ASMS 2004

Towards Global Proteomics, All Ions all the Time: A Proof-of-Principle of the Qualitative Ion Mapping Technology

Jeffrey C. Silva[1], Keith Richardson[2], Phillip Young[2], Richard Denny[2], Kieran Neeson[2], Therese McKenna[2], Craig Dorschel[1], Guo-Zhong Li[1], Marc Gorenstein[1], Timothy Riley[1], and Scott Geromanos[1]
[1]Waters Corporation, Milford, MA USA, [2]Waters Corporation, Manchester, UK

Introduction

Traditional tandem mass spectrometry (MS/MS) is a method of mass spectrometry whereby intact peptides (parent ions) generated from an enzymatic digest of a protein sample are mass filtered and collisionally dissociated to yield product or fragment ions which are subsequently mass analyzed. The resulting product ion spectra can be used to determine the sequence of the parent peptide and, through a database search, the identity of the originating protein. This method of analysis can be automated and it is generally performed in conjunction with a real time liquid chromatography separation system. Although this methodology is quite powerful, it has been shown to have limitations when characterizing complex protein mixtures. Many of these limitations are associated with the serial nature of the MS/MS data acquisition process which limits both the qualitative and the quantitative reproducibility of the information generated.

A new method of protein characterization of proteolytically-generated peptides has been developed which is based on coupling the alternate scanning of peptides at low and elevated energy[1,2] with a robust accurate mass LC/MS data acquisition mode[3]. This mode of data acquisition is very high in duty cycle and provides exact mass MS analysis of both the precursor and product ions for every detectable peptide. Chromatographic retention time is used to associate fragments to their corresponding precursors as a unique attribute of the Waters® Protein Expression System technology. Proprietary Protein Expression System Informatics software has been developed which allows this information to be processed in a quantitative and qualitative manner. This poster will focus on the qualitative identification capabilities of this software. ***Poster TPR 357*** *(Geromanos et al., "Towards Global Proteomics by Analysis of Exact Mass Retention Time Pairs: A proof-of-principle of the Quantitative Ion Mapping Technology") addresses the quantitative capabilities of the software and* ***Poster TPY 458*** *(C. Dorschel et al., "Protocols to Assure Reproducible Quantitative and Qualitative Analysis of Tryptic Digests of Complex Protein Mixtures for Global Proteomic Experiments") highlights the reproducibility of the Protein Expression System platform.*

Sample Preparation

Non-depleted human serum (Sigma, Cat. No. R9759) was tryptically digested according to a procedure employing a proprietary non-ionic detergent and in which cysteine residues are reduced with DTT and alkylated with iodoacetimide prior to digestion (see ***Poster TPY 458****). The resulting samples contain 0.94 – 1.13 ug human serum protein per microliter.*

Data Collection

- *LC/MS System: Waters Protein Expression System comprised of the Waters CapLC® System with the Waters Micromass® Q-Tof Ultima API Mass Spectrometer equipped with a NanoLockSpray™ Source operated at 12,000 mass resolving power*
- *Column: Waters NanoEase™ Atlantis™ dC18 Column, 300 µm x 15 cm*
- *Mobile Phase: A = 1% Acetonitrile in Water, 0.1% Formic Acid, B = 80% Acetonitrile in Water, 0.1% Formic Acid*
- *Gradient: 6% to 40% B over 100 min. at 4.4 µL/min, followed by 10 min. rinse (99% B) and 20 min. re-equilibration at initial conditions*

*Injections of 10.0 µL of the human serum sample were made directly onto the column. Mass spectral data were collected in alternating 1.8 second scans, with a 0.2 second interscan delay, at collision cell energies of 10 eV for the low energy mode and an energy ramp from 28 to 35 eV for the elevated scanning mode. The NanoLockSpray source was sampled every 10 seconds to obtain a scan of the exact mass standard [[Glu1]-Fibrinopeptide). The raw data were processed using functions in ProteinLynx™ Global SERVER and Protein Expression System Informatics software (***Poster TPR 354****, Li et al., "A Novel Algorithm to Track Ions or Peptides Found by LC/MS in Multiple Injections of Multiple Complex Biological Samples"). Additional data analysis was performed using Microsoft® Excel and Spotfire® DecisionSite™.*

Data Processing

A basic premise of the Protein Expression System Informatics software is that the exact mass MS data acquired at low energy is primarily composed of peptide precursor ion information and the exact mass MS data acquired at elevated energy is composed primarily of corresponding peptide product or fragment ion information. It is also understood that fragment ions associated with a particular parent ion are produced simultaneously with the parent ions as they elute from the reverse phase column.

*Protein Expression System Informatics filters the data acquired in the alternating scan experiment in a three-step process. The first step carefully aligns chromatographic retention time to associate fragment ions to each appropriate precursor ion (***Figure 1***). The second step compares the exact mass of each measured peptide precursor ion to a database containing the accurate mass of every peptide (including one missed cleavage) associated with every known protein in the proteome under study. Typically, peptides which are within an acceptable mass measurement tolerance (within +/- 5 ppm) are identified in a subset peptide data base (***Figure 2***). The accurate mass of each y" and b fragment ion is calculated for each of the peptides in this subset database and the measured exact mass fragment ion information obtained in the first processing step is compared to this theoretical accurate mass information (***Figure 3***). A probability-based search model that takes into account the number, type, and mass accuracy of the measured fragment ions is used to determine the best peptide sequence match.*

Figure 1.

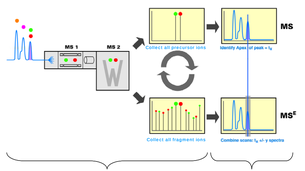
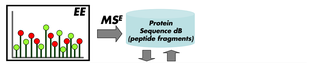


Figure 2.

Candidate Peptide Mass	Candidate Peptide Sequence b1 → bn / y"n ← y"1	Peptide Score	Protein Assignment
		+	Accession #1
		+	Accession #2
		+	Accession #3
		+	Accession #4
		+	Accession #5
		+	Accession #6
		+	Accession #7

Figure 3.



Candidate Peptide Mass	Candidate Peptide Sequence b1 → bn / y"n ← y"1	Peptide Score	Protein Assignment
		+	Accession #1
		++++++	Accession #2
		++	Accession #3
		++	Accession #4
		++	Accession #5
		++++++	Accession #6
		+	Accession #7

Results and Discussion

Figures 4 and 5 *illustrate typical results from an alternating low and elevated collision energy LC/MS analysis of non-depleted human serum tryptic digest. The upper panel of* ***Figure 4*** *illustrates the low energy base peak intensity chromatogram and the lower panel shows a low energy exact mass spectrum extracted at the highlighted retention time of 72.390 minutes.* ***Figure 5*** *illustrates the same information for the elevated energy function with the example spectrum extracted one scan later at 72.425 minutes. While many peptide ions co-elute during this moment in time, only a subset of the peptide ions apex at the same moment in time.* ***Figure 6*** *shows the low energy exact mass chromatograms for singly and doubly charge state ions for 4 example peptides that co-elute in the chromatographic region around 72 minutes. The singly and doubly charged ions associated with the same peptide clearly apex at the same chromatographic retention time.*

Figure 7 *examines the high energy data in the same retention time region. The center panel shows the exact mass chromatograms of a number of high energy ions some of which are clearly unrelated to each other. The left panel shows the low energy exact mass chromatogram for m/z 872.54 precursor at 71.96 minutes (lower trace) and six example high energy fragment ions that appear to be related since they chromatographically apex at the time. Similarly, the right hand panel shows six example high energy fragment ions that appear to be related to the low energy m/z 739.96 peptide at 72.60 minutes.*

Figure 4.

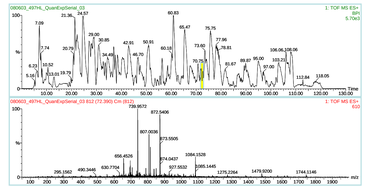


Figure 5.

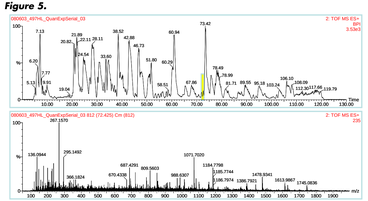


Figure 6.

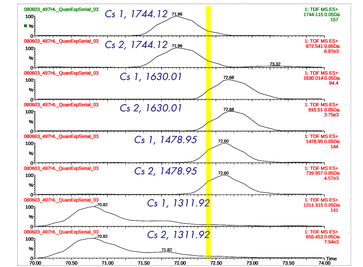
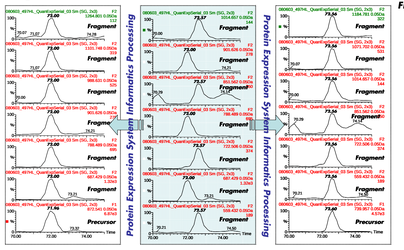


Figure 7.



As noted in the data processing section, Protein Expression System Informatics uses this principle of chromatographic apex alignment as a first filter to remove ions from the elevated energy spectra that may have a retention time similar to, but not coincident with a parent ion of interest (see ***Poster WPJ 161****, Gorenstein et al. "Statistical Study of Ions and Peptides Found In LC/MS/MS Analysis of Human Serum Digest" and* ***Poster TPR 354****, Li et al.).* ***Figure 8*** *illustrates this "cleaning" principle more graphically. The top panel shows the combined elevated energy spectra of all ions that elute at 72.425 minutes. The middle panel shows the elevated energy spectra after being cleaned of ions that do not coapex with the low energy doubly charged peptide, m/z 872.54, at 71.96 minutes. Similarly, the bottom panel shows the elevated energy spectra after removal of ions that do not co-apex with the low energy doubly charged peptide, m/z 739.96, at 72.60 minutes.*

In ***Table 1****, the qualitative identification data filtering steps illustrated in* ***Figures 2 and 3*** *have been applied to the 739.96 doubly charged peptide ion eluting at 72.60 minutes along with its chromatographically associated fragment ions. Each of the proteins listed in* ***Table 1*** *contains a tryptic peptide sequence that has an accurate mass within plus or minus 5ppm of the 739.96 measured mass. In* ***Tables 1 and 2****, columns 5 and 6 indicate those corresponding b- and y"-ions for each peptide, that are found in the elevated energy data to within plus or minus 10 ppm mass accuracy and plus or minus 0.05 minutes retention time. A "0" at a given amino acid position indicates the corresponding fragment ion was not found within the specified search tolerances, and a "1" at a given amino acid position indicates the corresponding fragment ion was found within the specified search tolerances. The measured exact mass fragment ion data associated with the 739.96 peptide parent ion has been compared to the theoretical y" and b fragment ions associated with each of the candidate parent ions in* ***Table 1****. The measured fragment ions associated with the MYLGYEYVTAIR clearly provide convincing evidence, with twelve matching fragment ions, to indicate that the 739.96 doubly charged peptide originates from the* ***Serotransferrin precursor*** *protein. Table 2, illustrates a similar analysis of the 872.54 doubly charged parent ion and its chromatographically associated fragment ions. In this example, there are 21 fragment ions that validate YAASSYLSLTPEQWK as the tryptic peptide to* ***Ig lambda chain C****.*

Table 3 *illustrates the top 82 proteins identified in a tryptic digest of undepleted human serum using the Protein Expression System profiling strategy. Protein sequence coverage is provided for each of the identified proteins.*

Figure 8.

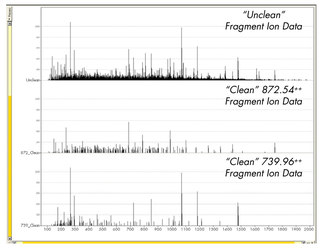


Table 1.

Description	Sequence	Actual MH+	Delta (ppm)	b-ions (b1.....bn-1)	y"-ions (y"1.....y"n-1)
Q96GA4 Similar to pre-pro-megakaryocyte potentiating factor	MEVSGDPGIPALHR	1478.7420	2.6	0000000000000	1000000000001
Q9H8J4 CDNA FLJ13564 fis clone PLACE1008201	AGPNSYAELITHR	1478.7387	0.3	0000000000000	1000000000000
Q9NVE3 CDNA FLJ10787 fis clone NT2RP4000481	ESDDGVWYYAPTK	1478.7373	-0.6	0000000000000	0000000000000
Q9BXL0* Ubiquitin carboxyl-terminal hydrolase 26	MTSGNISVSWPATK	1478.7308	-5.0	0000000000000	0000000000000
P02787 Serotransferrin precursor (Siderophilin)	***MYLGYEYVTAIR***	***1478.7348***	***-2.3***	***0111000000***	***1011111110111***
O60610 Diaphanous protein homolog 1	NSETPFTLEEAK	1478.7373	-0.6	0000000000000	0001000000000
O60522 Colon cancer antigen NYCO-45 (Fragment)	SALPYEINDSEIK	1478.7373	-0.6	0000000000000	0000000000000
P30449 Glycine amidinotransferase, mitochondrial precursor	WLSMKNTMLDEK	1478.7382	0.0	0000000000000	0000000000000
Q9HB07 MYG1 protein	AMEKVGEEFLQR	1478.7308	-5.0	0000000000000	1000000000000
Q9H099 Hypothetical 78.5 kDa protein	DLWHAAFFLSGSK	1478.7427	3.0	0000000000000	0000000000000
Q9UNI1 Elastase 1 precursor	NSWPSQASLQYR	1478.7387	0.3	0001000000000	1000000000000
P21266 Glutathione S transferase Mu 3	VDIENQVMDTR	1478.7308	-5.0	0000000000000	1000000000000
P55884 Eukaryotic translation initiation factor 3 subunit 9 (eIF3 etc)	YVYISVSWWSHK	1478.7427	3.0	0110000000	0000000000000

Table 2.

Description	Sequence	Actual MH+	Delta (ppm)	b-ions (b1.....bn-1)	y"-ions (y"1.....y"n-1)
O14773 Guanine nucleotide-binding protein beta subunit-5	MATEGHENETASIK	1743.8582	-2.0	0110000000000	1000000000100
Q9P2G7 KIAA1380 protein	DSSEIPGASVWHYAGK	1743.8701	4.8	0100000000000	1000000000000
P13796 Lipletin (Lymphocyte cytosolic protein 1)	GDEIEGIVAVVEDMSGLE	1743.8582	-2.0	0010000000000	0000000010000
Q96GG7 Interferon regulatory factor 1	LLEQSEWQPTWVDGK	1743.8548	-4.0	0000000000000	1010010000000
Q9H473 Golgi-spec. brefeldin A-resistance guanine nuf-1	NEEWMPEEQTGLVR	1743.8582	-2.0	0000000000000	0000000000000
Q9H0K9 Hypothetical 78.6 kDa protein	SHPLGLTDGGGFSEIK	1743.8548	-4.0	0010000000000	1000000000000
O75018 leucocyte Immunoglobulin-like Receptor 6A/B	TASHPGDTVTVENLIR	1743.8660	2.5	0000000100000	0000000000000
P01842 Ig lambda chain C regions	***YAASSYLSLTPEQWK***	***1743.8588***	***-1.6***	***0111111111100000***	***1111111111010***
Q9GZU1 CDNA FLJ22449 fis clone HRC09609	GFLLQNFEVGFMVWE	1743.8676	3.4	0000000000000	0000000000000
O43736 Integral membrane protein 2A (E25 protein)	GGEPNFLPVTEADIR	1743.8548	-4.0	0000001000000	0000000000000
Q9Y6I7 Fibroblast growth factor 2-interacting factor	NYGLEADATGQVSQHK	1743.8660	2.5	0000000000000	1100000000000
Q96IS3 KIAA0851 protein	THGGMDDWNSMR	1743.8669	3.0	0000000100000	0010001000000
O94808 Glucosamine-fructose-6P. aminotransferase 2	WATHGVPSAVNSHPQR	1743.8674	3.3	0000010000000	0100011000000

Table 3.

% Coverage	Description	Median Mass Error	% Coverage	Description	Median Mass Error
43	Fibrin alpha A pl.progenitor precursor	1.1	58	Hemoglobin alpha chain	0.3
12	Vitamin D binding protein precursor	2.4	45	Heat shock 70 kDa protein 1	2.3
14	Very low density lipoprotein receptor	2.8	55	Haptoglobin-1	2.9
24	Serum amyloid A protein precursor	2.6	55	Haptoglobin-2	2.8
58	Apolipoprotein a1/a4/e precursor	2.9	58	Haptoglobin related protein	2.9
45	Apolipoprotein a1/a4/e precursor	4.5	48	Factor VII	4.4
45	Complement factor B precursor validated	4.5	48	Complement factor B/EBP	4.4
45	Serum amyloid A protein 4	4.5	52	Complement factor H related protein 1	4.7
72	Serum albumin precursor	3.7	52	Complement factor H precursor H factor 1	4.7
72	Serum albumin precursor	3.7	52	Complement component C3 precursor	4.7
19	Plasma kallikrein precursor	4.8	38	Complement component C3 beta chain precursor	4.9
19	Plasma factor X precursor	4.8	38	Complement component C3 precursor	4.9
23	Serum amyloid p-component	3.7	52	Complement component C3 precursor	4.9
23	Phosphatidylcholine-sterol acyltransferase	3.7	52	Complement component C3 precursor	5.0
23	Phosphatidylcholine-sterol acyltransferase	3.7	52	Complement component C3 precursor	5.0
17	Leukocyte elastase	4.4	16	Collagen alpha 3 (V) chain precursor	4.3
14	Apolipoprotein b precursor	4.4	16	Collagen alpha 1 (V) chain precursor	4.3
24	Kininogen precursor	4.4	16	Coagulation factor IXa	4.7
19	Interleukin 2 receptor gamma precursor	4.4	16	Coagulation factor Xa	4.7
26	Interleukin 18 binding protein precursor	4.4	16	Coagulation factor VIIa	4.7
26	Interleukin 12b precursor	4.4	63	Chain A Human Serum Transferrin N Lobe	2.0
26	Hepatocyte growth activator heavy chain H5	4.4	63	Transferrin	2.0
26	Urinary tbc growth factor 5 precursor	4.4	63	Collagen X precursor	2.1
26	Insulin like growth factor IA/B precursor	4.4	63	Collagen IX precursor	2.1
26	Ig Mu chain receptor b-lee affinity	4.4	63	Collagen alpha 2 (IX) precursor	2.1
26	Ig Mu heavy chain	4.4	63	Apolipoprotein a1 precursor	2.1
68	Ig lambda	5.1	64	Apolipoprotein A-II	2.2
68	Ig lambda	5.1	64	Apolipoprotein A-II	2.2
52	Ig heavy chain V III	4.4	52	Antithrombin III	2.3
52	Ig heavy chain V III precursor	4.4	52	Antithrombin III precursor	2.3
52	Ig heavy chain V	4.4	52	Antithrombin	2.3
44	Ig gamma	4.4	63	Alpha-1-macroglobulin	2.3
44	Ig gamma	4.4	63	Alpha-1-antiproteinase	2.4
44	Ig gamma	4.4	63	Alpha-1-antitrypsin	2.4
63	Hemopexin	2.4	63	Alpha-1-microglobulin/ inter alpha trypsin inhibitor	2.3
63	Hemoglobin beta chain	2.4	63	Alpha-1-B glycoprotein	2.4
63	Hemoglobin gamma G chain	2.4	56	Alpha-1-antitrypsin precursor	2.4
63	Hemoglobin gamma A chain	2.4	56	Alpha-1-antitrypsin variant precursor	2.4
63	Hemoglobin delta chain	2.4	56	Alpha-1-acid glycoprotein 2 precursor	2.4
63	Hemoglobin beta chain	2.4	38	Alpha-1 type IV collagen	6.2

Summary

A novel alternating low and elevated energy exact mass LC/MS data acquisition strategy is proposed as an alternative to conventional data dependent LC/MS/MS analysis for protein identification. This strategy provides an enhanced duty cycle relative to data dependent MS/MS analysis and essentially yields MS and MS/MS like fragmentation information on all detectable peptide in a chromatographic analysis of a protein digest. Preliminary studies suggest that this new Waters Protein Expression System profiling strategy identifies more proteins with better sequence coverage than data dependent LC/MS/MS analysis.

1 Bateman, Robert Harold, et al. 2001. Methods and apparatus for mass spectrometry. United States patent application 20010052569 A1 December 20, 2001.
2 S Purvine, JT Eppel, EC Yi and DR Goodlett (2003) Proteomics 3: 847-850.
3 S Geromanos, A Dongre, G Opiteck, JC Silva (2003) Micromass Limited, PCT filing WO/054549.

©2004 Waters Corporation
720000896EN May 2004