HPLC 2004

Multidimensional LC/MS Approaches for the Analysis of Proteomes and Biofluids

K. Millea[2], A. B. Chakraborty[1], D. B. Wall[1], S. A. Cohen[1], I. S. Krull[2], S. J. Berger[1]
[1] Life Sciences R&D, Waters Corporation, Milford, MA; [2]Department of Chemistry, Northeastern University, Boston, MA

Overview

- Current proteomic methodologies used for the analysis of protein mixtures follow two different workflows to characterize proteins within a sample, "Top-Down" or intact protein analysis approaches and "Bottom-Up" or complex proteolytic digestion approaches.
- A multidimensional chromatographic approach incorporating both on-line and off-line mass spectrometry was used to combine both methodologies, and extract a greater amount of information from proteomic samples.
- Reducing dynamic range by chromatographic resolution of high abundance components from lower abundance components provides for better detection of the lower level components.
- Protein analysis of *E. coli* cytosol and human cerebrospinal fluid (CSF) was performed using multidimensional LC, with online ESI-Tof MS of the intact proteins, and MALDI Q-Tof analysis of collected fractions.
- "Top-Down" intact protein chromatographic fractionation was shown to produce simplified protein fractions, and can be combined with "Bottom-Up" analysis of the fractions for a more complete characterization of proteomic components.

Instrumentation

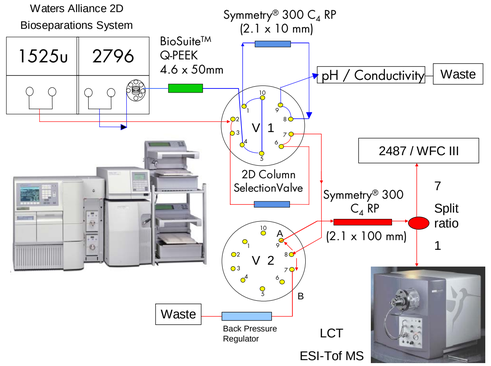


Figure 1: Multidimensional LC/ESI-Tof MS Configurations. A Waters Alliance® 2D Bioseparations System was used for online IEX/RP separations coupled to a Waters Micromass® LCT™ ESI-Tof MS and WFCIII fraction collector. 10-port 2-position valves were used to control solvent flow to allow trapping of analytes, column washing, separation and detection by the online mass spectrometer. A post-valve split permitted simultaneous fraction collection and ESI-Tof MS analysis. Collected fractions were proteolytically digested and analyzed using a Waters Micromass® Q-Tof Ultima™ MALDI MS system. MassLynx™ 4.0 was used for instrument control, data acquisition, and MS data processing. ProteinLynx™ Global Server was used for Q-Tof MS/MS data analysis, and protein database searching against the SwissProt database (http://www.expasy.org).

Results

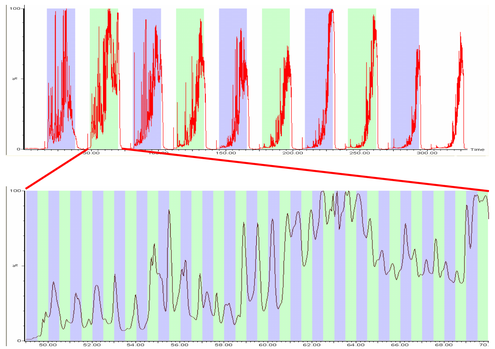


Figure 2: Comprehensive 2D (IEX/RP) Step Chromatographic Separation with Fraction Collection. 1.6 mg of *E. coli* cytosol was injected on the MDLC system for fractionation. Proteins were eluted from the 1D SAX column by 8-50 mM NaCl steps, followed by two further salt steps up to 1 M NaCl (pH 8). The 2nd dimension reversed phase separation was accomplished using a 10-55% ACN gradient containing 0.001%TFA/0.1% formic acid as a modifier. 30 sec fractions were collected during the 9 reversed phase cycles (colored bars). Collected fractions were lyophilized, reduced, alkylated and tryptically digested in the presence of 0.1% RapiGest™.

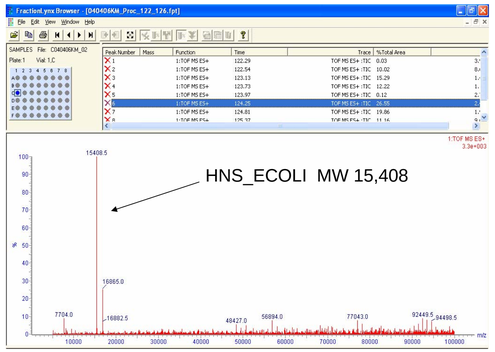


Figure 3: Automated ESI-Tof MS Spectral Analysis Using FractionLynx™: FractionLynx was used for automated data analysis of ESI-Tof data by performing peak identification, spectral summation, background subtraction, and MaxEnt1 deconvolution for identified TIC peaks. Multiple peaks and all collected fractions were analyzed in a single overnight processing run. The FractionLynx browser (above) depicts the deconvoluted spectrum for a protein with mass 15,408 Da, later identified as HNS_ECOLI during database searching of the Q-Tof MS/MS peptide digest data for that fraction.

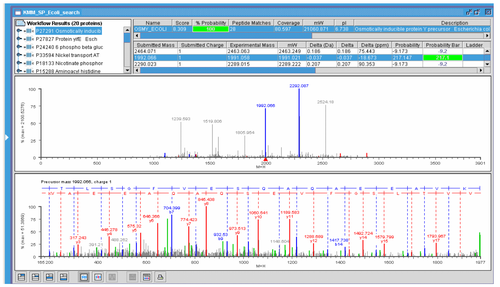
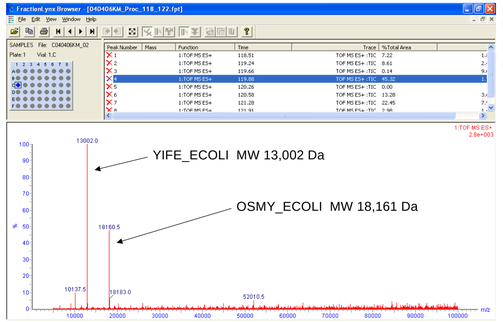


Figure 4: MALDI Q-Tof MS/MS Identification of *E. coli* protein OSMY_ECOLI by database searching with ProteinLynx™ Global Server v2.1. MALDI MS and MS/MS spectra were processed using ProteinLynx Global Server software to identify proteins present within the collected fractions. The results from the database searches were then integrated with the ESI-Tof data to provide a comprehensive picture of the proteins in collected fractions. In this example the protein OSMY_ECOLI was identified. An overall sequence coverage of 80% was obtained with an average mass error of 60 ppm, for the predicted 21kD protein.



Signal Sequence Processed Protein Calculated Mass : 18,161 Da

MTMTRLKISKTLLAVMLTSAVATGSAYAENNAQTTNESAGQK
VDSSMNKVGNFMDDSAITAKVKAALVDHDNIKSTDISVKTDQ
KVVTLSGFVESQAQAEEAVKVAKGVEGVTSVSDKLHVRDAK
EGSVKGYAGDTATTSEIKAKLLADDIVPSRHVKVETTDGVVQL
SGTVDSQAQSDRAESIAKAVDGVKSVKNDLKTK

Figure 5: Identification of a Processed Form of the *E. coli* Protein OSMY_ECOLI lacking its N-terminal Signaling Sequence. Combining the data from intact and digest MS analysis we were able to identify a truncated form of the *E. coli* protein OSMY_ECOLI lacking the N-terminal 28 amino acids. The loss of the leader sequence reduced the calculated mass from 21,074 Da to 18,161 Da in line with our observations. A search of the SwissProt database confirms the processing site predicted from our mass data (Protein OSMY_ECOLI Accession Number P27291).

#	Sprot Name	Accession #	#	Sprot Name	Accession #	#	Sprot Name	Accession #
1	ACFD_ECOLI	Q46837	34	KAD_ECOLI	P05082	67	SLT_ECOLI	P03810
2	ADHP_ECOLI	P39451	35	KITH_ECOLI	P23331	68	SODC_ECOLI	P53635
3	AGP_ECOLI	P19926	36	LHR_ECOLI	P30015	69	SODM_ECOLI	P00448
4	ALF1_ECOLI	P71295	37	MALG_ECOLI	P07622	70	STFR_ECOLI	P76072
5	ALKH_ECOLI	P10177	38	MALK_ECOLI	P02914	71	SUCD_ECOLI	P07459
6	ALN_ECOLI	P77671	39	MAOC_ECOLI	P77455	72	SYH_ECOLI	P60906
7	ARPB_ECOLI	P76205	40	MDH_ECOLI	P06994	73	TALA_ECOLI	P78258
8	BASR_ECOLI	P30843	41	META_ECOLI	P07623	74	TNAA_ECOLI	P00913
9	BCSC_ECOLI	P37650	42	NARG_ECOLI	P09152	75	WECF_ECOLI	P56258
10	BGAL_ECOLI	P00722	43	NARL_ECOLI	P10957	76	WZYE_ECOLI	P27835
11	CATE_ECOLI	P21179	44	NDK_ECOLI	P24233	77	XDHA_ECOLI	Q46799
12	CH10_ECOLI	P05380	45	NIRB_ECOLI	P08201	78	XYLR_ECOLI	P37390
13	CH60_ECOLI	P06139	46	OSMY_ECOLI	P27291	79	YAGA_ECOLI	P37007
14	CISY_ECOLI	P00891	47	PAAC_ECOLI	P76079	80	YAGY_ECOLI	P77188
15	CLPB_ECOLI	P03815	48	PBPC_ECOLI	P76577	81	YBBK_ECOLI	P77367
16	CRP_ECOLI	P03020	49	PDXA_ECOLI	P19624	82	YBCH_ECOLI	P37325
17	CSPC_ECOLI	P36996	50	PFLB_ECOLI	P09373	83	YCCJ_ECOLI	P46131
18	CSRA_ECOLI	P31803	51	PFLD_ECOLI	P32674	84	YCIN_ECOLI	P46132
19	DAPE_ECOLI	P24176	52	PGPB_ECOLI	P18201	85	YDBA_ECOLI	P33666
20	DAPF_ECOLI	P08885	53	PNTA_ECOLI	P07001	86	YDEV_ECOLI	P77432
21	DBHA_ECOLI	P02342	54	PPIB_ECOLI	P23869	87	YDHR_ECOLI	P77225
22	DCEA_ECOLI	P80063	55	QOR_ECOLI	P28304	88	YEBU_ECOLI	P76273
23	DCLZ_ECOLI	P52095	56	RBSB_ECOLI	P02925	89	YEBV_ECOLI	P76274
24	DEAD_ECOLI	P23304	57	RECQ_ECOLI	P15043	90	YEFG_ECOLI	P37749
25	DEOC_ECOLI	P00882	58	RFAD_ECOLI	P17963	91	YEGE_ECOLI	P38097
26	DKGA_ECOLI	Q46857	59	RFAY_ECOLI	P27240	92	YEGK_ECOLI	P76395
27	DNAK_ECOLI	P04475	60	RHSC_ECOLI	P16918	93	YEGP_ECOLI	P76402
28	DPO2_ECOLI	P21189	61	RHSD_ECOLI	P16919	94	YFHH_ECOLI	P37767
29	EUTB_ECOLI	P19635	62	RL10_ECOLI	P02408	95	YFID_ECOLI	P33653
30	FABG_ECOLI	P25716	63	RL19_ECOLI	P02420	96	YFIM_ECOLI	P46126
31	FKBA_ECOLI	P45523	64	RL24_ECOLI	P60624	97	YIFE_ECOLI	P27827
32	FRUR_ECOLI	P21168	65	RL25_ECOLI	P02426	98	YGEA_ECOLI	P03813
33	FTNA_ECOLI	P23887	66	RL27_ECOLI	P02427	99	YGFT_ECOLI	Q46820

Figure 6: Collated List of Proteins Identified. A partial list of proteins identified using MALDI MS/MS is shown above, most having been confirmed by ESI-Tof. ESI-Tof MS provides additional information on processing and posttranslational modification that can be useful in the study of complex proteomes.

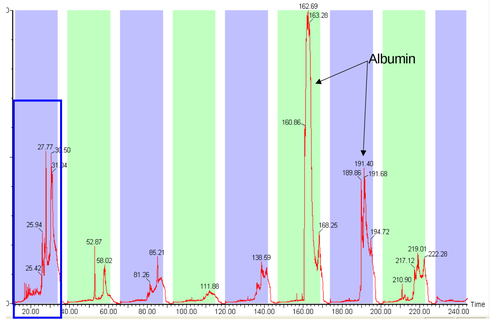


Figure 7: Comprehensive 2D (IEX/RP) Step Chromatographic Separation of Human Cerebrospinal Fluid. 540 ug of human CSF protein was injected on the MDLC system for fractionation. Proteins were eluted off the 1D SAX column by 50 mM NaCl steps at pH 8. The 2D reversed phase separation was done using a 5-45% ACN gradient in 0.005%TFA/0.1% formic acid. Fraction collection was done as shown above. The 2D separation protocol was effective in separating the highly abundant serum albumin protein from other mixture components. This allowed for analysis of the underlying low level proteins and peptides within the sample. A closer view of the boxed region is shown in Figure 8.

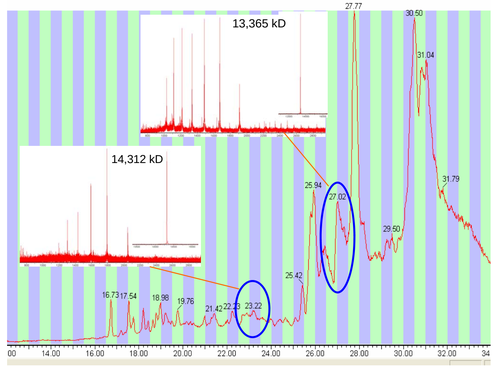


Figure 8: MDLC/ESI-Tof-MS of a Human CSF SAX Fraction. Proteins of lower abundance can be identified within the sample matrix once separated from serum albumin. The above figure shows one of the second dimension reversed-phase separation steps, with collected fractions indicated by the colored bars. The figure insets show raw and deconvoluted spectra of two proteins observed during the run. MaxEnt1 deconvolution yielded intact protein masses of 13,365 kD and 14,312 kD for the two components. Proteolytic digestion of the collected fractions is ongoing, and will be used to characterize the proteins present, using a similar MALDI Q-Tof based peptide analysis strategy.

Conclusions

- Complex proteomic samples were fractionated using multidimensional chromatography to reduce sample complexity and increase the effective dynamic range prior to intact protein and tryptic digest mass analysis.
- On-line ESI-Tof MS detection of proteins during the chromatographic fractionation allowed for intact protein mass data to be combined with tryptic peptide digest Q-Tof MS/MS analysis of the collected fractions.
- Combining intact mass and proteolytic digest information allows for increased confidence in protein identification and assignment of modifications for proteins in their biologically relevant form.
- The combined analysis, yielding both intact ESI-Tof MS and tryptic digest MALDI MS/MS data, was shown to unambiguously identify biologically significant proteolytic processing sites of several periplasmic *E. coli* proteins.
- Chromatographic resolution of high abundance components from lower abundance components provides for better detection of low level components in intricate sample matrices.
- Samples with extreme dynamic range (e.g. many biofluids) may still require depletion of abundant proteins to permit optimum column loading for detection of lower abundance components.
- Chromatographic separations offer a high degree of flexibility for proteomic analyses for dealing with samples exhibiting a wide range of sample complexity.