ASMS 2004

Statistical Study of Ions and Peptides Found In LC/MS, LC/MS^E Analysis of Human Serum Digest

Marc V. Gorenstein, Guo-Zhong Li, Scott Geromanos, Jeffrey C. Silva, Craig A. Dorschel, Robert S. Plumb, Chris L. Stumpf, and Timothy Riley
Waters Corporation, Milford, MA USA

Introduction

Data-analysis algorithms are critical to the success of proteomic investigations. This paper describes, first, two algorithms and their application to LC/MS data. It then describes the application of these algorithms to a novel, LC/MS-based, proteomics method.

The first data-analysis algorithm detects the ions obtained in an LC/MS separation. This convolution-based technique is designed to measure 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 t_r. All its ions must also elute at t_r. Variations from t_r are due only to measurement error.

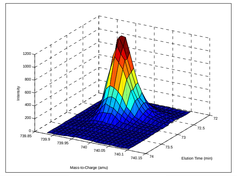
By selecting ions that have the same retention time (to within measurement error), the algorithm simplifies spectra. Such simplified spectra can more clearly reveal the unique, multi-ion signature of peptides.

Both algorithms are applied to a novel, LC/MS-based, exact-mass, proteomics method.[1,2,3] This method collects spectra that alternate uniformly in time between a (low-energy) MS-mode and an (elevated-energy) MS^E–mode. There is no MS spectral selection applied prior to MS^E fragmentation. Thus, the MS^E spectra contain fragment ions of all precursor peptides. Rather than relying on MS selection, this alternating-energy method simplifies spectra using retention-time selection.

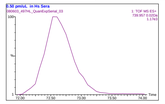
The convolution technique detects all the precursor ions and all the fragment ions seen in both modes. The retention time of a fragment ion must be the same as the retention time of its respective precursor. Given an MS precursor, the spectral-simplification method selects only MS^E ions that have the same retention times (to within measurement error) as that precursor peptide. The MS^E spectra are greatly simplified, and the remaining, accurately-mass-measured, fragment ions can be searched against a database.

An ion in an LC/MS analysis

Figure 1A

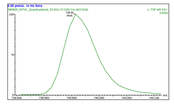


1B



Elution time (min)

1C



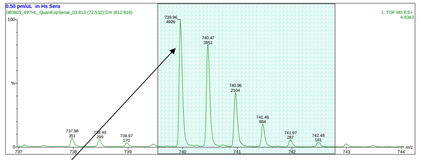
m/z (amu)

Figure 1A shows a three-dimensional plot of the raw counts for a single ion obtained in an LC/MS analysis of human serum digest. Conventionally, we visualize such a peak either by a chromatographic trace (**B**) or by a spectrum (**C**).

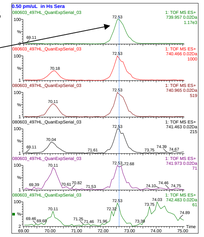
The three-dimensional plot suggests that it is the *apex* location that carries the key information associated with an ion. The precise *(x,y)* location of the apex gives the retention time and m/z of the ion. The sum of the counts in the peak gives the intensity of the ions.

Example: Detection of ions from a tryptic peptide

Figure 2A



2B



^{12}C MYLGYEYVTAIR

To illustrate the ion detection algorithm, we analyze a peptide in human serum. **Figures 2A** shows a spectrum extracted at 72.5 minutes that contains the signature of a Z=2$^+$ ion. These six ions are from a tryptic peptide of the Serotransferrin precursor protein in human serum.[2] The peptide's sequence is MYLGYEYVTAIR. At Z=2, the m/z value for the ^{12}C, monoisotopic ion is 739.96 amu.

Figure 2B shows six nominal mass chromatograms, one for each isotopic mass of this peptide. As indicated by the blue line, the retention time of each ion has the same value of 72.53 min, indicating that the apex of each ion falls within the same, single scan.

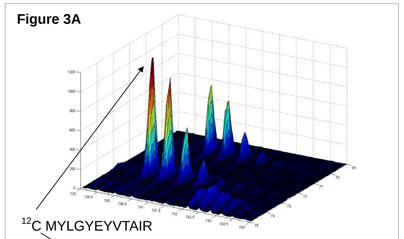
Figure 3A shows the 3-dimensional representation of this data obtained by assembling the spectra into a matrix form. The vertical axis is counts; the x– and y-axes are time and m/z. The mass and time range in **Figure 3A** correspond to the indicated mass range in **Figure 2A** and the time range in **Figure 2B**.

The matrix of intensities are convolved with a propriety, 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 ration 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 3B** 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.

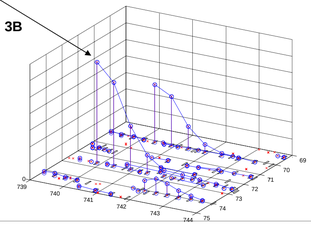
The ions that correspond to peptides are indicated by lines connected to the tops of the sticks. The ^{12}C monoisotope of MYLGYEYVTAIR is indicated.

Figure 3A



^{12}C MYLGYEYVTAIR

3B



Application to alternating LC/MS and LC/MS^E

To obtain a complete chromatographic profile, LC/MS acquisition must collect a sufficient number of spectra across a peak. In the alternating LC/MS, LC/MS^E technique[1-3], spectra are sampled at a high enough duty cycle so that both the MS and the MS^E scans sample each chromatographic peak multiple times.

Figure 6A shows a chromatographic peak sampled a total of about 10 times across the peak's full-width at half height. **Figure 6B** and **Figure 6C** show that, taken separately, the MS and the MS^E scans each adequately sample the same chromatographic peak.

In the alternating-scan technique, the MS^E-mode spectra contain fragment ions obtained from *all* precursor peptides; there is no MS spectral selection applied prior to MS^E fragmentation.

This lack of spectral selection may seem to make these MS^E spectra difficult to interpret. A conventional analysis would have to contend with a high-density of fragment ions in the MS^E spectra.

Figure 6

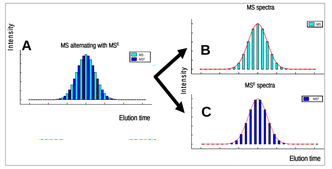


Figure 7

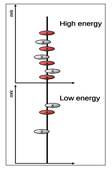
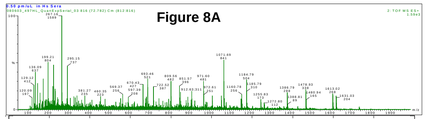


Figure 7 shows how spectral simplification can be used to select only those ions associated with a precursor, removing those that are not associated. Rather than relying on MS selection, this alternating-energy method simplifies spectra using retention time selection.

Figure 8A shows the MS^E spectrum at 72.5 min that corresponds to the MS spectrum of **Figure 5A**. **Figure 8B** contains the simplified spectrum corresponding to **Figure 5B**. Ions that fall within ±20 ppm of the Y– and B– fragment ion of MYLGYEYVTAIR are colored red. **Figures 9A,B** and **Figures 10A,B** focus in on the mass ranges that contain two fragment ions. These figures show the dramatic reduction in ions that results from retention-time based, spectral simplification.

Figure 8A



8B

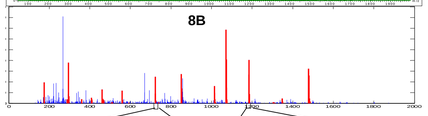
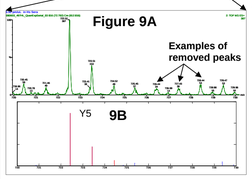
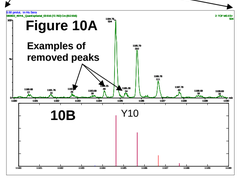


Figure 9A



9B

Figure 10A



10B

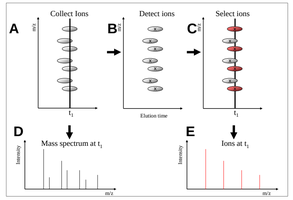
Simplification of spectra by retention time selection

Figure 4 illustrates the spectral simplification algorithm. **Figure 4A** shows a contour plot that contains seven ions that elute at a range of m/z values. Four of these ions elute at t_r . **Figure 4D** shows the mass spectrum extracted at retention time t_r. Because three other ions partly co-elute with these four, all *seven* ions appear in the extracted spectrum. Traditionally, it is from this, single, extracted spectrum that an ion's m/z and intensity are obtained.

In the proposed method, ions parameters are obtained from the data matrix. **Figure 4B** shows the apex locations of each ion with an X, illustrating the output of the ion-detection algorithm described above. The simplification algorithm selects only those ions whose retention times is at, or near, the value t_r, as show in **Figure 4C**.

Of the seven ions, only four that pass this criteria. **Figure 4E** displays these four selected ions, plotting them at the m/z values and with intensities obtained from the convolution-based detection method.

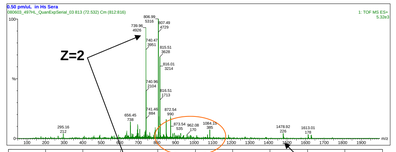
Figure 4



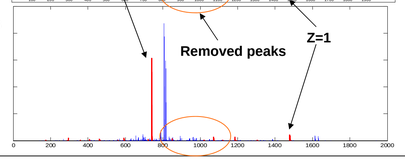
To illustrate spectral selection with real data, **Figure 5A** shows the complete, combined low-energy spectrum centered at 72.5 minutes. At baseline, the chromatographic-peak-width is 1.0 minutes. Thus spectral peaks will appear in the spectrum for *any* ion whose retention time falls within ±0.5 minutes of 72.5 min.

Figure 5B shows the simplified spectrum obtained by selecting only those ions whose retention times fall within ±0.1 minute of 72.5 min. Thus about 80% of the peaks in **(5A)** are excluded in **(5B)**. E.g., note the peaks near 1000 amu in **(5A)** that do not appear in **(5B)**. The Z=2 and Z=1 ion clusters for the MYLGYEYVTAIR appear in both plots.

Figure 5A



5B



Conclusions

- Ions are detected by their apex location
- Apex location measures ion's m/z, retention time, and intensity
- Exact mass and retention-times provide unique signature for peptides
- Retention time selection simplifies spectra
- Simplified spectra provide increased confidence in peptide identification

Referenced ASMS 2004 Posters:
[1]*MPX 452: J. Silva, et al., Towards Global Proteomics, All Ions all the Time: A Proof-of-Principle of the Qualitative Ion Mapping Technology*
[2]*TPR 354: G.Z. Li, et al., A Novel Algorithm to Track Ions or Peptides Found by LC/MS in Multiple Injections of Multiple Complex Biological Samples*
[3]*TPR 357: S. Geromanos, et al., Towards Global Proteomics by Analysis of Exact Mass Retention Time Pairs: A proof-of-principle of the Quantitative Ion Mapping Technology*

720000897EN May 2004