Improved protein identification: Off-line multidimensional LC/MS as an effective tool for proteomics research

Application Note

Ralf Moritz
Martin Vollmer
Edgar Nägele

Abstract

Compared to fully automated on-line two-dimensional LC/MS of digested proteome samples, an off-line two-dimensional approach greatly increases chromatographic resolution resulting in more significant protein identification results. In this Application Note we present off-line 2D LC/MS as an alternative technique to on-line 2D LC/MS for the analysis of complex proteome samples. The first dimension involves off-line fraction collection with the Agilent 1100 Series micro fraction collection system. The second dimension uses the Agilent Nanoflow Proteomics Solution together with the Agilent Spectrum Mill MS Proteomics Workbench for LC/MS/MS analysis and subsequent protein identification.

Agilent Technologies

Introduction

Today, automated on-line 2D LC/MS/MS of digested proteins is generally applied for proteome characterization[1]. Usually on-line 2D LC/MS/MS is performed using a strong cation exchange (SCX) column in series with a reversed phase (RP) column. In the course of analysis tryptic peptides are eluted stepwise by injecting salt plugs of increasing ionic strength from the SCX column in the first dimension. In the second dimension these peptides are first trapped on a reversed phase (RP) enrichment column and finally separated on an analytical RP column[2]. It has been shown that this methodology, especially in nano scale, is capable of resolving large proteomes[3] as well as identifying a subset of proteins expressed under special conditions[4] in differential proteomics research.

In contrast to this well established on-line methodology off-line 2D LC/MS/MS significantly increases chromatographic resolution of highly complex proteome samples. This improvement is attributed to continuous gradient SCX-chromatography in the first dimension without intermitting RP-chromatography intervals. As a consequence a higher peak capacity and therefore a higher number of identified proteins is associated with the off-line 2D LC/MS/MS approach.

In this Application Note off-line 2D LC/MS is presented as an alternative technique for analysis and identification of complex proteome samples. The Agilent 1100 Series micro fraction collection system[5] was used for fraction collection in the first dimension (SCX). In the second dimension (RP) the Agilent Nanoflow Proteomics Solution[6] in combination with the Agilent Spectrum Mill MS Proteomics Workbench[7] was used for reliable protein identification. The results obtained with the on-line and the off-line approaches are compared and discussed.

Experimental

Equipment

Agilent 1100 Series micro fraction collection system: Capillary pump with micro vacuum degasser, thermostatted well-plate autosampler, thermostatted column compartment with micro 2-position/6-port micro valve, DAD with 500 nL cell and thermostatted micro fraction collector (figure 1 A).

Agilent 1100 Series Nanoflow Proteomics Solution: Nanoflow pump with micro vacuum degasser, thermostatted micro well-plate autosampler, 2-position/6-port micro switching valve box with holder, quaternary pump, LC/MSD Trap XCT with orthogonal nanospray ion source (figure 1B).

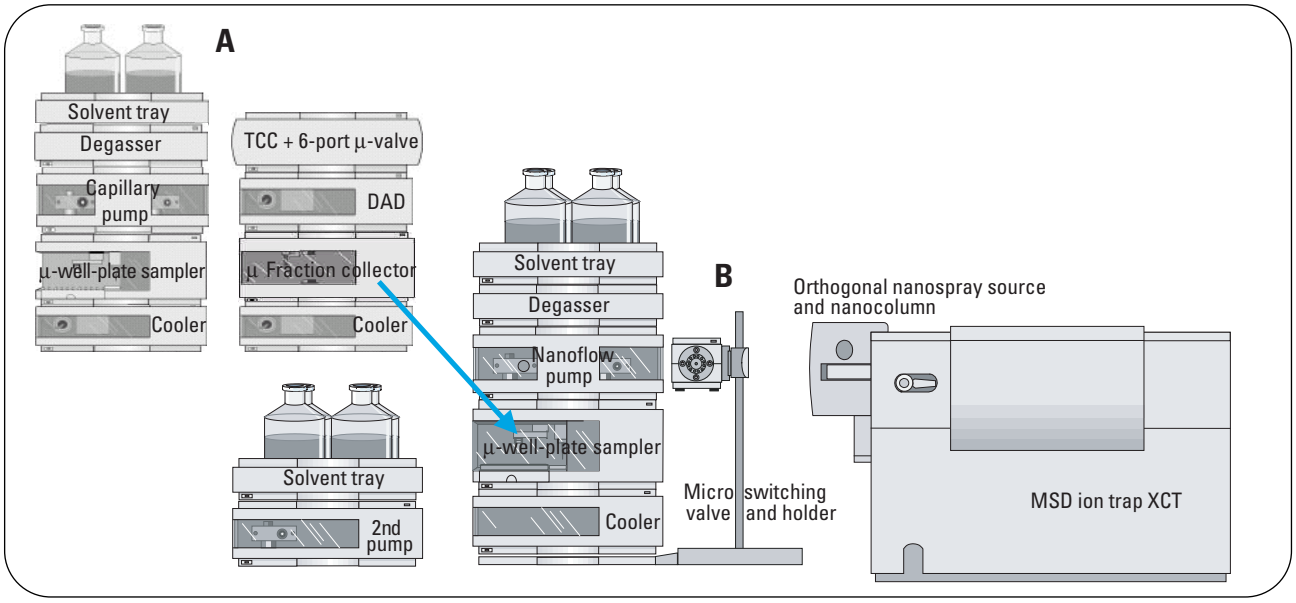


Figure 1
Off-line 2D LC/MS system. A: Micro fraction collection system. B: Agilent Nanoflow Proteomics Solution

Software:
Agilent ChemStation rev. A.10.01 and Spectrum Mill MS Proteomics Workbench software.

The typical workflow for an off-line 2D LC proteome analysis is shown in figure 2. The first dimension starts with the elution of tryptic peptides from a SCX column with a linear salt gradient and the subsequent collection of fractions in the micro liter range with the micro fraction collector (figure 2 A). The well-plate, containing all collected fractions was then transferred directly to the second dimension – the Agilent Nanoflow Proteomics Solution (figure 2 B). After reinjection the peptides from each fraction are concentrated on a short C18 enrichment column located between the ports of the micro valve. The enrichment column was then switched into the nanoflow path, the concentrated peptides were eluted, further separated by a gradient with increasing organic solvent on an analytical nanobore column and sprayed directly into the MS.

Chromatographic method

Off-line approach:	
1st dimension (SCX chromatography):	Agilent 1100 Series micro fraction collection system
Solvents:	A: 5 % AcN +0,03 % formic acid, B: 500 mM NaCl, + 5 % AcN + 0.03 % formic acid.
Gradient:	0 min 0 % B, 5 min 0 % B, 35 min 20 % B, 38 min 100 % B, 41 min 100 % B.
Flow:	5 µL/min.
Column:	Agilent BioSCX Series II, 0.30 x 35 mm, 3.5 µm particles
Autosampler:	5 mL sample injection volume
Time-based fraction collection mode:	1st fraction 0 - 5 min, 2nd to 13th fraction each 3 min, liquid contact control mode, fraction size 15 µL, cooling 4 °C .
2nd dimension (RP Chromatography):	Agilent 1100 Series Nanoflow Proteomics Solution
Nanoflow pump:	
Solvents:	A = water + 0.1% formic acid, B = AcN + 0.1% formic acid
Gradient:	0 min 5 % B, 5 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:	300 nL/min
Columns:	Zorbax 300SB C18, 75 µm x 150 mm, 3.5 µm particles Zorbax 300SB C18, 0.3 x 5 mm, 5 µm particles
Micro valve:	Enrichment column switch: 5 min
Autosampler:	15 µL sample injection volume
Quaternary pump:	
Solvent:	water + 3% AcN + 0.1 % formic acid, flow: 20 µL/min
On-line approach[2]:	
	Agilent 1100 Series Nanoflow Proteomics Solution
Nanoflow pump:	
Solvents:	A = water + 0.1 % FA, B = AcN + 0.1 % FA.
Flow:	300 nL/min.
Gradient:	0 min 5 %B, 5 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.
Columns:	Agilent BioSCX Series II, 0.30 x 35 mm, 3.5 µm particles Zorbax 300SB C18, 75 µm x 150 mm, 3.5 µm particles Zorbax 300SB C18, 0.3 x 5 mm, 5 µm particles
Micro valve:	
Enrichment column switch:	0 min in-line with SCX; 5 min in-line with nanocolumn; 85 min in-line with SCX
Autosampler:	
Injection volume:	5 µL sample injection volume, 20 µL salt solution injection volume
Salt steps injected on SCX (NaCl):	0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 300, 500 mM

MS conditions for both experiments:

Source:	positive orthogonal nanospray source	
Drying gas flow:	3 L/min	
Drying gas temp:	225 °C	
Skim 1:	40 V	
Capillary exit:	135 V	
Trap drive:	80 V	
ICC:	on	
	Target:	40,000
	Max. accu. time:	150 ms
	Averages:	4
Automatic MS/MS:	Number or precursors:	3
	Isolation width:	1.15 V
	Fragmentation amplitude:	4
	Preferred charge state:	+2
	SmartFrag:	On, 30-200%

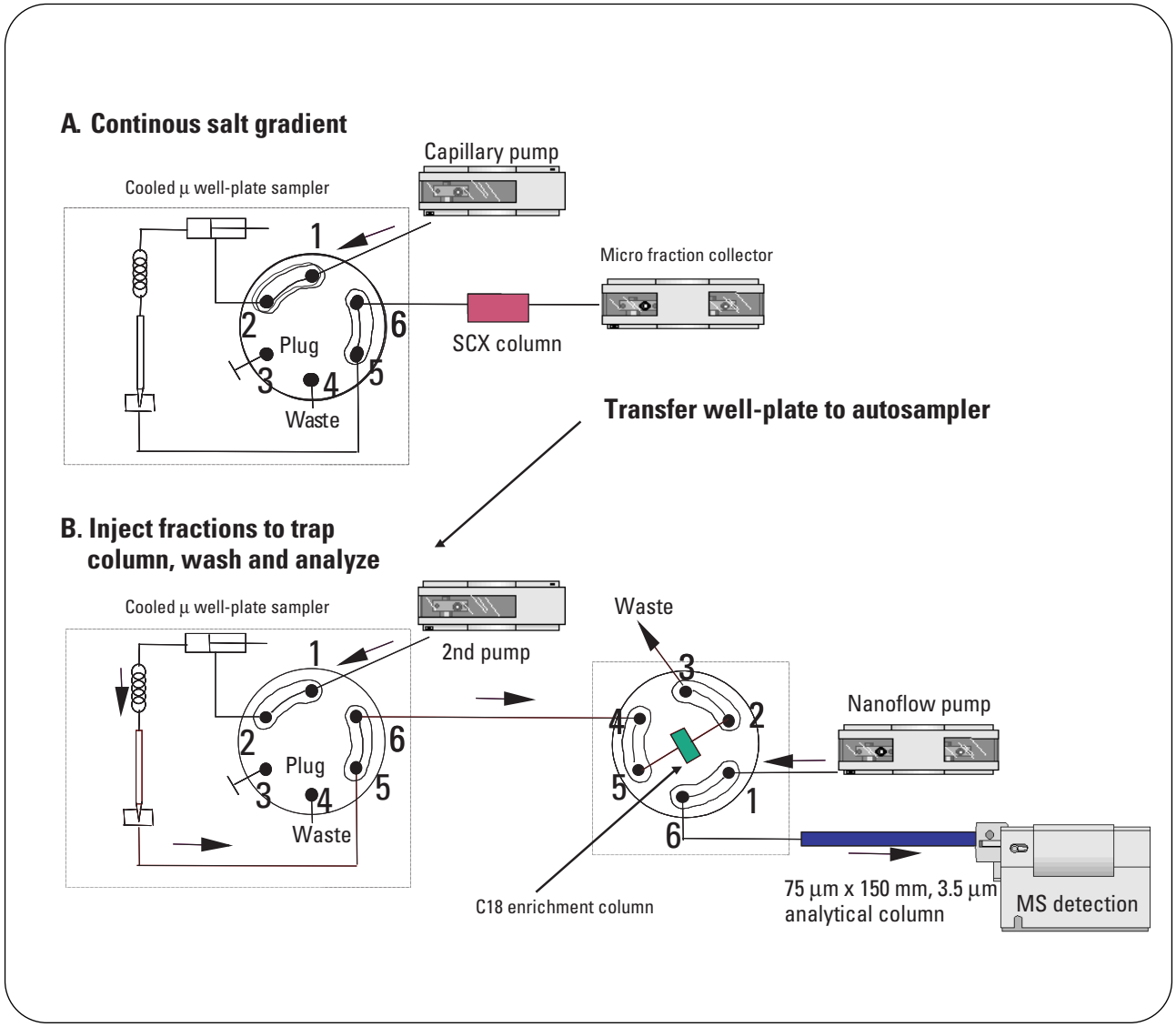


Figure 2
Workflow for proteome identification by off-line 2D LC/MS

Sample preparation

Lyophilized yeast cells *(S. cerevisiae)*, resuspended in cooled 50 mM NH_4HCO_3 containing 8 M urea were disrupted in a bead beater (0.5-mm glass beads). After centrifugation to remove cell debris proteins in the supernatant were reduced with 1 mM DTT at 37 °C for 1h, alkylated in the dark with 10 mM iodoacetamide for 30 min at RT, ultrafiltrated for buffer exchange and tryptically digested with TPCK trypsin at 37 °C for 16 h. Finally, the sample was lyophilized in a SpeedVac and dissolved in 5 % AcN, 0.03 % formic acid prior to analysis.

Results and discussion

Collection principle

In order to collect small volumes in the lower µL range reliably a unique collection principle was developed which is illustrated in figure 3. During the entire collection process, the fraction collector ensures that the droplet forming at the capillary tip is in constant contact with the surface of the deposited liquid. At the start of fraction collection, the droplet being formed is precisely deposited at the bottom of the well. As the well is filled the capillary tip moves continuously upwards, keeping constant contact with the solvent surface. When fraction collection is finished the capillary tip moves quickly upwards, ensuring that the droplet currently being formed at the capillary tip is delivered to the well. The well filling velocity during fraction collection is not only calculated from the current flow rate but also from the geometry of the actual well type. Such a unique collection principle guarantees air bubble-free, precise deposition of low volume fractions in a broad variety of well plate types and vessels from different manufacturers. In addition, cross contamination between collected fractions is effectively prevented.

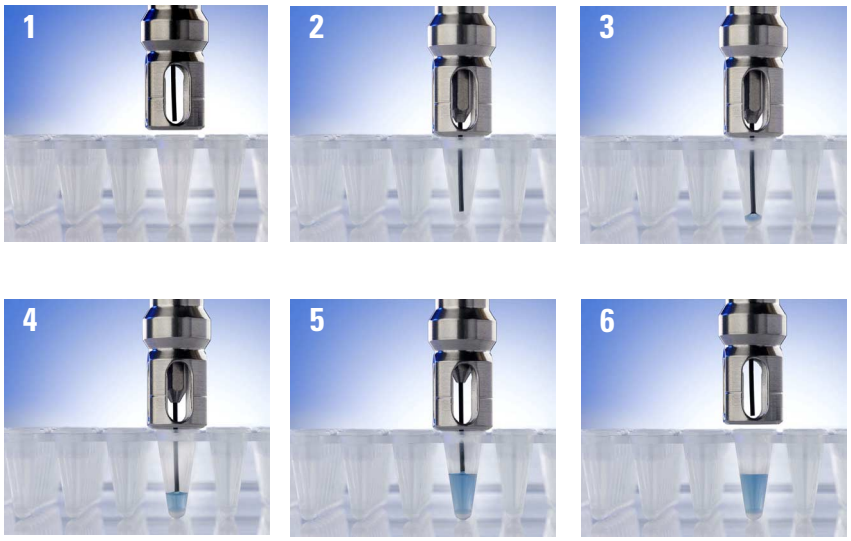


Figure 3
Collection principle for small volumes in the lower µL range. During the collection process the capillary stays in constant contact with the solvent. This approach leads to air bubble free accurately collected fractions without cross contamination.

Data evaluation

Trypsin-digested *Saccharomyces cerevisiae* proteome was subjected to off-line and on-line 2D LC-MS/MS analysis. For both approaches the same conditions were chosen for solvents, sample and columns. For the on-line version the individual salt step concentrations were

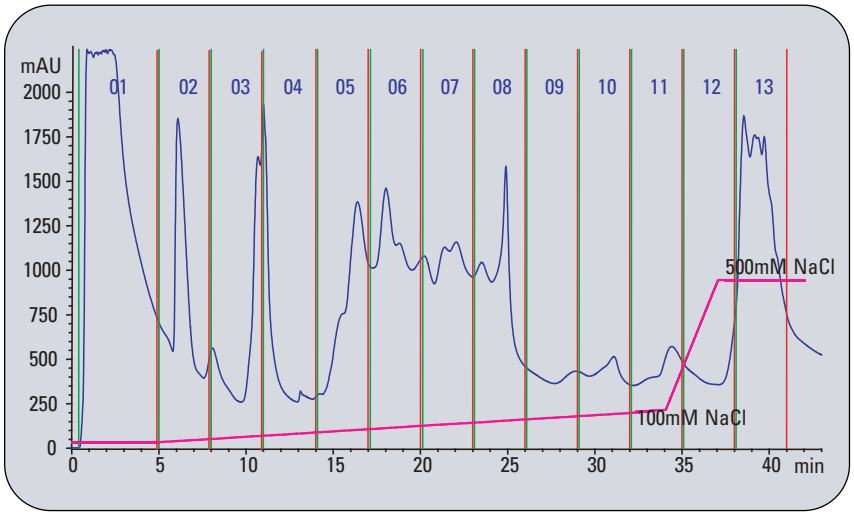


Figure 4
Time-based collected fractions from the first dimension (SCX) and the applied salt gradient.

adjusted such that they practically corresponded to the fractions collected in the off-line approach. After evaluation of the MS/MS data with the Agilent Spectrum Mill software using the identical stringent parameters for both approaches the following results were obtained.

For the nano on-line 2D LC-MS/MS methodology, which comprises injected salt steps within the 1st dimension and a nano LC RP separation in the 2nd dimension 101 proteins were identified. For the off-line methodology that was performed by applying continuous gradient elution during the 1st dimension and a nano LC RP separation in the 2nd dimension, 144 proteins were identified. These results were confirmed in three repeated analyses. Proteins with highest scores for both analyses indicate that basically the same high abundance proteins were identified in both experiments (table 1). However, the values of the scores are different and in general higher in the off-line than in the on-line approach. This clearly demonstrates that the higher chromatographic resolution obtained with a continuous linear salt gradient leads to well separated peaks, and finally to a higher number of detected peptides in the MS.

On-line 2D LC/MS analysis of the yeast proteome				
Score	**Peptides**	**Spectra**	**Coverage**	**Protein name**
150,46	10	45	36	Phosphopyruvate hydratase
123.75	8	52	25	Phosphoglycerate kinase
114.54	8	21	24	Pyruvate decarboxylase
98.57	7	76	23	Translation elongation factor eEF-1
81,53	6	16	21	Pyruvate kinase
65.13	4	16	13	Hexokinase A
62.63	5	39	16	Alcohol dehydrogenase I
55.54	4	16	23	Phosphoglycerate mutase 1
48.01	4	6	18	Triosephosphate isomerase
47.39	3	7	22	Eukaryotic translation initiation factor 5A-2
33.54	3	3	5	Heat-shock Protein HSC82
32.65	2	15	14	BMH2 Protein
31.18	2	2	13	Ribosomal protein S6
29.24	2	7	6	Phosphogluconate dehydrogenase
24.70	2	2	27	Superoxide dismutase
23.30	2	2	20	Cytochrome C oxidase
18.24	1	1	10	Malate dehydrogenase
17.65	1	3	6	40S Ribosomal protein S5
16.97	1	3	3	Citrate synthase
16.79	1	2	5	Glycerol-3-phosphate dehydrogenase
16.33	1	2	4	Glucose kinase

Off-line 2D LC/MS analysis of the yeast proteome				
Score	**Peptides**	**Spectra**	**Coverage**	**Protein name**
161.50	12	27	42	Phosphopyruvate hydratase
150.45	10	32	32	Phosphoglycerate kinase
112.92	8	14	14	Pyruvate decarboxylase
97.77	6	8	19	Glucose kinase
93.12	7	17	23	Translation elongation factor eEF-1
73.11	5	7	13	Citrate synthase
71.66	5	8	19	Pyruvate kinase
64.03	4	6	14	Hexokinase A
63.04	4	19	21	Glyceraldehyde-3-phosphate dehydrogenase
56.07	4	5	23	Eukaryotic translation initiation factor 5A-2
55.24	4	8	13	Phosphoglycerate mutase 1
54.04	4	5	20	60S Ribosomal protein L4-A
43.59	3	4	8	Phosphogluconate dehydrogenase
43.35	3	5	18	BMH1 Protein
42.49	3	9	15	Alcohol dehydrogenase
41.15	3	5	11	Glycerol-3-phosphate dehydrogenase
39.39	3	4	6	Aconitate hydratase
37.57	3	3	15	Ribosomal protein S4
36.99	3	6	11	Ketol-acid reductoisomerase
33.77	3	3	19	Guanine nucleotide-binding protein
30.14	2	3	27	Superoxide dismutase

Table 1
Proteins with highest scores for on-line and off-line analysis of the yeast cell lysate.

The chromatogram obtained for the first dimension with the marked fractionations and the applied gradient is shown in figure 4. This chromatogram clearly shows that the majority of the peptides is eluted for a salt concentration up to 100 mM NaCl from the SCX column and collected in fractions 2-9. The MS base peak chromatograms corresponding to fraction 5-8 are shown in figure 5. As a consequence of the better resolution the detection of more peptides per single protein leads to higher sequence coverage and therefore to a more significant protein identification. Furthermore, a better resolution leads to the detection of additional lower abundant proteins for the off-line approach. Due to a higher

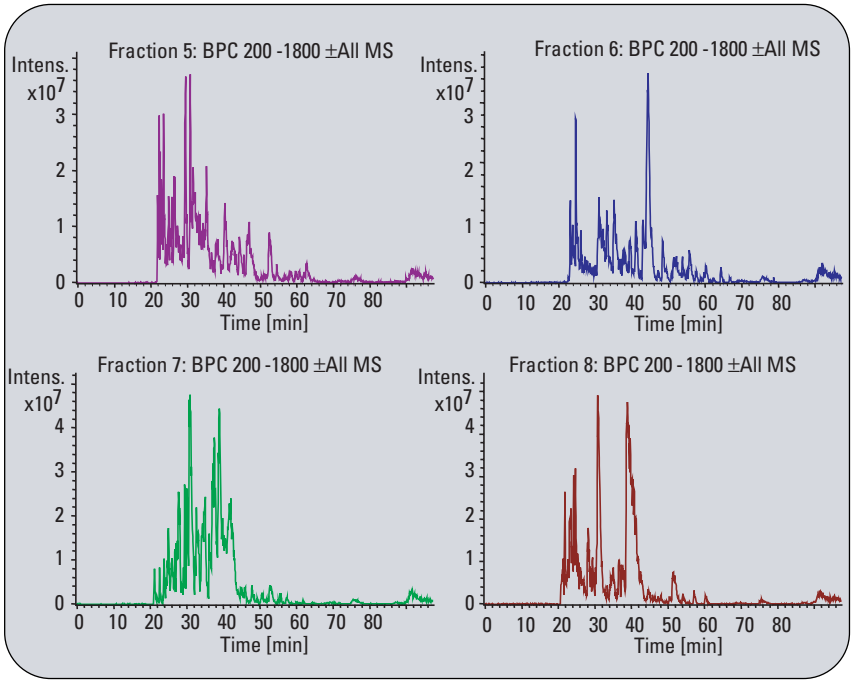


Figure 5
Base peak chromatograms from SCX fractions fractions after RP separation

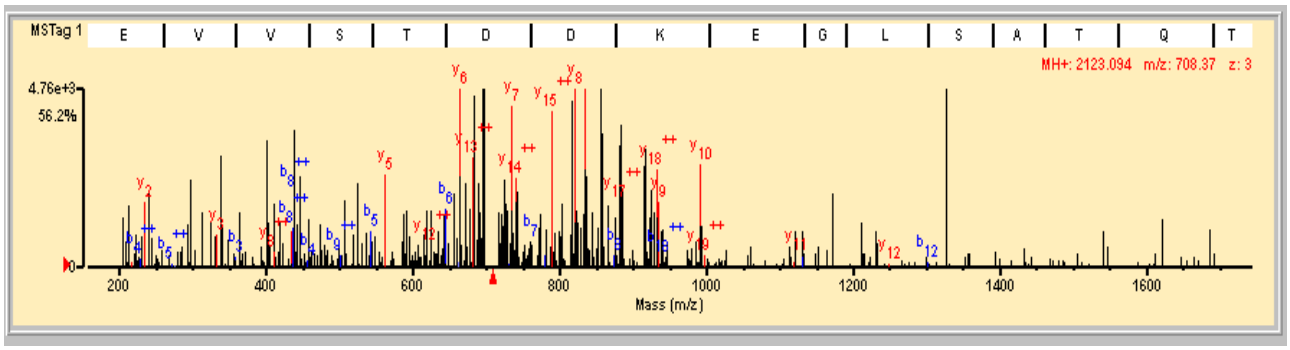


Figure 6
Peptide from Orothate phosphoribosyltransferase. Protein number 140 in the off-line experiemnt, Score 9.28, Peptides 1, Spectra 1, Sequence coverage 8 %, MW 24.6 kDa, pI 5.80.

chromatographic resolution in the off-line method peptides and corresponding proteins which fall below significance level in the on-line procedure were identified. This is demonstrated with the identification of Orothate phosphoribosyltransferase, which is protein number 140 in the off-line experiment and not found in the on-line experiment. The MS/MS fragmentation pattern of the identified peptide gives a reliable identification of this particular protein (figure 6).

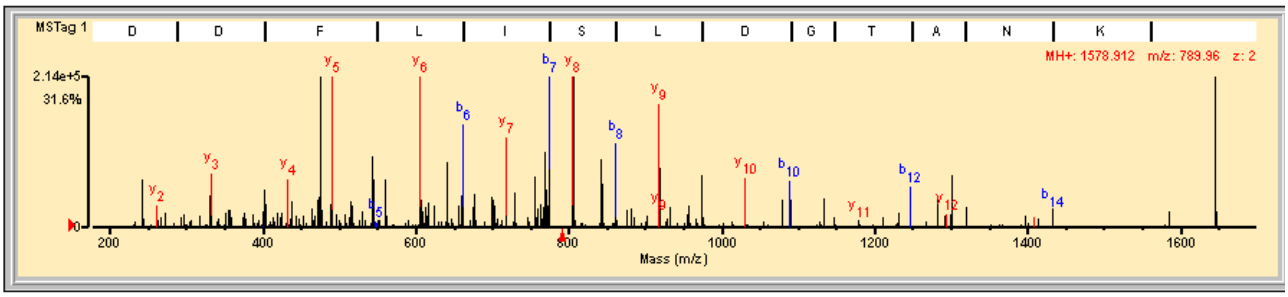
The Agilent Spectrum Mill software provides an excellent tool to compare complex proteome analyses from multidimensional separations. Results from individual salt step injections and single fractions as well as distribution of peptides in different fractions can be displayed and reviewed easily. This is shown for the distribution of the tryptic peptides over the

Sequence	m/z Measured	MH+ Matched	FRAC1-3 Score	FRAC4 Score	FRAC5 Score	FRAC6 Score	FRAC7 Score	FRAC8 Score	FRAC9 Score	FRAC10 Score	FRAC11 Score	FRAC12-13 Score
(K)AAQDSFAAGWGVMVSHR(S)	597.49	1789.84						10.60				
(K)AVDDFLISLDGTANK(S)	790.13	1578.80			14.90 10.83 13.71 9.79 14.19	14.25						
(R)GNPTVEVELTTEK(G)	708.94	1416.72			12.75							
(R)IEEELGDNAVFAGENFHHGDKL(-)	814.69	2441.14									18.84	
(K)IGLDCASSEFFK(D)	687.51	1373.64			12.43							
(K)NVNDVIAPAFVK(A)	644.07	1286.71			12.34 13.21 11.83							
(K)NVPLYKHLADLSK(S)	749.95	1497.84						9.09	12.49			
(R)SGETEDTFIADLVVGLR(T)	911.52	1821.92			10.00							
(R)SIVPSGASTGVHEALEMR(D)	921.50	1840.92					15.50 12.74 9.62	13.19 16.50 12.30				
(K)TAGIQIVADDLTVTNPK(R)	878.37	1755.95		13.90 18.35 16.79								
(K)VNQIGTLSESIK(A)	645.09	1288.71			12.20							
(K)WLTGPQLADLYHSLMK(R)	625.09	1872.97							9.23			

Table 2

Distribution of the tryptic peptides over the collected fractions from the protein Phosphopyruvate hydratase identified in the off-line experiment.

collected fractions from the protein Phosphopyruvate hydratase identified in the off-line experiment (table 2). This peptide distribution-also demonstrates the performance of the SCX separtion. The quality of the linear gradient separation is indicated by the fact that most of the peptides are only found in one fraction. SingleMS/MS spectra can be retrieved in a fast manner for manual inspection of the automated peptide and protein validation settings. As an example the single peptide MS/MS spectrum of Phosphopyruvate hydratase that was identified by 2D off-line is shown in figure 7. All identified fragments of the y- and b-series are indicated. The lower panel shows the detailed results after database search of all identified peptides from the protein. The proteins' identity, as well as its sequence, the sequence coverage, the species and the database accession number is displayed.



1 MAVSKVYARS VYDSRGNPTV EVELTTEKGV FRSIVPSGAS TGVHEALEMR DGDKSKWMGK GVLHAVKNVN DVIAPAFVKA 80
81 NIDVKDQKAV DDFLISLDGT ANKSKLGANA ILGVSLAASR AAAAEKNVPL YKHLADLSKS KTSPYVLPVP FLNVLNGGSH 160
161 AGGALALQEF MIAPTGAKTF AEALRIGSEV YHNLKSLTKK RYGASAGNVG DEGGVAPNIQ TAEEALDLIV DAIKAAGHDG 240
241 KIKIGLDCAS SEFFKDGKYD LDFKNPNSDK SKWLTGPQLA DLYHSLMKRY PIVSIEDPFA EDDWEAWSHF FKTAGIQIVA 320
321 DDLTVTNPKR IATAIEKKAA DALLLKVNQI GTLSESIKAA QDSFAAGWGV MVSHRSGETE DTFIADLVVG LRTGQIKTGA 400
401 PARSERLAKL NQLLRIEEEL GDNAVFAGEN FHHGDKL 437

The matched peptides cover **42%** (184/437 AA's) of the protein.

Protein Name: **phosphopyruvate hydratase (EC 4.2.1.11) 1 [validated]**
Species: **yeast**
NCBInr Accession #: **1323462**

Figure 7
MS/MS spectrum of one identified peptide from Phosphopyruvate hydratase. The total sequence coverage is shown below the spectrum.

Conclusion

Off-line 2D LC-MS/MS represents a powerful alternative to on-line methodologies for protein identification from complex proteomes. In such an approach the trypsin-digested peptides are eluted with a continuous salt gradient from a SCX column and collected as small fractions. These fractions are subsequently separated on a RP column, analyzed by ESI-MS/MS and identified with the Agilent Spectrum Mill MS Proteomics Workbench. In contrast, the conventional on-line technique elutes in the first dimension with discrete salt steps. In the second dimension the eluted peptides are directly separated on a RP column, analyzed by ESI-MS/MS and finally identified.

In this study we were able to demonstrate that in the off-line approach about 40 % more proteins were identified than in a comparable on-line approach. Furthermore, we could show that the scores for the identified proteins by both approaches are higher when applying the off-line technique. This means that the probability for false positive protein identification decreases when applying the off-line technique and therefore makes analyses more reliable. The total number of identified proteins in an off-line analysis can be increased further by performing the SCX dimension with a longer shallow gradient. This, however, requires the collection of additional fractions and significant longer analysis time. Of course, using less stringent parameters in the database search will also increase the number of identified proteins. However, these hits are presumably false positive identified proteins and their corresponding peptide fragmentations pattern needs a careful, time consuming manual inspection to decide their validity.

Furthermore, the off-line methodology adds more fexibility and a high degree of automation to a multidimensional chromatographic approach. Besides SCX and RP other additional separation or prefractionation techniques can be combined. Furthermore, after one dimension the collected samples might be stored for later repeated analysis or subjected to chemical or enzymatic modification before they are transferred to the next dimension.

A comparison of the applicability of both approaches indicates that the off-line approach is best suited for highly complex samples or when a separation step for further processing is necessary. In such a case the higher investment and more complex handling for an off-line system is justified.

References

1.
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 analysis". *J. Chromatogr. B 752: 281-291*, **2001**.

2.
Vollmer M, Nägele E, Hörth P. "Two-dimensional nano LC/MS for complex proteomics application". *Agilent Technolgies Application Note*, *Publication number 5988-9287EN*, **2003**.

3.
Washburn MP, Wolters D, Yates JR. "Large-scale analysis of the yeast proteome by multidimensional protein identification technology". *Nat. Biotechnol. 19: 242-247*, **2001**.

4.
Nägele E, Vollmer M, Hörth P. "Proteome profiling in *E. coli*: Effect of heat-shock conditions on protein expression pattern". *Agilent Technolgies Application Note*, *Publication number 5988-8629EN*, **2003**.

5.
Agilent micro-fraction collection system, *Agilent Product Brochure*, *Publication Number 5988-9614EN*, **2003**.

6.
Agilent Nanoflow Proteomics Solution, *Agilent Product Brochure*, *Publication Number 5988-3612EN*, **2002**.

7.
Agilent Spectrum Mill MS Proteomics Workbench, *Agilent Product Brochure*, *Publication Number 5988-9443EN*, **2003**.

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

www.agilent.com/chem/proteomics

Published August 1, 2003
Publication Number 5988-9913EN

Agilent Technologies