ACCURATE MASS LC/MS PEPTIDE MAPS: DATA PROCESSING FOR TRACE COMPONENT IDENTIFICATION

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

BETH L. GILLECE-CASTRO[1]; Marc V. Gorenstein [1]; Daniel Golick[1]; Keith Richardson [2]; Barry Dyson [2]; Scott Berger [1]; Jeff Mazzeo [1]; Thomas E. Wheat [1]; Diane Diehl[1]
Waters Corporation; [1]Milford MA, USA and [2]Manchester UK

OVERVIEW

- Software tools were developed for application to protein characterization.
- Two novel algorithms were developed to detect and deconvolute peptide signals in LC/MS maps.
- A matching algorithm relates the observed peptides to the protein structure, including modifications.

INTRODUCTION

The primary structure of a protein can be characterized by peptide mapping, and the same analytical technique can be used to identify modifications to the protein structure. Highly resolving chromatography, accurate mass LC-MS, and software tools have been combined to more efficiently correlate peptide maps with protein structure. Even optimized chromatography still requires confirmation of peak identity and purity so it is useful to couple the separation to the exact mass measurements possible with an oa-ToF mass spectrometer. Peptides can be identified based on molecular weight, and co-elutions can be detected. This additional information links the chromatographic pattern to the structure of the protein. Additionally, the amount of trace degradation or contamination can be assessed. In particular batches can be compared for the amounts of degradation.

Complete interpretation of complex LC/MS chromatograms with accurate mass measurement is time-consuming and labor intensive. New specialized software has been developed for these large data sets. The peaks are detected by the Apex3D algorithm to deconvolute multiply-charged ions and combine isotopes. This processed data is matched to the structural features of the proteins with rigorous comparison and search algorithms. The combination of UPLC, oa-Tof MS, and advanced software act synergistically to improve the interpretation of peptide maps.

METHODS

Samples
MassPREP™ Phosphorylase b Digestion Standard
MassPREP™ Hemoglobin Digestion Standard
MassPREP™ Peptide Standards
Add 250 µL water to each vial. Vortex. Final concentration of digest is 4pmol/µL
Oxidized digest
Add 250µL 0.01% hydrogen peroxide to 1 vial Phosphorylase b MassPREP™ Digestion Standard. Vortex. Incubate at room temperature for 2 hours.

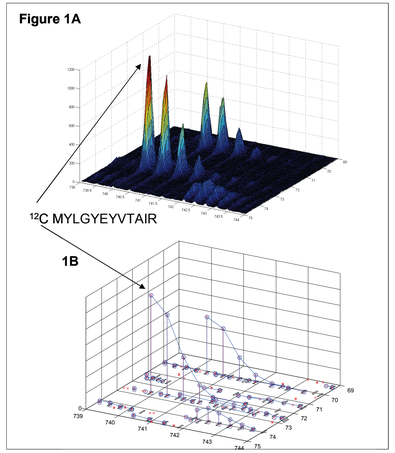
Instruments
Waters® ACQUITY UPLC™System including
ACQUITY UPLC™BEH300 C_{18} 1.7µm 2.1x100mm Column
Waters Micromass LCT Premier Mass Spectrometer or
Waters Micromass Q-tof Premier Mass Spectrometer

Software
Masslynx 4.1
Biopharmalynx 1.1, Beta

METHODS

Peak Detection
The first data-analysis algorithm detects the ions obtained in an LC/MS separation. A convolution-based technique measures three key properties of each ion: retention time, mass-to-charge ratio, and intensity.
The second algorithm simplifies spectra by selecting only those ions whose retention times fall within restricted ranges. As an example, consider a peptide that elutes at retention time *tr*. All its ions must also elute at *tr*. Variations from *tr* are due only to measurement error.
By selecting ions that have the same retention time (to within measurement error), the algorithm simplifies spectra. Such simplified spectra can more clearly reveal the unique, multi-ion signature of peptides.

Figure 1A



Ion Detection: Figure 1A shows a 3-dimensional representation of LC/MS data obtained by assembling spectra into a matrix form. The vertical axis is counts; the x– and y-axes are time and m/z.
This matrix of intensities is convolved with a proprietary, 2-dimensional filter. The filter coefficients are chosen so that the apex location of the convolved data optimally estimates the retention time and mass-to-charge ratio of the respective ion. (Gorenstein, Plumb Stumpf, patent pending.) At the apex, the filter output gives the ion's response in area-counts.
Thus the apex location of the filtered data determines the three key ion parameters: retention time, m/z, and intensity.
***Figure 1B shows the results obtained from the detection algorithm. Each ion is represented as a "stick" located at the apex of the convolved data. The** (x,y) location of the stick gives the ion's retention time and m/z, and the height of the stick is the ion's intensity. The second algorithm then combines all sticks related by retention time, isotopic m/z difference, and charge state. The resulting table and chromatographic representations of the data have one stick for each peptide with the combined intensity from all isotopes and charge states.*

RESULTS

Figure 2 — Analysis of LC/MS Peptide Map

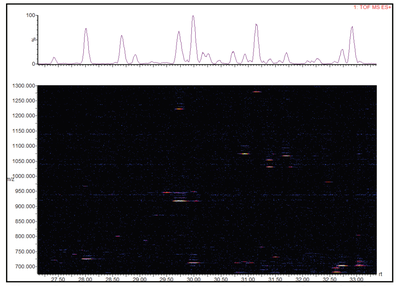


Figure 2A: Contour Map showing the time-m/z-intensity data matrix for a tryptic digest of bovine hemoglobin between 27 and 33min. Interpretation of this data set and its relation to

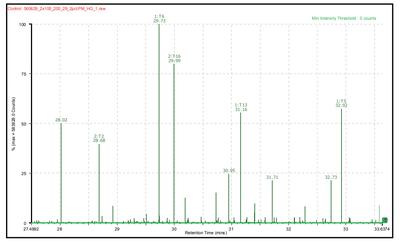


Figure 2B: Processed LC/MS Peptide Map from Figure 2A. Each peptide is represented as a "stick" located at the apex retention time. The intensity of the stick is the sum of all of the m/z ions for isotopes and charge states for each component peptide.

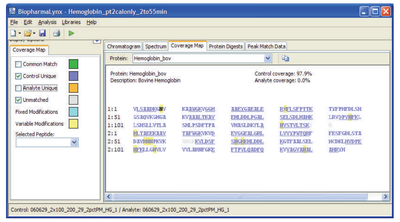


Figure 2C: Coverage Map from Figure 2B. Each peptide is matched to the known sequence using exact mass measurement and a hierarchal search.

Figure 3 — Comparing LC/MS Peptide Maps

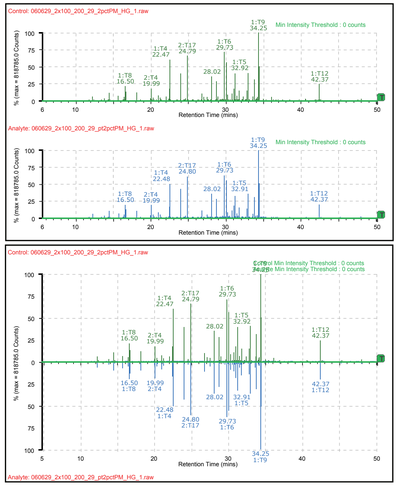


Figure 3A and 3B: Processed LC/MS Peptide Maps of Spiked Digests. The two digests have been spiked with different amounts of peptide standard to create a simulation of two batches of protein with differences in structure. The stacked and mirror image plots are time-aligned to provide alternative views for manual inspection.

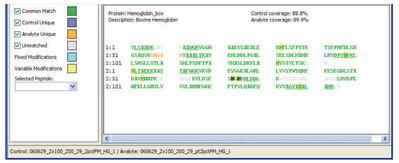


Figure 3C: Comparative Coverage Maps LC/MS Peptide Maps of Spiked Digests. The visual comparisons in Figures 3A and 3B are more easily interpreted when each map is subjected to the sequence matching algorithm. Color coding is used to identify peptides that are common to both digests as well as those that are only observed in the control digest or the analyte digest.

Figure 4 — Fragment and Dimer Identification

Source conditions can be altered to enhance fragmentation of peptides. In this example a low level of fragmentation is present to confirm the sequence as shown. Electrospray source conditions may also produce dimer ions when concentrations of peptides are high.

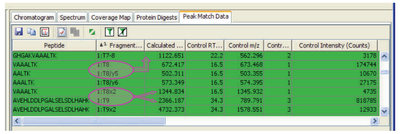


Figure 4: Two fragments of T8, y5 and y6, confirm the identification suggested by the accurate mass measurement. Low to moderate levels of dimer formation in the ion source are observed for peptides T8 and T9. Both fragments and dimers are automatically interpreted by the software.

Figure 5 — Trace Contaminant Detection

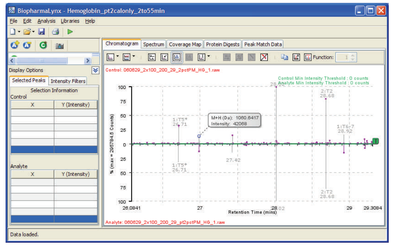


Figure 5A: Comparative LC/MS Peptide Maps of Spiked Digests. The specific differences between the digests as shown in the coverage maps can be related to the chromatogram by extracting a difference plot. Spiked peptides are highlighted in purple in the difference plot. The automatically labeled peaks are higher in the control than the analyte as expected since the control was spiked at 2% and the analyte at 0.2%.

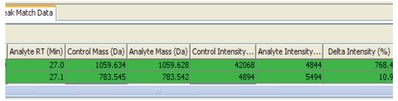


Figure 5B: Quantitative Comparisons of LC/MS Peptide Maps of Spiked Digests. The differences between the digests as shown in Figure 5A are clearly identified by the exact mass measurement. The relative intensities for the two different spike levels are in reasonable agreement with the expected values.

Figure 6 — Modification and Disulfide Identification

Forced degradation, such as oxidation, produce highly modified peptides. Multiple oxidations combined with deamidation were observed in this tryptic digest of rabbit phosphorylase b.

Disulfide bonds are identified from non-reduced digest samples and can be compared to a reduced sample.

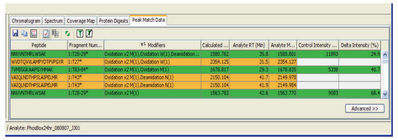


Figure 6A: The oxidized sample, "analyte", shows high intensity pepetides with multiple oxidations on both Met and Trp residues. Lines which are green show that the control sample was partially oxidized. Lines which are gold show that some new oxidation states are observed after treatment with peroxide.

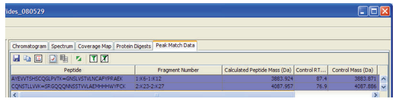


Figure 6B: Disulfide-bonded peptides are shown in this example of a Lys-C digest. Lines in the table for the control non-reduced sample are blue because the analyte recuced sample had no matching peaks.

CONCLUSIONS

- **Retention time alignment and intensity normalization allow accurate comparisons between runs and batches.**
- **The protein coverage of peptide maps can be readily assessed and compared using software tools.**
- **Less than 0.5% of trace contaminants and modifications can be detected, and the amounts compared between runs and batches.**
- **MS source conditions can be set to take advantage of in-source fragmentation for sequence confirmation.**
- **Disulfide bond linkages can be confirmed.**
- **Software tools provide efficient matching of chromatographic peaks with known structural features of the protein as well as modifications to that structure.**

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS