An Integrated Approach to Improve Sequence Coverage and Protein Identification by Combining LC-MALDI MS/MS and Nano-LC/MS/MS

Application Note

Ning Tang and Christine Miller
Agilent Technologies

This material was presented in a poster at the 53rd ASMS Conference on Mass Spectrometry in 2005. Research presented in posters at scientific conferences may include results from instruments or products that are not yet commercially available.

Introduction

Electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) have proven to be complementary as different peptides are typically seen with the two ionization techniques, however, very little was known about the characteristics of the peptides that prefers each ionization technique. Coupling of the RP-HPLC with MALDI via a MALDI spotting device has provided an automated and robust solution for analyzing medium complexity samples using MALDI. In this study, we have investigated the improvements in proteome coverage for complex protein samples by combining nano-LC/MS/MS with LC-MALDI/MS/MS. The integrated approach with better sequence coverage facilitated protein identification.

Experimental

Three dimensions of separations were employed to simplify the human serum sample before it was analyzed by the mass spectrometry (see Figure 1). The first dimension of separation was immunodepletion of six of the most abundant proteins in human serum using an antibody column. The second dimension was strong cation exchange (SCX) fractionation of the tryptic peptides. The third dimension was reversed-phase chromatography of the SCX fractions before mass spectrometry analysis.

Materials

HPLC grade solvents were obtained from Burdick & Jackson. Human serum was purchased from Genomics Collaborative (Cambridge, MA). It was from a healthy anonymous male donor. The immunodepletion column and reagents, the α-cyano-4-hydroxy-cinnamic acid (CHCA) matrix solution, and the in-gel digestion kit were from Agilent Technologies.

Immunodepletion

Human serum was immunodepleted using the Agilent Multiple Affinity Removal system (MARS), which removes six of the highest abundance proteins in serum. The depleted serum was then denatured, reduced, alkylated and digested with trypsin.

Offline SCX separation

Offline capillary LC prefractionation was done using a Zorbax Bio-SCX Series II column, 50 x 0.8 mm, with an ammonium formate gradient. The gradient went from 0–100 mM in 120 minutes, then at 500 mM at 134 minutes and finally to 1 M at 140 minutes. Fifty fractions were collected. Fractions were dried, washed twice with H_2O to remove the ammonium bicarbonate, and then dried before LC-MALDI and direct MALDI analysis.

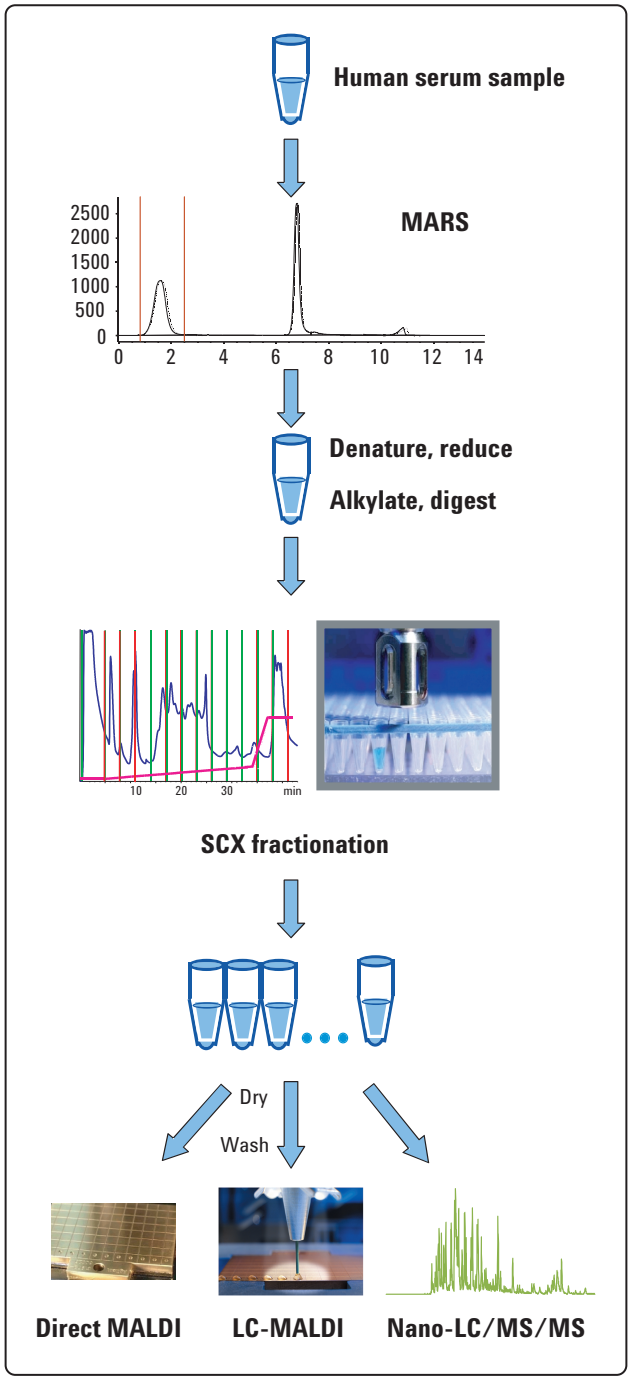


Figure 1. Experimental design

LC-MALDI MS/MS

Samples were separated using a capillary HPLC equipped with a thermostatted micro fraction collector/spotter (Agilent Technologies). The flow rate was 1 µL/min. The matrix CHCA (2 mg/mL diluted with 50% 2-propanol/1% acetic acid) was added to the sample at 1 µL/min using a syringe pump connected to a T-connector. The spotter spotted every minute with 2 µL of total volume on the MALDI target. The target was then analyzed using the LC/MSD Trap XCT Plus ion trap mass spectrometer with a PDF-MALDI source.

Nano-LC/MS/MS

All LC/MS analyses were done using the Agilent Protein Identification Solution which consisted of an Agilent 1100 Series nanoflow liquid chromatograph (LC) system coupled to an Agilent 1100 Series LC/MSD Trap XCT Plus.

Database search

Data processing was done using Spectrum Mill MS proteomics workbench software. The spectra were searched against the IPI-human database using carbamidomethylation as the Cys modification. The search was done with three iterative cycles. The first search was done in the identity mode against the IPI-human database with trypsin cleavage specified and 2 missed cleavages allowed. The validated search results were saved as a new database for the next search cycle. The second search cycle was done against the smaller saved results database to search for single KQMSTY modifications. The results were then validated and saved. The third search cycle was done against the saved results database with the enzyme selection set to none.

Results

Comparison of LC-MALDI MS/MS with direct MALDI MS/MS

The SCX fractions were run by both MALDI MS/MS and LC-MALDI MS/MS. The number of proteins and peptides identified by each technique were summarized in Table 1. LC-MALDI identified 3–4 fold more proteins and 5–6 fold more peptides compared to direct MALDI. Sequence coverage of identified proteins increased by LC-MALDI, which also provided better confidence in the protein ID.

Table 1. Comparison of the number of proteins and peptides identified by LC-MALDI MS/MS and direct MALDI MS/MS

	# of proteins	# of peptides
Direct MALDI	23	56
LC-MALDI	76	313

Combine results from LC-MALDI MS/MS and nano-LC/MS/MS

We have identified 129 proteins in total combining the results from both techniques. Out of the 129 proteins, 40 of them were identified by both techniques. Forty proteins were only identified by LC-MALDI MS/MS and 49 proteins were uniquely identified by nano-LC/MS/MS. Combining the two techniques has increased the number of proteins identified as well as sequence coverage and database search scores.

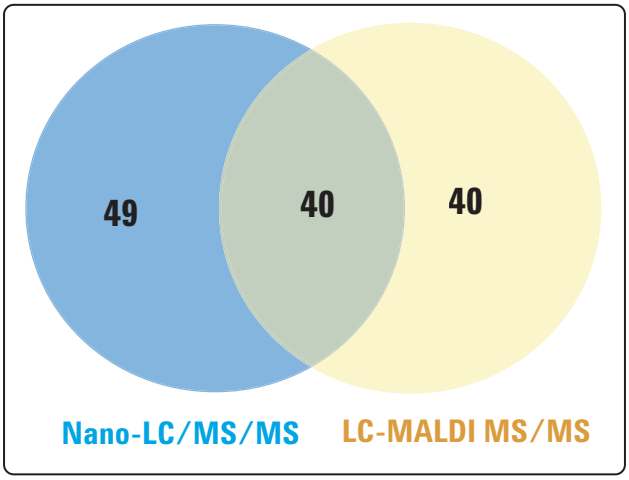


Figure 2. Combining the results from nano-LC/MS/MS and LC-MALDI MS/MS

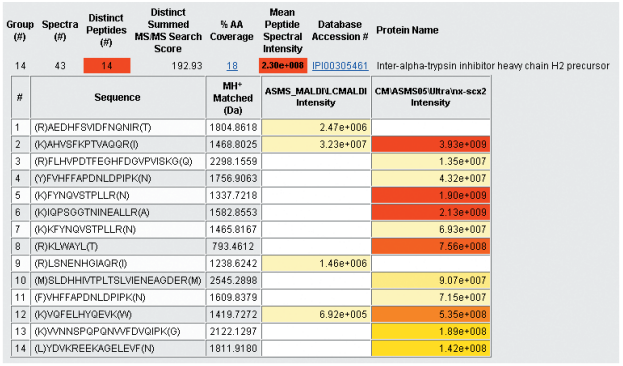
Comparison of LC-MALDI MS/MS with nano-LC/MS/MS

The number of proteins and peptides identified by each technique were summarized in Table 2. LC-MALDI MS/MS and nano-LC/MS/MS identified similar number of proteins. However, nano-LC/MS/MS identified 2–3 fold more peptides, therefore, it provided higher sequence coverage and more confident protein ID. Combining the two techniques, we have identified more proteins with more peptides.

Table 2. Comparison of the number of proteins and peptides identified by LC-MALDI MS/MS and nano-LC/MS/MS

	# of proteins	# of peptides
Nano-LC/MS/MS	89	876
LC-MALDI	76	313
Combination	129	1044

Each technique identified unique peptides among the proteins found by both methods. Combining the results increased the sequence coverage of the proteins. Figure 3 is an example of Spectrum Mill database search results of one of the protein that was identified as inter-alpha-trypsin inhibitor heavy chain H2 precursor.



Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	Database Accession #	Protein Name
14	43	14	192.93	18	2.30e+008	IPI00305461	Inter-alpha-trypsin inhibitor heavy chain H2 precursor

#	Sequence	MH+ Matched (Da)	ASMS_MALDI\LCMALDI Intensity	CM\ASMS05\Ultra\nx-scx2 Intensity
1	(R)AEDHFSVIDFNQNIR(T)	1804.8618	2.47e+006	
2	(K)AHVSFKPTVAQQR(I)	1468.8025	3.23e+007	3.93e+009
3	(R)FLHVPDTFEGHFDGVPVISKG(Q)	2298.1559		1.35e+007
4	(Y)FVHFFAPDNLDPIPK(N)	1756.9063		4.32e+007
5	(K)FYNQVSTPLLR(N)	1337.7218		1.90e+009
6	(K)IQPSGGTNINEALLR(A)	1582.8553		2.13e+009
7	(K)KFYNQVSTPLLR(N)	1465.8167		6.93e+007
8	(R)KLWAYL(T)	793.4612		7.56e+008
9	(R)LSNENHGIAQR(I)	1238.6242	1.46e+006	
10	(M)SLDHHIVTPLTSLVIENEAGDER(M)	2545.2898		9.07e+007
11	(F)VHFFAPDNLDPIPK(N)	1609.8379		7.15e+007
12	(K)VQFELHYQEVK(W)	1419.7272	6.92e+005	5.35e+008
13	(K)VVNNSPQPQNVVFDVQIPK(G)	2122.1297		1.89e+008
14	(L)YDVKREEKAGELEVF(N)	1811.9180		1.42e+008

Figure 3. Spectrum Mill display of all the peptides identified by LC-MALDI MS/MS and nano-LC/MS/MS

Discussion

We compared our results with the nonredundant list developed by *Anderson et al.* from a combination of four separate sources (see Table 3). Thirty seven of the proteins we identified were also identified by those other 4 sources. These proteins we have identified with relatively high sequence coverage and Spectrum Mill database search scores. They represented the high abundant proteins and are more likely to be identified most of the time regardless of the different separation techniques and mass spectrometry that was used. Twenty one proteins we identified were also identified by three other sources. They have lower sequence coverage and Spectrum Mill scores. Twelve proteins were identified by two other sources and fifty nine proteins were identified by less than two other sources. They have relatively low sequence coverage and low spectrum mill scores.

Table 3. Correlation of Agilent results with nonredundant list comprised of data from four separate sources

ID by other sources	# of proteins	Average # of peptides per protein	Average sequence coverage (%)	Average Spectrum Mill scores
4	37	15	26	210
3	21	11	12	143
2	12	3	11	36
<2	59	4	7	40

Conclusions

- LC separation coupled with MALDI has significantly increased the number of proteins identified.
- LC-MALDI MS/MS and nano-LC/MS/MS have each identified unique peptides among the proteins found by both techniques and combining the results gives increased sequence coverage for most of the proteins.
- Each technique has identified unique proteins. LC-MALDI MS/MS has identified 40 unique proteins and nano-LC/MS/MS has identified 49 unique proteins. Combining both techniques, 129 proteins were identified.
- Combining both techniques increased the confidence in the protein identification and gave the optimum sequence coverage for the proteins.

References

1. Nagele, E. and Vollmer, Coupling of nanoflow liquid chromatography to matrix-assisted laser desorption/ionization mass spectrometry: real-time liquid chromatography run mapping on a MALDI plate, *M. Rapid Commun. Mass spectrum.* 2004; **18:** 3008
2. Anderson, N. L. *et al.*, The human plasma proteome, *Mol. Cell. Proteomics,* 2004; **3.4:** 311–326.

www.agilent.com/chem

Printed in the U.S.A. July 29, 2005
5989-3500EN

Agilent Technologies