

# Profiling Approach for Biomarker Discovery using an Agilent HPLC-Chip Coupled with an Accurate-Mass Q-TOF LC/MS

## Application Note

### Authors

Sudha Rajagopalan and Ravindra Gudihal  
Agilent Technologies India pvt. Ltd.  
Bangalore India

Keith Waddell  
Agilent Technologies, Inc.  
Santa Clara, CA USA

### Abstract

A profile-directed biomarker discovery approach involves initial differential expression analysis followed by targeted identification of differentially expressed proteins. In the present study, a very small amount (100 femtomole and 10 femtomole) of horseradish peroxidase was spiked into human plasma and a profile-directed approach was used for the identification of the spiked protein. Excellent reproducibility in retention time and intensity measurement, and outstanding accuracy in mass measurement and sensitivity over a wide dynamic range, were achieved using various Agilent platforms such as a Multiple Affinity Removal (MARS) column, a 1200 Series HPLC-Chip/MS and a 6520 Accurate-Mass Quadrupole Time-of-Flight (Q-TOF) LC/MS. Differential analyses using MassHunter Mass Profiler and MassHunter Mass Profiler Professional (MPP) software enabled the comparison between spiked and non-spiked plasma samples. Subsequent targeted MS/MS analysis on a Q-TOF LC/MS followed by a database search using Spectrum Mill helped identify the spiked protein.

### Introduction

Human plasma is extensively used for clinical diagnosis and early disease detection. Plasma is composed of classical blood proteins as well as tissue leakage proteins present in a wide dynamic range of concentrations. Highly sensitive, accurate and reproducible techniques are required to identify biomarkers in such complex biological samples. Mass spectrometry is used extensively in biomarker discovery and validation processes, especially in cancer biomarker studies.<sup>1</sup> Due to the volume and complexity of data generated in a biomarker study, sophisticated software tools are necessary for data analysis and visualization. This application note describes a profile-directed label free quantitation approach for the identification of a spiked plant protein in a human plasma sample. Excellent reproducibility in retention time and intensity measurements, along with accuracy in mass measurement and sensitivity over a wide dynamic range of concentrations offered by Agilent's 1200 Series HPLC-Chip/MS and 6520 Accurate-Mass Q-TOF LC/MS, have been demonstrated in the present work. Agilent also offers advanced software tools required for data analysis and statistical comparison.



**Agilent Technologies**

## Study design

Human plasma was depleted to remove 14 highly abundant proteins before digestion. A digest of a plant protein, horseradish peroxidase, was spiked at different concentration levels into the digested plasma and analyzed in four replicates using an Agilent 6520 Accurate-Mass Q-TOF LC/MS configured with a 1200 Series HPLC-Chip/MS. Replicate plasma digest samples without spiked peroxidase were also analyzed as controls. Molecular features extracted from peroxidase-spiked and non-spiked LC/MS runs were compared using Mass Profiler and MassHunter MPP. These features that were statistically differentiated in the spiked samples were identified using targeted MS/MS analysis on a Q-TOF LC/MS followed by database searching against the SwissProt database using Spectrum Mill. The experimental work flow and various Agilent platforms used in the study are shown in **Figure 1**.

## Experimental

### Sample preparation

Human plasma samples and peroxidase were obtained from Sigma. Fourteen of the highly abundant proteins in 125  $\mu$ L of human plasma were depleted using Agilent's 10 x 100 mm MARS column as described.<sup>2</sup> Proteins that did not bind to the affinity column (eluted in the flow-through) were collected and buffer-exchanged with 100 mM ammonium bicarbonate using 5 kDa molecular weight cut-off spin concentrators.

The protein concentration in the depleted plasma was measured using a DC protein assay kit (Bio-Rad) as described by the vendor. A 50  $\mu$ g aliquot of the depleted plasma was reduced with dithiothreitol (DTT), alkylated using iodoacetamide and digested using 25:1 (v/v) trypsin (Agilent). A quantity of 250 picomole of peroxidase was similarly reduced, alkylated and digested. Equal amounts (2  $\mu$ g) of digested plasma samples were spiked with varying amounts (0, 100 femtomole and 1 picomole) of peroxidase digest. A 200 ng aliquot of the plasma digest (spiked or non-spiked) was loaded onto the chip in each LC/MS analysis.

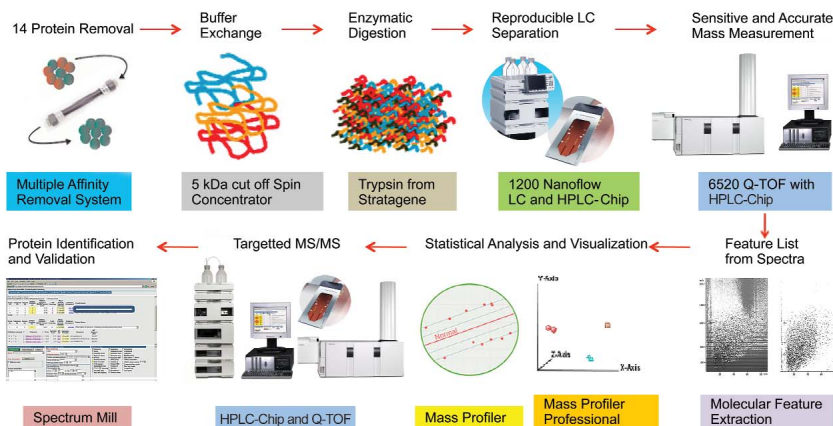


Figure1. Workflow used for the biomarker model study.

## LC/MS analysis

All LC/MS experiments were performed on an Agilent 1200 Series HPLC-Chip/MS interfaced to an Agilent 6520 Accurate-Mass Q-TOF LC/MS. The LC system consisted of a capillary pump for sample loading, a nanoflow pump and a thermostated microwell-plate autosampler. The HPLC-Chip configuration consisted of a 160 nL enrichment column and a 150 mm x 75  $\mu$ m analytical column (Zorbax 300SB-C18). Mobile phases employed were: A). 0.1% formic acid in water and B). 90% acetonitrile with 0.1% formic acid. An 80 min long gradient method was used for the LC separation. Sample loading onto the enrichment column was done at 3% B. The gradient used for the analytical column began at 3% B, was raised to 45% B at 45 min, 70% B at 55 min, maintained at 70% B until 58 min, raised to 95% B at 68 min and then brought back to 3% B at 75 min. The column was equilibrated for 5 min before subsequent runs. Samples were loaded at 4  $\mu$ L/min flow rate and eluted at 400 nL/min.

An Agilent 6520 Accurate-Mass Q-TOF LC/MS operating in high resolution (4 GHz) positive ion mode was used for all experiments. The MS source conditions were:

source temperature: 305°C  
capillary voltage: 2,000 V  
fragmentor voltage: 175 V  
drying gas flow rate: 5 L/min

Data was acquired between  $m/z$  300-3,000 at a scan rate of 1 spectra/sec for all samples in the profiling experiments. The number of LC/MS runs performed and samples used in the profiling experiment are described in **Table 1**. In the targeted MS/MS experiment, data was acquired from 300-3,000  $m/z$  with an acquisition rate of 4 spectra/sec in MS mode and  $m/z$  region 50-3,000 with an acquisition rate of 5 spectra/sec in MS/MS mode. Complete system control was achieved using Agilent MassHunter data acquisition software (B.02.00).

Table 1. Details of LC/MS runs performed in the profiling experiment.

Sample No	Amount of peroxidase spiked on-column	Plasma digest loaded on chip in each LC/MS run	Number of replicates
1	0	200 ng	4
2	10 femtomole	200 ng	4
3	100 femtomole	200 ng	4

## Data analysis

Data analysis was performed using the following software packages: Agilent MassHunter Qualitative Analysis (version B02.00), Mass Profiler (version B01.00), MPP (version B02.00), and Spectrum Mill (version B03.00). LC/MS data were extracted and evaluated using a specialized molecular feature extractor (MFE) algorithm in MassHunter Qualitative Analysis software. Mass Profiler and MPP were used for statistical evaluation of technical reproducibility and for comparison of spiked versus non-spiked (differential) samples. Spectrum Mill was used for protein identification from differential features.

## Results and Discussion

Total ion chromatograms of four replicate LC/MS analyses of sample 1 are shown in **Figure 2** demonstrating the excellent reproducibility achieved in the LC/MS runs.

MFE takes raw data as the input and outputs a list of molecular features. A molecular feature represents a chemical entity such as a compound or a peptide. MFE reports a feature combining the abundance information of all its isotope clusters, different ion species such as multiple charge states, as well as dimers and adducts. Nearly 25,000 molecular features were observed in the LC/MS runs of plasma digest demonstrating the complexity of the sample (**Figure 3**). Retention times, masses and abundances of all features from each LC/MS run were compiled by MFE into feature lists.

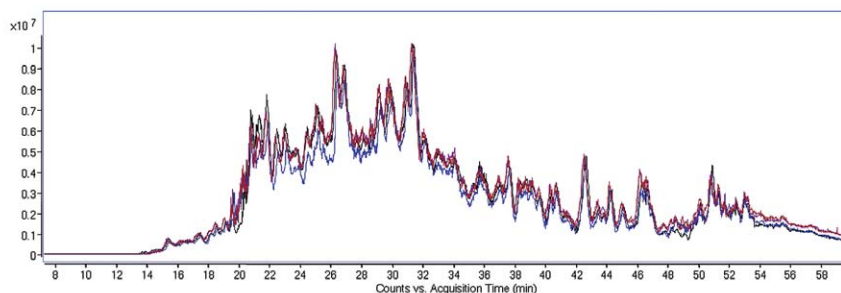


Figure 2. Total ion chromatogram of four replicate LC/MS runs of plasma digest.

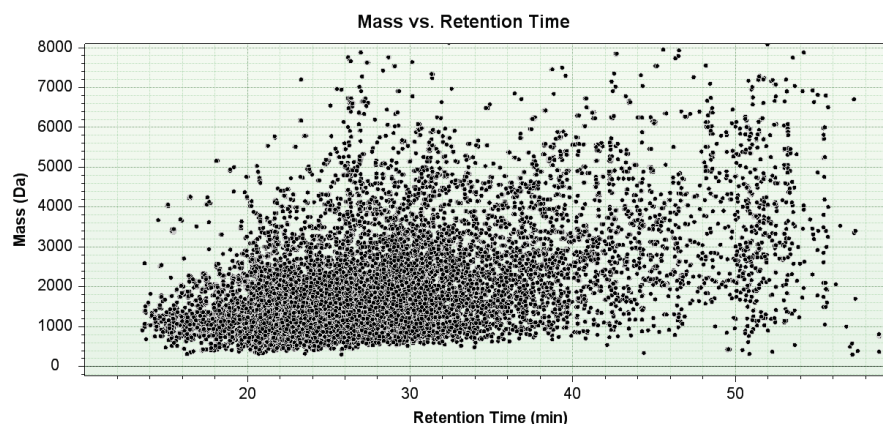


Figure 3. Mass versus retention time plot of the molecular features observed in LC/MS runs of plasma digest.

Statistical analyses were performed on molecular features to check the reproducibility of the LC/MS analysis. A limited portion of the molecular feature list from four replicate LC/MS runs of plasma digest is shown in **Table 2**. Retention time deviation among the four replicate LC/MS runs is within  $\pm 4$  sec (maximum being 0.06 min), which demonstrates the excellent retention time reproducibility achieved using Agilent's microfluidic-based

HPLC-Chip. Precision of  $\pm 3$  ppm is achieved in mass measurement. The relative standard deviation of abundance values is less than 1% (maximum being 0.67%), which shows that the variation in measured intensities of the features between replicate runs is minimal. Excellent technical reproducibility achieved in the LC/MS analysis makes it suitable for comparison of samples with small biological differences.

Table 2. Portion of the molecular feature list obtained from four replicate LC/MS runs of plasma digest.

ID	RT Value (min)	RT S.D. (min)	Mass Value (Da)	Mass S.D. (Da)	Abundance Value	Abundance R.S.D.	# of replicates in which feature is observed
1	42.566	0.048	3676.683	0.0091	36736530	0.15	4
2	46.208	0.051	2365.185	0.0053	32621650	0.44	4
3	46.671	0.043	2237.092	0.005	24123180	0.09	4
4	44.294	0.044	3520.587	0.0088	21518870	0.11	4
5	50.837	0.039	3065.391	0.0059	17312580	0.17	4
6	46.537	0.039	2612.318	0.0055	10599180	0.14	4
7	29.708	0.027	1565.76	0.0044	8590402	0.67	4
8	36.686	0.037	1418.698	0.004	8310790	0.08	4
9	47.022	0.047	4034.905	0.0088	8189842	0.15	4
10	37.091	0.060	3079.582	0.0082	7957212	0.1	4
11	39.577	0.057	1473.818	0.0043	7098632	0.05	4
12	36.914	0.045	2689.265	0.0073	5942096	0.08	4
13	42.693	0.037	3345.398	0.0086	5780965	0.13	4
14	50.078	0.034	2097.069	0.0038	5372915	0.18	4
15	35.485	0.019	1489.658	0.0043	4678271	0.07	4

## MassHunter MPP for checking the technical reproducibility

In a biomarker discovery process, the technical variations (variations during sample preparation and analysis) should be small to ensure that differences between samples reflect true biological differences between them. In the present study, there should be clear differentiation between the replicates of plasma digest with equal amounts of peroxidase than between these with a different amount of peroxidase.

Agilent's MPP software offers advanced visualization and statistical tools for this differential analysis of multiple samples. Within MPP, various clustering algorithms and visualization tools are available to identify samples with similar mass abundance patterns. Principal component analysis (PCA) is a clustering tool often applied to reduce the dimensionality of complex data sets. When samples are plotted according to their correlation coefficient for the first three PCs on a 3D scatter plot, clustering of the sample can be seen as shown in **Figure 4**. The brown squares in Figure 4 represent the replicates of sample 1, while the blue and red dots represent replicates of samples 2 and 3, respectively. Clustering of replicates in a PCA scatter plot confirms that the replicates of each sample are more similar to one another than replicates of other samples.

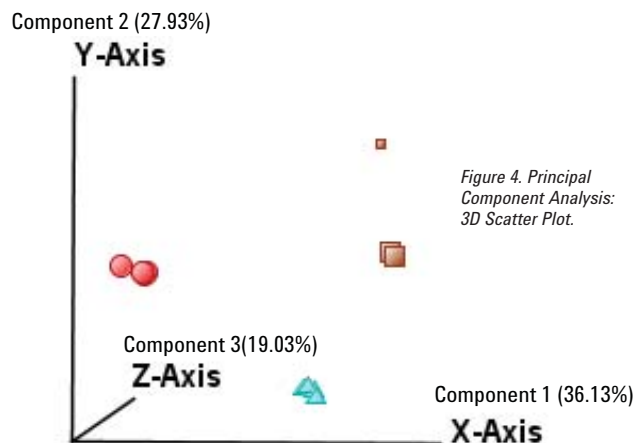


Figure 4. Principal Component Analysis: 3D Scatter Plot.

## Differential analysis

The protein profiling approach in biomarker discovery comprises of two steps: rapid differential expression analysis of samples in MS-only mode followed by identification from targeted MS/MS data of differentially-expressed putative markers. Mass Profiler software enables retention time and  $m/z$  alignment for features across samples, intensity normalization, and t-test statistics for identification of significant differences between samples. Molecular features extracted from replicate LC/MS analyses of sample 1 were compared with features from replicates of sample 2 using Mass Profiler. Differential features were searched against

the SwissProt database using peptide mass fingerprints of Spectrum Mill software. A screen shot of search results is shown in **Figure 5**. Peroxidase C1A precursor from horseradish is listed at the top among the identified proteins with a dynamic probability score of 0.0305.

A peptide mass fingerprint search against the SwissProt database using Spectrum Mill software on the differential features from the comparison of samples 1 and 3 also identified the protein peroxidase (data not shown). Six of the peroxidase peptides identified in the database search are shown in **Table 3**.

Figure 5. Database search of differential features in the comparison of samples 1 and 2.

<p>Sample ID (comment): 10fndif  Database searched: SwissProt  Molecular weight search (10000 - 40000 Da) selects 195305 entries.  Full pI range: 356194 entries.  Combined molecular weight and pI searches select 195305 entries.  PMF search selects 5 entries.</p>									
Parameters used in Search									
Rank	Dynamic Probability Score	Static Probability Score	# (%) Masses Matched	Mass Error Mean (Std Dev) (ppm)	Protein Coverage	Protein MW (Da)/pI	Species	Accession #	Protein Name
1	0.0305	4.21	4/26 (15%)	-29.5 (4.0)	9%	36825.3/5.67	ARMRU	P00433	Peroxidase C1A precursor - Armoracia rusticana (Horseradish) (Armoracia laphatifolia)
2	0.0376	5.15	4/26 (15%)	-29.5 (4.0)	9%	36847.3/6.21	ARATH	Q9LH89	Peroxidase 32 precursor - Arabidopsis thaliana (Mouse-ear cress)
3	0.0528	0.659	4/26 (15%)	-37.8 (11.1)	41%	10846.7/9.52	ECOLI	Q47149	Uncharacterized protein yafQ - Escherichia coli (strain K12)
4	1.04	12.3	4/26 (15%)	-31.8 (10.4)	11%	36298.3/5.25	BACSU	P42319	Uncharacterized protein ysjH - Bacillus subtilis
5	1.29	4.21	4/26 (15%)	-23.5 (12.1)	12%	28259.2/5.65	HYPNA	Q9C1C1	30S ribosomal protein S2 - Hyphomonas neptunium (strain ATCC 15444)

Table 3. Differential features obtained in the comparison of samples 1 and 3.

Feature No.	RT Value (min)	RT S.D. (min)	Mass Value (Da)	Mass S.D. (Da)	Abundance Value	Abundance R.S.D.	No. of runs observed
1	15.827	0.036	742.4529	0.0013	298859.6	1.64	4
2	18.849	0.025	802.4151	0.0016	816061	1.15	4
3	16.006	0.028	934.4484	0.0007	48349.07	1.51	4
4	18.275	0.056	958.4867	0.0006	539690.5	1.18	4
5	20.257	0.02	1474.707	0.003	211224.4	1.09	4
6	20.006	0.018	1585.787	0.0033	435146.3	1.19	4



The concentration of spiked peroxidase is ten-fold higher in sample 3 as compared to sample 2 described in Table 1. Hence, abundance levels of peroxidase peptides are expected to be up regulated in sample 3. Features extracted from samples 2 and 3 were compared using Mass Profiler. Log<sub>2</sub> abundance of 10 femtomole spiked samples was plotted against log<sub>2</sub> abundance of 100 femtomole spiked samples (**Figure 6**). Filtering the results for an 8-fold difference in abundance identified many of the peroxidase peptides, some of which are marked in Figure 6.

The mass spectrum of sample 3 between 18.09-18.24 min is shown in **Figure 7**. The doubly charged ion of a peroxidase peptide at *m/z* 480.2452 (mass value 958.49) in samples 2 and 3 are also shown in Figure 7. As shown in inset b, the intensity of the peak in a 10 femtomole peroxidase spiked sample is ten-fold less than the corresponding peak in a 100 femtomole peroxidase spiked sample (inset a).

Targeted MS/MS analysis of sample 3 was performed using the differential features as an inclusion list for precursor selection. MS/MS data from the targeted analysis was searched against the Swiss Prot database using Spectrum Mill software. A screen shot of the search results is shown in **Figure 8**. The occurrence of peroxidase at the top of the list of identified proteins confirms the identification of peroxidase. Sequences of three of the identified peroxidase peptides are shown in this figure.

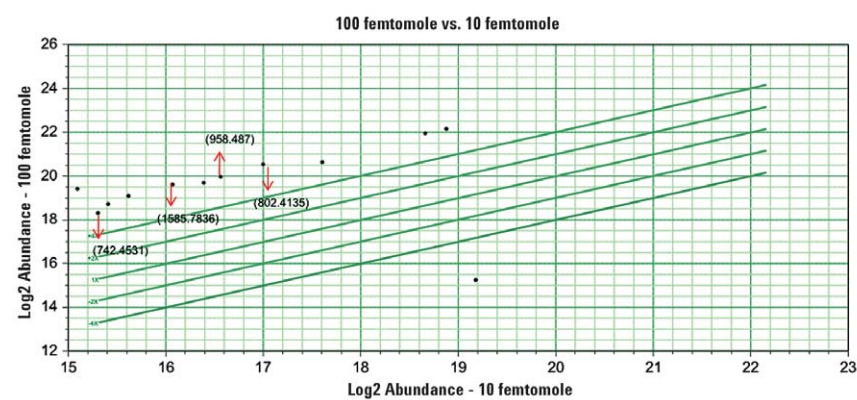


Figure 6. Comparison of sample 2 and 3.

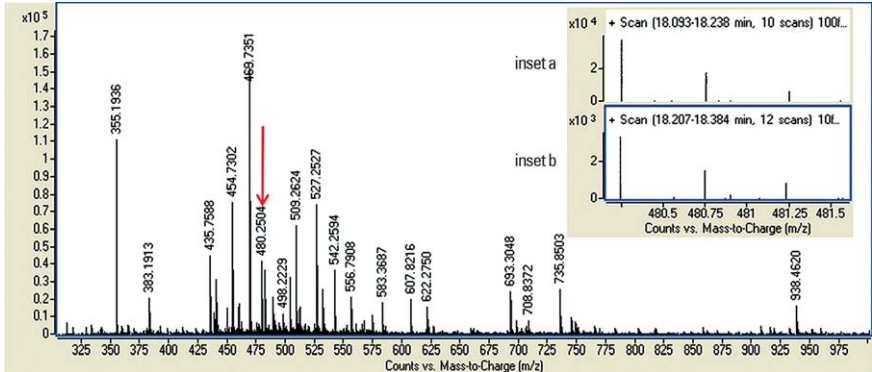


Figure 7. Mass spectrum of a 100 femtomole peroxidase spiked sample between 18.09 and 18.24 min. Inset shows the doubly charged ion at *m/z* 480.2 in the spectra of (a) 100 femtomole peroxidase spiked sample and (b) 10 femtomole spiked sample.

Agilent Spectrum Mill - Protein/Peptide Summary									
<a href="#">Spectrum Mill</a>   <a href="#">Summary Settings</a>   <a href="#">Autovalidation</a>   <a href="#">Easy MS/MS</a>   <a href="#">MS/MS Search</a>   <a href="#">Spectrum Summary</a>   <a href="#">Build TIC</a>   <a href="#">Tool Belt</a>   <a href="#">Help</a>									
protein groups ready for display proteinGroupingMethod: oneSharedPeptideSubgrouped 2 Proteins listed									
Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	Database Accession #	Protein Name	
1	1.1	3	3	43.57	8	2.74e+005	<a href="#">Q9LHB9</a>	Peroxidase 32 precursor - Arabidopsis thaliana (Mouse-ear cress)	
1	1.1	3	3	43.57	8	2.74e+005	<a href="#">P00433</a>	Peroxidase C1A precursor - Armoracia rusticana (Horseradish) (Armoracia laphatifolia)	
2	2.1	3	3	43.54	5	5.22e+005	<a href="#">P01011</a>	Alpha-1-antichymotrypsin precursor - Homo sapiens (Human)	
Totals:		9	9						
Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	Database Accession #	Protein Name	
1	1.1	3	3	43.57	8	2.74e+005	<a href="#">Q9LHB9</a>	Peroxidase 32 precursor - Arabidopsis thaliana (Mouse-ear cress)	
Validation category #	Filename	z Score	Fwd-Rev Score	SPI (%)	Spectrum Intensity	Sequence	MIT Matched (Da)		
✓ V C R	1 100femtomol1.3710.3710.2	2	13.16	3.11	88.3 2.02e+005	(R) IAAAILR (L)	743.477		
✓ V C R	2 100femtomol1.4798.4916.2	2	17.24	6.31	95.9 1.41e+005	(R) RGNITPLTGTGGQIR (Q)	1586.832		
✓ V C R	3 100femtomol1.4347.4370.2	2	13.17	3.14	93.2 4.78e+005	(R) DTTVNELR (S)	969.516		

Figure 8. Database search of MS/MS data from targeted analysis.

The MS/MS spectrum of the peroxidase peptide DTIVNELR (mass value 958.49) is shown in **Figure 9**. Even though the peak at  $m/z$  480.2504 is not the most intense peak in the spectrum of sample 3 (shown in Figure 7), a strong MS/MS spectrum is obtained in the targeted MS/MS analysis. Nearly 20 ions are co-eluting at the time of peak elution at  $m/z$  480.25 as seen in Figure 7. As this peak is the seventh in intensity, under classic data dependent conditions (top six from a MS), this peak may not get selected for MS/MS fragmentation.

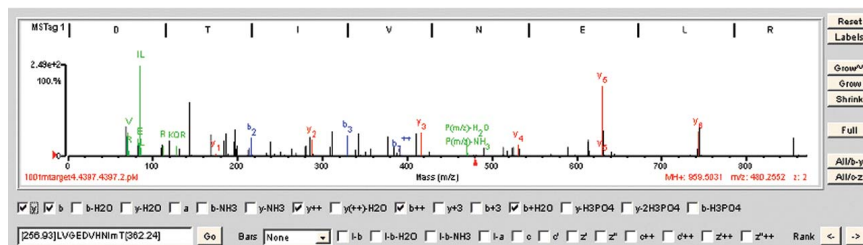


Figure 9. MS/MS spectrum for the peptide DTIVNELR in targeted MS/MS analysis.

In the biomarker discovery process, it is necessary to be able to identify small changes in the concentration of a few proteins in the presence of a large number of proteins in the sample. In this study, digest of a single plant protein spiked at a low concentration of 10 femtomole in 200 ng of human plasma digest could be identified, which demonstrates the outstanding sensitivity achieved

using Agilent's 6520 Accurate-Mass Q-TOF LC/MS configured with the HPLC-Chip. Concentrations of the spiked peroxidase peptides are comparable to the natural abundance levels of tissue leakage proteins as shown in **Figure 10**. In the figure, the normal range of abundance of classical plasma proteins is marked in black, tissue leakage proteins are blue and interleukins

etc. are in purple boxes. 14 of the high abundant classical plasma proteins (removed using Agilent's MARS column in this study) are enclosed in a green box. The concentration of a spiked peroxidase peptide level is marked with a red dotted line showing that it is in the tissue leakage region where biomarkers can be expected to be present.

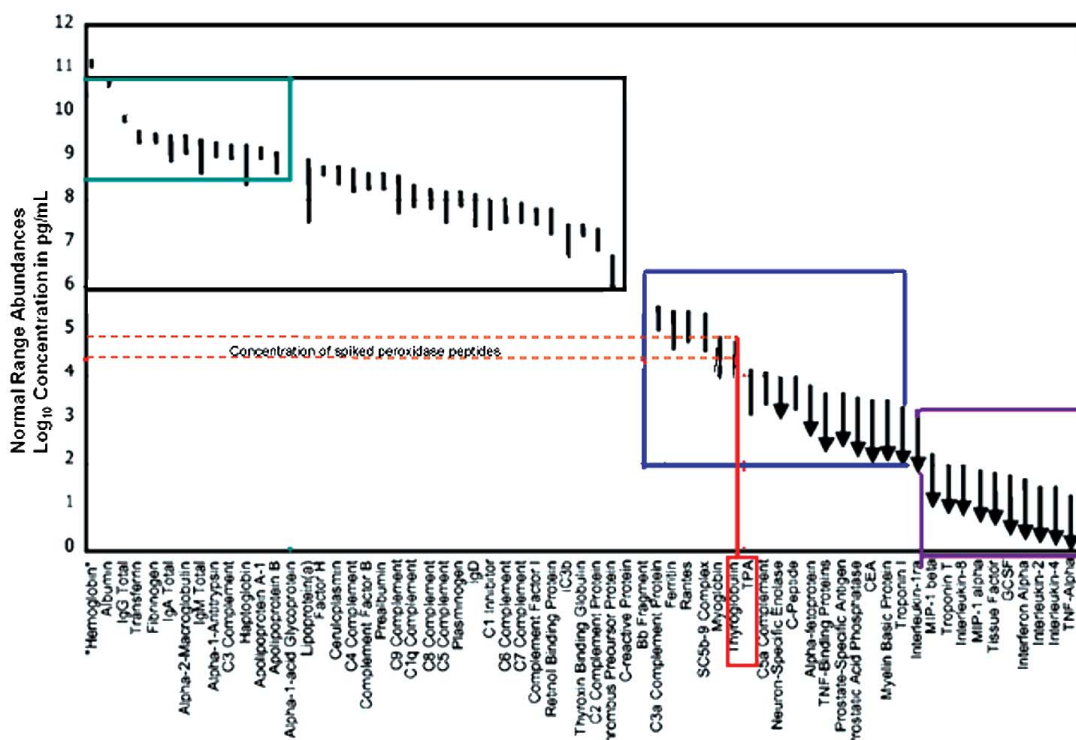


Figure 10. Plasma proteome map showing the abundance range of some known proteins used in clinical diagnosis. (Figure adopted from Anderson et al.<sup>3</sup>)

## Conclusions

The profiling approach described in this model study demonstrates outstanding sensitivity, exceptional accuracy and excellent reproducibility achieved using Agilent's 6520 Accurate-Mass Q-TOF LC/MS system coupled with Agilent's 1200 HPLC-Chip/MS system.

Removal of 14 highly abundant proteins using an Agilent MARS column enables in-depth analysis of low abundant proteins.

The MFE algorithm in the MassHunter Qualitative Analysis software helps detect low abundant peaks in the mass spectrum, increasing the sensitivity of detection.

A quality control check using MPP software is useful in evaluation of the technical variation and improves the reliability of differential analysis.

Differential analysis using Mass Profiler and MPP software enables the identification of a few differentially expressed features among a large number of similarly expressed features making it ideal for the biomarker discovery process.

Targeted MS/MS analysis on the differential features using Agilent's 6520 Accurate-Mass Q-TOF LC/MS demonstrates the ability to select the less abundant ions for fragmentation in the presence of highly abundant ions, which may not be possible in the conventional data dependent proteomic approach.

Low amounts (10 femtomole) of a protein spiked in a highly complex human plasma sample can be identified using the LC/MS based label free quantitation method.

## References

1. Isaaq HJ, Conrads TP, Prieto DA, Tirumalai R, Veenstra TD, SELDI-TOF MS for diagnostic proteomics. *Anal Chem* **2003**, 75, 148–155A.
2. Agilent publication 5989-7839EN: Human Serum and Plasma Protein Depletion – Novel High-Capacity Affinity Column for the Removal of the “Top 14” Abundant Proteins.
3. Anderson NL and Anderson NG, The human plasma proteome, *Molecular & Cellular proteomics* **2002**, 845-867.

[www.agilent.com/chem/proteomics](http://www.agilent.com/chem/proteomics)

This item is intended for Research Use Only. Not for use in diagnostic procedures. Information, descriptions, and specifications in this publication are subject to change without notice.

Agilent Technologies shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this material.

© Agilent Technologies, Inc. 2009  
Published in the U.S.A. October 20, 2009  
5990-4404EN



**Agilent Technologies**