Enhanced and Automated Analysis for Intact Protein Data by Peak Modeling Charge Deconvolution Method for Mass Spectrometry

Xiaoling Wu[1], Linda DeNoyer[2], Ning Tang[1], Maithilee Samant[1], Keith Waddell[1], [1]Agilent Technologies, Santa Clara, CA, [2]Spectrum Square Associates, Ithaca, NY

ASMS 2012
TP 488

Agilent Technologies

Introduction

In LC/MS data analysis, the maximum entropy deconvolution method has been widely used for transforming multiple-charged spectrum into zero-charged mass spectrum. With a deconvoluted spectrum, the complexity of the data is reduced significantly; however, the maximum entropy deconvolution method can introduce unwanted artifacts through extraneous signals. We present an advanced data processing methodology that can resolve this problem by applying an automatic peak modeling technique, delivering the most probable deconvolution results that are virtually free of artifacts. This peak modeling deconvolution method results in improved signal-to-noise ratios and enhanced resolution of protein peaks for precise analysis.

Experimental

Materials

Monoclonal antibody IgG1 was diluted with 3% acetonitrile (ACN) and 0.1% formic acid (FA) in water to the following concentration: 5000ng/µL, 1000ng/µL, 500ng/µL, 100ng/µL, 10ng/µL, 5ng/µL. 1µL of sample was injected.

LC/MS analysis

All samples were analyzed using an Agilent 1290 HPLC coupled to Agilent 6550 Q-TOF mass spectrometer with Agilent Jet Stream. Agilent Poroshell 300SB-C8 Column 1x75mm was used.

Mobile phase:
solvent A: 0.1% formic acid (FA) in water;
solvent B: 0.1% FA in 90% acetone nitrile (ACN) and 9.9% water.
Flow rate: 0.15ml/min
Gradient: 0-3min: 3% B
3-8min: 90% B
8-10min: 90%B
Q-TOF parameters:
Vcap: 5000V
Drying gas flow rate: 12L/min
Drying gas temperature: 290 °C
Sheath gas flow rate: 12L/min
Sheath gas temperature: 350 °C
MS scan rate: 1 scan/second
Mass range: 700 – 7000 m/z

Methods

Two different methods of charge state deconvolution in Agilent MassHunter Workstation BioConfirm software were used for comparison.

Maximum Entropy Charge Deconvolution

The maximum entropy charge deconvolution is a well-known and powerful method for electrospray mass spectra of intact proteins deconvolution. This method transforms a m/z raw spectrum of one or more intact proteins, usually from averaging the mass spectra across the chromatographic elution time of the protein(s) of interest, into the most probable zero-charge mass spectrum (in Dalton units). Maximum entropy deconvolution works very reliably for pure protein data or relative simple protein mixture.

Peak Modeling Charge Deconvolution

This advanced algorithm starts with a maximum entropy deconvolution. Based on the maximum entropy result, it can automatically generate peak models without manual intervention and apply these models through fitting and validating procedures, providing a highly resolved zero-charge spectrum and a set of mass error assessments for each peak.

Results

We have examined protein data in the range of 12-150 kDa as a means of assessing the peak modeling charge deconvolution methodology. The benefits of this improved methodology are illustrated by the results of an IgG1 sample data.

Monoclonal antibody IgG1

A serial dilution from 5ng to 5000ng IgG1 was analyzed on 6550 Q-TOF mass spectrometer with 1.0 mm column. The sensitivity of 6550 iFunnel Q-TOF is 10x - 50x better than older technology. An example of 10 ng mAb IgG1 is shown in Figure 1. Figure 1a is the full charge envelope of IgG1 raw data. Figure 1b shows a zoomed in spectrum at Z=47, which is the highest abundance charge state in the raw data. Figure 1c shows a zoomed in spectrum at Z=39. the lower abundance charge states have much lower signal-to-noise ratios and peaks are severely overlapped,

The reconstructed zero-charge spectra from maximum entropy deconvolution and peak modeling deconvolution are shown in Figure 2.

Results and Discussion

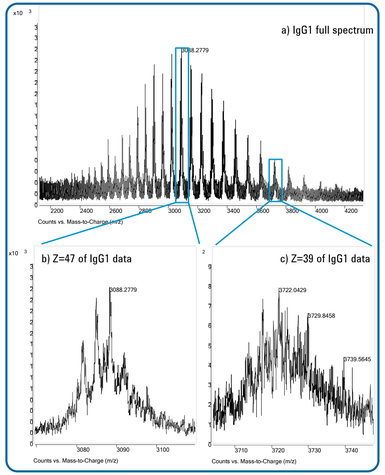
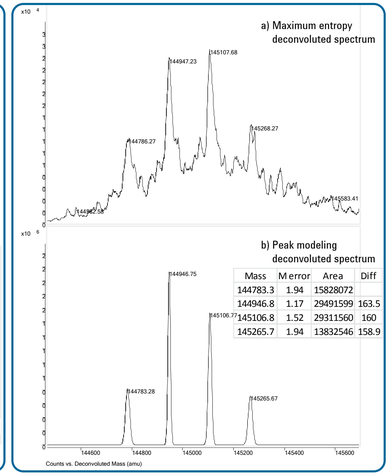


Figure 1. Raw data of 10ng IgG1 acquired on an ESI-QTOF instrument. a) Full spectrum. b) Zoomed in spectrum at Z=47. c) Zoomed in spectrum at Z=39.



Mass	M error	Area	Diff
144783.3	1.94	15828072	
144946.8	1.17	29491599	163.5
145106.8	1.52	29311560	160
145265.7	1.94	13832546	158.9

Figure 2. Deconvoluted spectra. a) Maximum entropy deconvoluted spectrum. b) Peak modeling deconvoluted spectrum.

The data is automatically baseline corrected using the same baseline algorithm. Therefore, any differences between the results can not be attributed to different baseline corrections.

Figure 2a shows the maximum entropy deconvoluted spectrum. The maximum entropy deconvolution result showed improved mass resolution compared to the measured raw data, especially for the less-resolved small features. Deficiencies in this method, however, arise from two instances: 1) harmonics and artifacts are generated in deconvoluted results, and 2) protein peaks cannot be well-separated due to the limited resolution and low signal-to-noise ratio for low amount of protein being analyzed.

The new peak modeling deconvolution is less prone to noise because it automatically generates peak models from the measured data and uses the peak models to fit the data. Any data that doesn't fit the models is rejected as noise. Therefore, the result is much cleaner than the maximum entropy deconvolution result because only the fitted peaks are reported. The new peak modeling method produces almost no artifacts due to the automatic peak modeling technique and several other enhancements, including the ability to accommodate calibration errors and physical rules so that only masses consistent with the experiment rules are reported. The mathematically plausible masses that are inconsistent with the rules are filtered out. Figure 2b shows the peak modeling deconvolution result. The new result has enhanced resolution and overlapped peaks are well-resolved. This shows the new peak modeling methodology is robust even for very poor S/N data and severely overlapped peaks.

Results and Discussion

The peak table in Figure 2b shows the peak modeling deconvoluted masses, mass error, peak area and adjacent mass differences. The expected common mass difference should be 162 Da. The mass difference values are in agreement with the expected value. The M error is the standard deviation calculated from each charge state of the protein. In figure 2b, the peak width represents mass error and the peak area represents the relative quantitation. The peak width corresponds to the confidence in the mass measurement by the peak modeling deconvolution algorithm rather than the true peak width.

Relative quantitation analysis

We collected IgG1 data over a wide range of concentrations, ranging from 5000 ng/µl to 5 ng/µl, with 5 replicates for each concentration. The raw data was deconvoluted using the peak modeling deconvolution method. The deconvoluted spectra at different concentrations are shown in Figure 3. The average peak area from 5 replicates are calculated for each peak. Figure 4 shows the calibration curves for the top 4 deconvoluted peaks at 144784, 144946, 145107 and 145269 Da. Calibration curves for the deconvoluted peak area were found to be linear in three orders of magnitude.

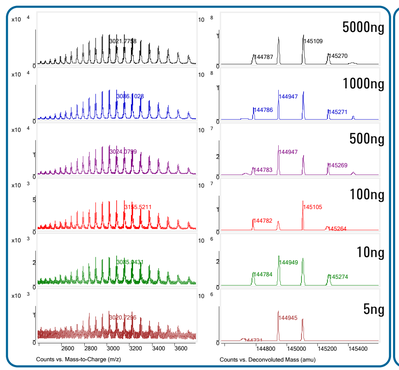


Figure 3. Raw data and peak modeling deconvoluted spectra of IgG1 at different concentrations. From top to bottom are spectra for concentrations 5000, 1000, 500, 100, 10 and 5 ng/µl.

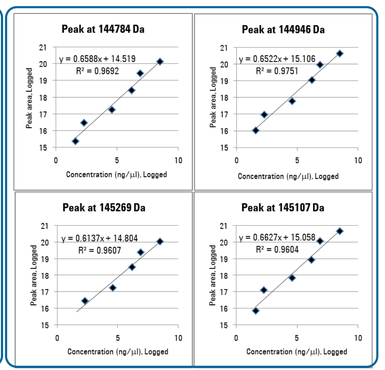


Figure 4. Calibration curves for the top 4 deconvoluted peaks. From top left, clockwise, are calibration curves for peaks at 144784, 144946, 145107 and 145269 Da.

Conclusions

- Two charge deconvolution methods are compared - Maximum entropy deconvolution and peak modeling deconvolution.
- Peak modeling deconvolution can provide a much cleaner deconvolution result that is free of artifacts. The result has improved signal-to-noise ratios and reports the standard deviation of the mass measurement.
- Peak modeling deconvolution result has an enhanced resolution and overlapped peaks are well-resolved. This has enabled the differentiation of small modifications from the main heterogeneous glycoprotein profile with much greater clarity.
- Peak modeling deconvolution result can be used for relative quantitative analysis for determining protein concentrations.