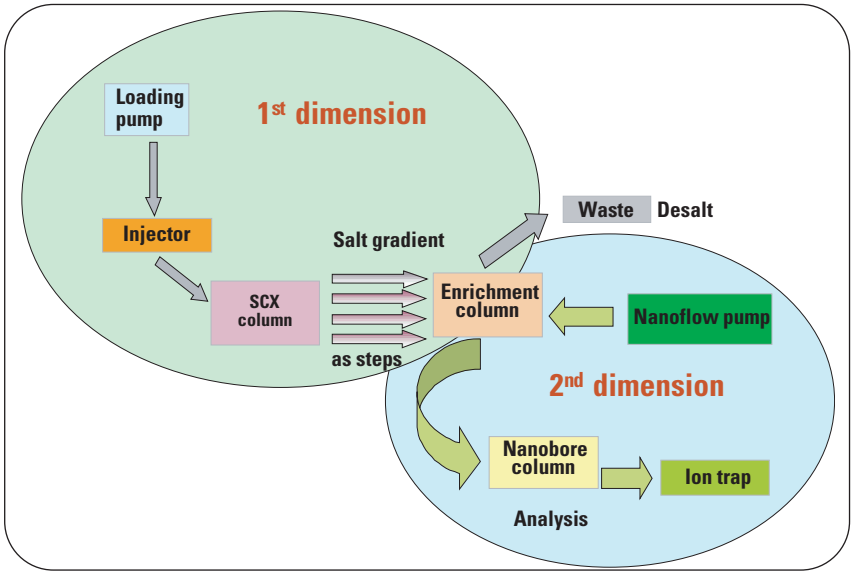


Figure 1
Principle of two-dimensional LC with the combination of SCX and RP chromatography

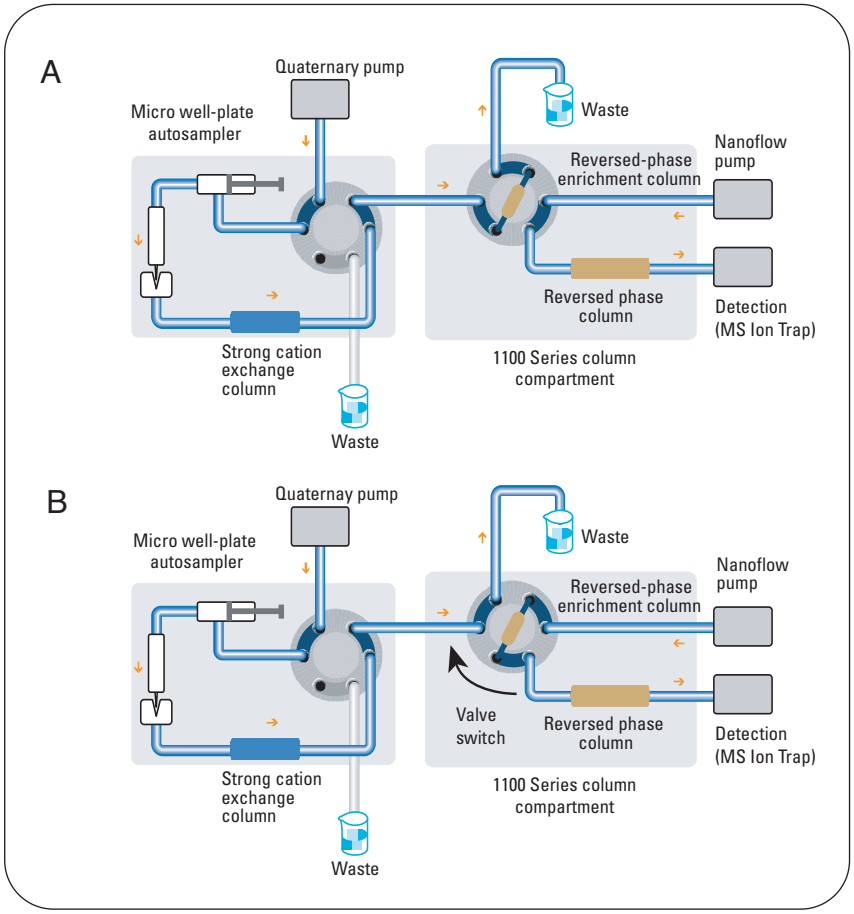


Figure 2
Flow path of the Agilent Nanoflow Proteomics Solution.
A: Sample loading. Elution from SCX and trapping on enrichment column
B: Valve switch in column compartment, elution from enrichment column separation on RP and MS analysis.

column outlet was connected to the 2-position/6-port micro valve in the autosampler. The non-binding peptides were directed to the second 2-position/6-port micro valve in the column compartment and enriched on a short C18 enrichment column, which was mounted in between two ports of this valve. During peptide elution from the SCX column, the enrichment column was inline with the SCX column and the eluent of the enrichment column was directed to waste, which allowed washing of salts and other non-binding contaminants that would have been disadvantageous for later MS analysis. Elution from SCX is obtained stepwise by injecting increasing concentrations of ammonium formate from the autosampler (20 µl portions). Salt concentrations of 20, 40, 60, 80, 100, 150, 200, 300, 500, 1000 mM were used to elute peptides from the first dimension with the quaternary pump. 3 % acetonitrile with 0.1 % formic acid was used as mobile phase. The flow program for the quaternary pump was performed as followed: 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. After each salt injection the enrichment column was switched in the nano flow path where mobile phase was directed from the nanoflow pump through the valve to the RP nanocolumn (figure 2B). The outlet of the column was directly connected to the sprayer needle in the ionsource of the MS instrument. By switching the valve in the column compartment after 10 min, the enrichment column is transferred into the nano flow path resulting in reversed flow through this column. The increasing concentration of organic solvent eluted the sample, which was enriched on this column and further separation was achieved on the analytical reversed-phase nanocolumn (solvent: A = H_2O + 0.1 % formic acid; B = AcN + 0.1 % formic acid). The nanoflow pump delivered the following gradient: 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. The flow rate of the nanoflow pump was 450 nL/min. In the following procedure the trapping column was switched back into the solvent path of the SCX column for the next elution step.

Mass spectrometry and data analysis

The outlet of the RP nanocolumn was connected online to the sprayer needle in the orthogonal nanospray ion source of the Agilent ion trap and eluting peptides were analyzed directly by data-dependent MS/MS to obtain mass and sequence information for database analysis. High quality MS and MS^n spectra were taken with automatic scan functions including Auto-MS^n and Active Exclusion. The data files from the different fractions were extracted automatically with the Spectrum Mill software and extracted peptide MS/MS spectra were subjected to database search using the NCBI database in the Spectrum Mill software. Individual database searches were performed for the glucose and lactose sample and only search results indicated as significant with a significant score were taken into consideration. In addition, mass spectra, ion series of fragmentation patterns and sequence coverage from identified proteins involved in lactose and glucose metabolism were inspected manually and investigated for plausibility. The details of the two-dimensional nano LC/MS method are summarized in table 1.

Results and discussion

By performing 2D nano LC in combination with ESI iontrap MS/MS, analysis of *E. coli* cellular extracts originating either from a lactose or glucose grown culture 305 and 450 proteins were identified respectively, from single experiments at a high confidence level.

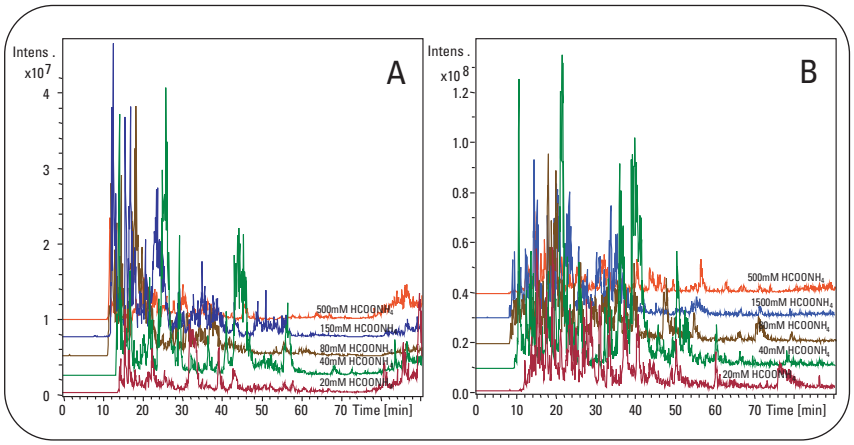


Figure 3
Base peak chromatograms obtained by MS analysis after reversed phase chromatography of individual SCX elution steps for both *E. coli* cultures. A) *E. Coli* culture grown on glucose. B) *E. Coli* culture grown on lactose.

High chromatographic resolution was achieved for both culture conditions as shown in the base peak chromatograms obtained by MS analysis after reversed phase chromatography of individual SCX elution steps (figure 3). Even at the salt step injection's high concentrated end useful peptides were found. Among positively identified proteins, cytosolic and membrane proteins, metabolic enzymes and structural proteins were equally represented (figure 4). Regarding the size of the identified proteins, there is a slight bias towards larger proteins due to the higher number of tryptic peptides generated, and the corresponding likelihood to detect one or several peptide fragments of single proteins by MS analysis. Glucose and lactose are mainly metabolized by glycolysis and the tricarbonic acid cycle. These enzymes are essential for both sugar metabolic pathways and should therefore be present and detectable in both proteomes. Using the Agilent Nanoflow Proteomics Solution, where the protein mixture is sequentially eluted online from a SCX column to an enrichment column and further separated by reversed phased chromatography on a nanobore C18 reversed phase column, the presence of all of these enzymes (with the exception of phospho glycerate kinase) could be demonstrated for both culture conditions (table 2). The metabolic pathway of glycolysis with all the involved enzymes is outlined in figure 5. All enzymes involved and identified by database search achieved confident scores. For most of the proteins several peptides were detected which gave a sequence coverage up to 40 %. The fact that all of these glycolytic enzymes were identified for both conditions demonstrate the quality of the analysis and imply that differences observed in protein pattern account for differences in protein expression between the two conditions. In contrast, enzymes only or predominantly necessary for lactose uptake and conversion to glucose-6-phosphate, as a common metabolite for both pathways, were exclusively

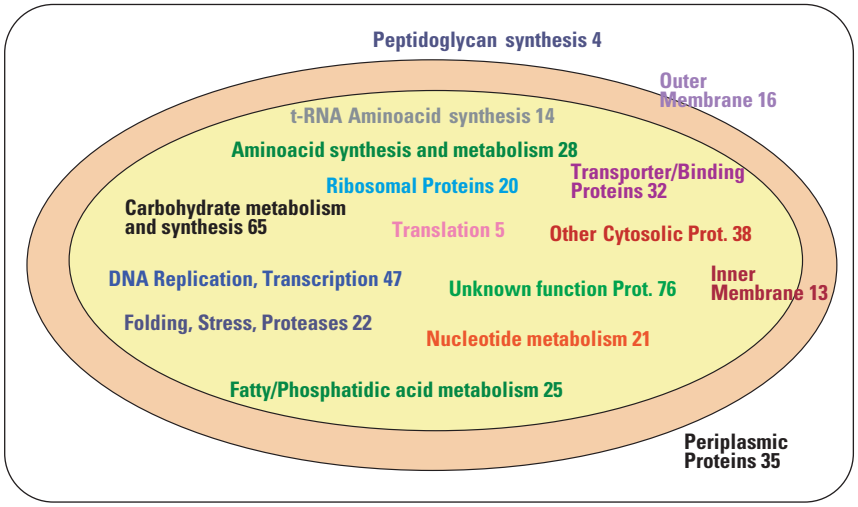


Figure 4
Number of positively identified proteins. Cytosolic- and membrane proteins, metabolic enzymes and structural proteins are equally represente

Columns:	
SCX:	PolyLC Inc., PolySulfoethyl Aspartamide, 0.32 x 50 mm, 5 µm
Enrichment column:	Zorbax 300SB C18, 0.3 x 5 mm, 5 µm
Analytical column:	Zorbax 300SB C18, 75 µm x 50 mm, 3.5 µm
Nanoflow pump:	
Solvent:	A = 0.1 % formic acid in water; B = 0.1 % formic acid in acetonitrile
Gradient:	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
Flow:	450 nL/min
Micro valve:	
Enrichment column switch:	0 min in-line with SCX; 10 min in-line with nanocolumn; 85 min in-line with SCX
Autosampler:	
Injection volume:	20 µL Sample, 20 µL salt solution
Salt steps injected on SCX ($HCOONH_4$):	20, 40, 60, 80, 100, 150, 200, 300, 500, 1000 mM
Quaternary pump:	
Solvent:	0.1 % formic acid in 3 % acetonitrile in water
Flow:	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

MS conditions:

Source: orthogonal nanospray source
Drying gas flow: 3 L/min
Drying gas temp: 225°C
Skim 1: 40 V
Capillary exit: 135 V
Trap drive: 80V
ICC: on, target 40,000; Max.accu. time: 150 ms
Averages: 4

Automatic MS/MS:
Preferred charge state: +2
Number or precursors: 3
Isolation width: 1.15 V
Fragmentation amplitude: 4
SmartFrag: On, 3-200 %

Table 1
Method for 2D nano-LC/MS

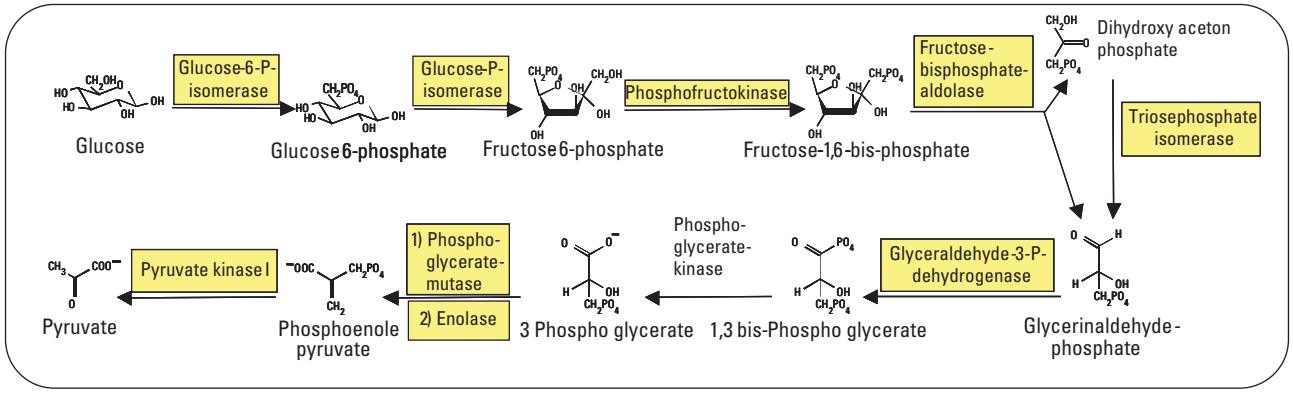


Figure 5
Metabolic pathway of the glycolysis from glucose to pyruvate. All proteins identified in both bacteria cultures are displayed in a box.

detected in proteomes originating from the lactose culture (lower part in table 2). As shown on the pathway for lactose metabolism in *E. coli* (figure 6), key enzymes are beta-Galactosidase, UDP-Galactose-epimerase and Glucose phosphate uryditansferase. The first enzyme catalyzes the hydrolysis of the disaccharide to glucose and galactose while the latter are involved in the conversion of Galactose-1-phosphate into Glucose-1-phosphate. The title illustration shows the crystal structure model of the lactose metabolism enzymes beta Galactosidase and UDP-Galactose 4- Epimerase aquired by X-ray diffraction (www.rcsb.org/pdb). These enzymes were detected in the lactose proteome with a significant level. Selected MS/MS spectra for peptides originating from from proteins according to the glycolysis and to the lactose metabolism show comprehensive and consecutive y- and b-series fragmentation patterns (figure 7).

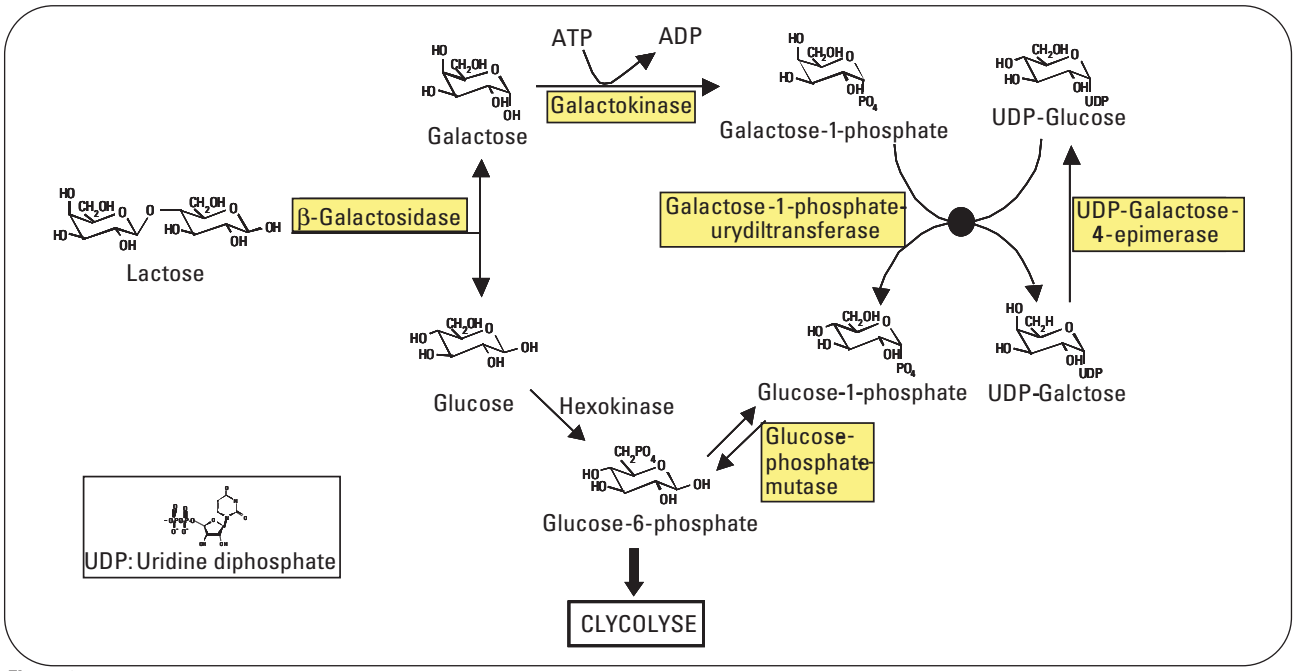


Figure 6
Metabolic pathway for lactose degradation. All metabolic enzymes were identified only in the lactose culture.

Conclusion

The Agilent 1100 Nanoflow Proteomics Solution provides an excellent, highly automated tool for online two-dimensional nano LC/MS, combining sequential SCX step gradient separation with reversed phase chromatography by an intermediate enrichment step. Wolters et.al.[5] estimated, according to a theoretical calculation of Giddings et.al.[10], that the total peak capacity of their 2D HPLC MudPit system including mass spectrometry is 23000 peptides (assuming 15 SCX fractions and a RP gradient length of 90 min). This peak capacity may be sufficient for many proteomics applications with medium to high complexity such as bacterial proteomes as in this study and for organelles or for subcelluar fractions, which need resolution for 500 to 1000 proteins. If, however, the focus is on complete protein expression patterns from eukaryotic cells or body fluids with proteins encompassing a wide dynamic range, additional prefractionation[11] and orthogonal separation techniques on protein or peptide level will be necessary. These can include the removal of high abundance proteins or tagging chemistries to reduce initial complexity without losing relevant information. Alternatively, the resolution of the first dimension could be increased significantly by switching from stepwise online elution to linear gradient offline separation with an intermediate microfraction collection device.

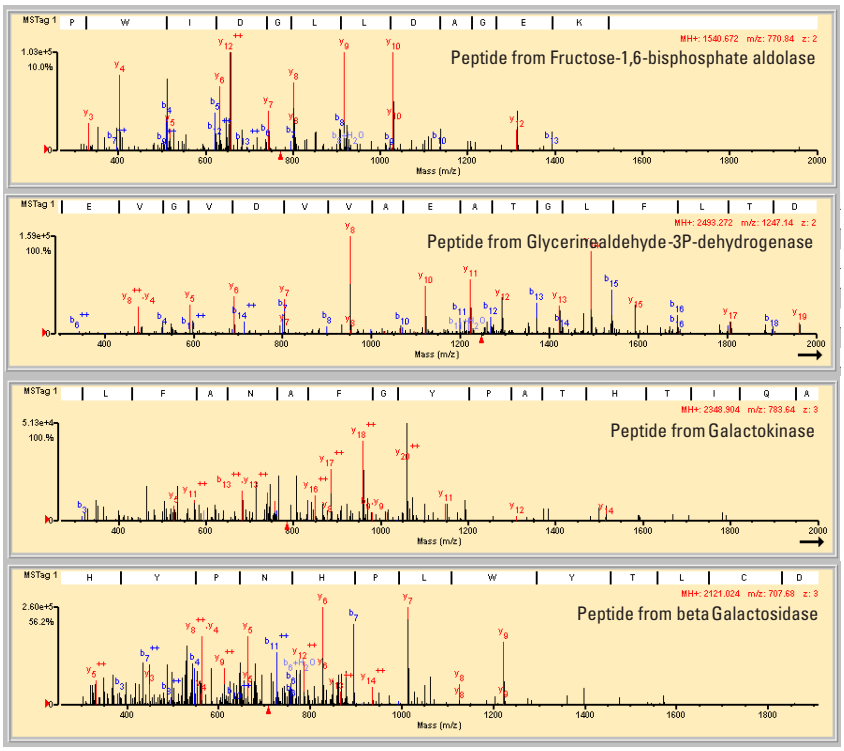


Figure 7
Fragmentation pattern for peptides from Glycolysis- and Lactose metabolism proteins.

Spectra (#)	Distinct Peptides (#)	Distinct Summed MS-Tag Score	% AA Coverage	Protein MW (kDa)	Protein pI	Protein name
Proteins from glucose and lactose culture						
1	3	11.29	4	35.7	5.68	Glucose-6-phosphate isomerase
1	1	3.86	4	35.7	5.68	Glucose-6-phosphate isomerase
2	2	20.90	4	61.5	5.85	Glucose phosphate isomerase
1	1	9.02	2	61.5	5.85	Glucose phosphate isomerase
1	1	3.20	4	32.4	5.25	Phosphofructokinase
4	8	28.68	20	39.1	5.52	Fructose-bisphosphate aldolase
4	5	35.31	15	39.1	5.52	Fructose-bisphosphate aldolase
1	1	3.56	6	25.5	5.58	Triosephosphate isomerase
1	1	4.44	6	25.5	5.58	Triosephosphate isomerase
7	15	50.92	29	35.5	6.61	Glyceraldehyde-3-phosphate dehydrogenase
8	60	74.72	40	35.5	5.73	Glyceraldehyde-3-phosphate dehydrogenase
3	41	40.08	18	28.5	5.85	Phosphoglycerate mutase
6	73	74.09	40	28.5	5.85	Phosphoglycerate mutase
4	22	40.84	12	45.6	5.32	Enolase
5	14	54.65	17	45.6	5.32	Enolase
1	2	12.83	4	50.7	5.77	Pyruvate kinase I
3	6	41.40	10	50.7	5.77	Pyruvate kinase I
Proteins from lactose only culture						
1	1	7.87	5	41.4	5.28	Galactokinase
1	1	3.48	2	85.6	5.94	beta Galactosidase
1	1	8.01	6	32.9	5.11	Galactose-1-phosphate uridylyltransferase
1	1	7.22	4	37.2	5.89	UDP-galactose 4-epimerase
1	1	7.60	5	47.5	5.71	Glucosephosphatemutase
1	1	5.21	2	47.0	9.43	D-Galactonate transporter
2	2	12.54	6	46.5	9.20	Lactose permease
1	1	11.29	4	35.7	5.68	D-Galactose/D-Glucose binding protein

Table 2
Proteins identified from the *E. coli* grown on lactose or glucose. Proteins from the glycolysis from *E. coli* grown on glucose are marked dark gray and glycolysis proteins from the lactose culture are marked in light gray. Proteins related to the lactose metabolism detected only in the lactose culture are listed in the lower part of the table.

References

1.
Pennington SR, Dunn MJ (Eds), Proteomics: from protein sequence to function, *Oxford: BIOS-Scientific Publishers Limited*, **2001**.

2.
O'Farrel PH, High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem*, *250: 4007*, **1975**.

3.
Klose J, Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, *26: 231-243*, **1975**.

4.
Davis MT, Beierle J, Bures ET et al., Automated LC-LC-MS/MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteome analyses. *J. Chromatogr. B*, *752: 281-291*, **2001**.

5.
Wolters DA, Washburn MP, Yates JR III. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.*, *73: 5683-5690*, **2001**.

6.
Florens L, Washburn MP, Raine JD et al., A proteomic view of the Plasmodium falciparum life cycle, *Nature*, *419: 520-526*, **2002**.

7.
Washburn MP, Wolters D, Yates JR III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.*, *19: 242-247*, **2001**.

8.
"Agilent Nanoflow Proteomics Solution" *Agilent Technologies Brochure*, *publication number 5988-3621EN*, **2002**.

9.
"Agilent Spectrum Mill MS Proteomics Workbench", *Agilent Product Brochure*, *Publication Number 5988-9443EN*, **2003.**

10.
Giddings JC, Multidimensional chromatography: techniques and applications. *In Cortes HJ (ed) New York: Marcel Dekker*, *chapter 1*, **1990.**

11.
"Agilent 1100 Series Micro Fraction Collection System", *Agilent Product Brochure*, *Publication Number 5988-9614EN*, **2003.**

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

www.agilent.com/chem/proteomics

Published September 01, 2003
Publication Number 5988-9862EN

Agilent Technologies