Protein Confirmation Using the LC/MSD TOF and the Agilent TOF Protein Confirmation Software

Application Note

Donghui Yi, Irene Lee, and John Fjeldsted
Agilent Technologies

Introduction

The combination of liquid chromatography, electrospray ionization, and time-of-flight mass spectrometry (LC-ESI-TOF MS) is well-established for low-ppm mass analysis of small molecules and peptides. While there is a need for similar mass accuracy for analysis of proteins, this has been more difficult to achieve. ESI analyses of proteins result in multiply-charged mass peaks that are deconvoluted to produce an average protein molecular weight. During deconvolution, any mass measurement errors for the multiply-charged mass peaks are multiplied by the mass peak charges. Calculation of an average protein molecular weight is further complicated by the presence of unresolved post-translational modifications and microheterogeneities.

This work describes use of the Agilent TOF Protein Confirmation software, which features an innovative deconvolution algorithm that gives protein molecular weights with very high accuracy. The software includes both interactive and automated operating modes. It can both confirm the presence of target proteins and supply molecular weights of unexpected proteins detected in the sample.

Agilent Technologies

Experimental

Protein standards were purchased from Sigma. Aqueous stock solutions were prepared at 1 nmol/mL with 0.01% trifluoroacetic acid and were stored at –20°C. All analyses were accomplished using an Agilent LC/MSD TOF (time-of-flight) mass spectrometer. The coupling between the LC and the TOF was via an electrospray ionization (ESI) source with dual nebulizers—one nebulizer for the LC eluent and one nebulizer for the internal reference mass compounds. All data was reference mass corrected prior to being written to disk.

The LC/MSD TOF data was processed with the Agilent TOF Protein Confirmation software, which provides three major capabilities:

- Calculation of molecular weights of target proteins via a protein sequence editor
- Interactive deconvolution and protein confirmation
- Automated deconvolution and protein confirmation

LC Conditions

Instrument:	Agilent 1100 Series LC
Column:	ZORBAX Poroshell SB300-C18, 1 x 75 mm, 5 µm particle size
Mobile phase A:	0.1% formic acid in water
Mobile phase B:	0.1% formic acid in acetonitrile
Gradient:	
20% B at 0 min	
100% B at 5.5 min	
Flow rate:	0.55 mL/min

MS Conditions

Instrument:	Agilent LC/MSD TOF with ESI source
Ionization mode:	Positive ion
Capillary voltage:	4000 V
Nebulizer:	45 psig
Drying gas:	12 L/min at 350°C
Scan range (u):	300–2000
Fragmentor:	225 V
Skimmer:	60 V
Octopole RF:	250 V
PMT:	700 V

Results and Discussion

The following describes use of and results from each of the three major components of the TOF Protein Confirmation software.

Calculation of molecular weights

The protein sequence editor was used to calculate the monoisotopic and average molecular weights of the proteins that were analyzed. This allowed comparison of the theoretical molecular weights with experimental results. This editor is shown in Figure 1. After the sequence was entered or copied in (e.g., from a public protein database) and modifications and linkages were added, then the calculator button was clicked to compute the molecular weight. The ability to a copy in the sequences and to customize modifications and linkages added to the convenience of the software.

Interactive protein confirmation

The interactive protein confirmation mode is very useful when there are only a few samples to process, or when the sample has co-eluting peaks that necessitate manual spectral selection to achieve optimum deconvolution results. It is also useful for iterative adjustment of deconvolution settings.

Data from the protein samples was processed using interactive protein confirmation. All molecular weights were calculated with low-ppm errors. For the cytochrome c example shown in Figure 2, the theoretical average mass was 12229.86, versus 12229.98 determined experimentally. This analysis produced an error of only 0.12 Da, or 10 ppm.

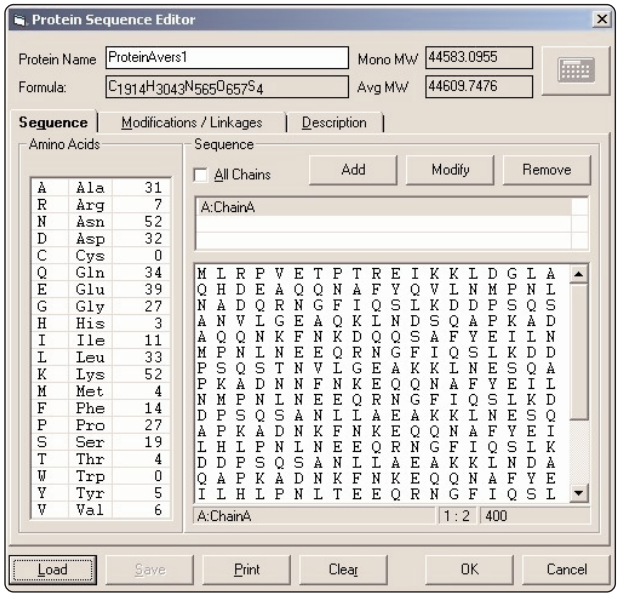


Figure 1. Protein sequence editor for calculation of protein molecular weights

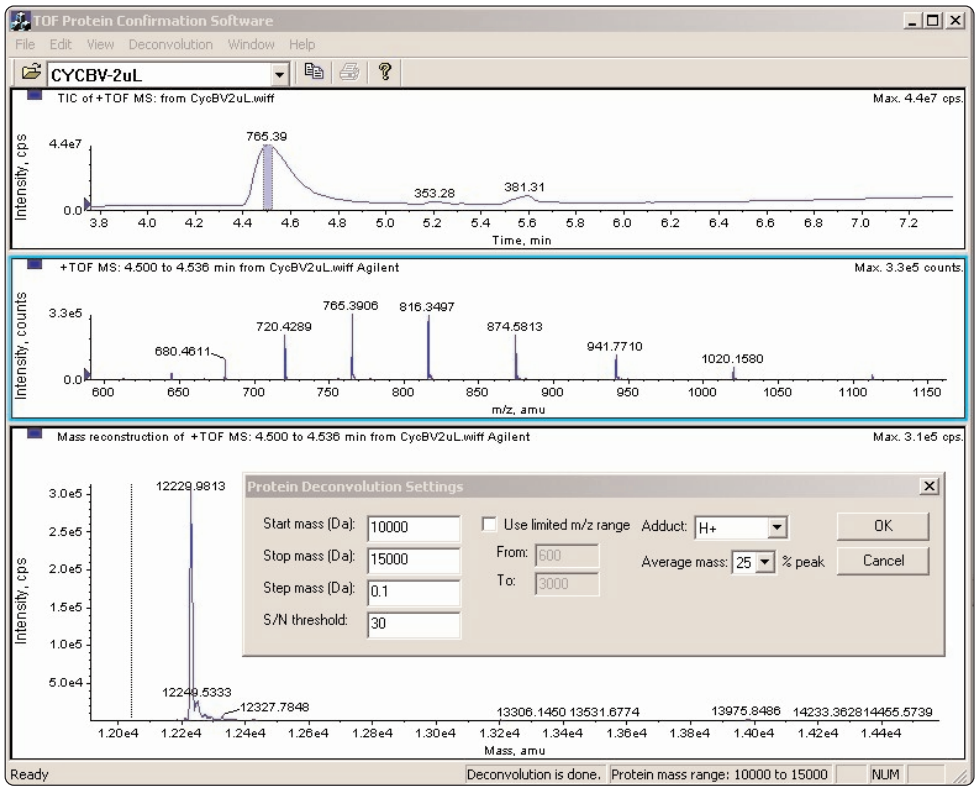


Figure 2. Interactive spectral deconvolution and protein confirmation

Like the automated mode described next, the interactive mode provided the following results:

- Molecular weight of primary/target component(s)
- List of molecular weights of other proteins detected
- Sets of multiply-charged ions associated with each protein molecular weight

Automated protein confirmation

The automated protein confirmation mode is very useful for high-throughput analyses. It can process data automatically as it is acquired, or process data that was acquire previously. It can apply a single deconvolution method to all samples, or apply individual deconvolution methods to each sample. The automated protein confirmation mode will also automatically generate reports.

The automated protein confirmation mode was used to process data from the protein standards. The first step was to set up a protein analysis method, as shown in Figure 3. Method parameters determined how much of the chromatographic peak was averaged to generate the multiply-charged spectrum, and they established the deconvolution and report options.

The second step was to set up a processing sequence, or worklist, as shown in Figure 4. The worklist included data files, corresponding protein analysis methods, and the expected protein molecular weights. The molecular weights were typed in, or were calculated from single-letter amino acid sequences that were copied in or transferred directly from the protein sequence editor. The software calculated the difference between the theoretical molecular weights and the experimental deconvoluted molecular weights,

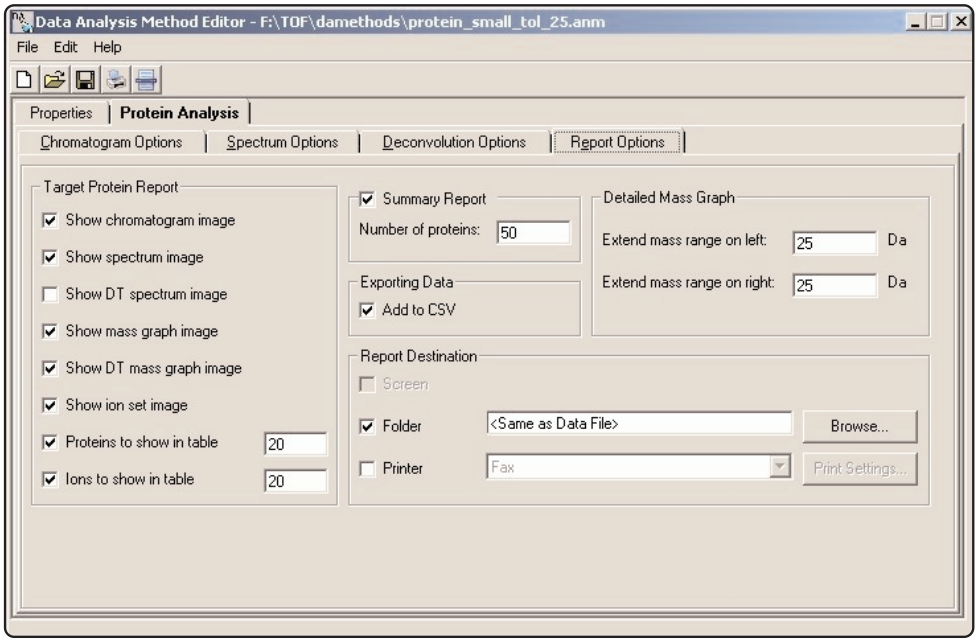


Figure 3. Setting up a processing method for automated protein confirmation

		Sample Name	DA Method	Data File	Expected MW
1	✓	InsulinRC	protein_small_tol.anm	Insulin17.2pm.wiff	5807.631
2	✓	cytochromec 1pmc	protein_small_tol.anm	cytochromec (recalibrated).wiff	12229.92
3	✓	Streptavidin	protein_big_tol.anm	STRE1.66pm.wiff	13271.04
4	✓	MGBRC	protein_small_tol.anm	MGB1.41pm.wiff	17329.60
5	✓	B-GALA	protein_small_tol_25.anm	BGALA1.7pm.wiff	116352.0

Figure 4. Setting up a processing sequence for automated protein confirmation

and established which detected proteins were target proteins and which were "other" proteins (i.e., unidentified modifications, sample impurities).

As with the interactive mode, the automated protein confirmation provided extremely accurate results. For example, the LC/MSD TOF analysis of β-lactoglobulin is shown in Figures 5 and 6. Figure 6 shows a portion of the results of the automated processing for this sample. The TOF Protein Confirmation software revealed the two expected proteins (the A and B variants), both measured to within 0.08 Da (4 ppm) error.

A number of additional proteins were processed using the automated software. To avoid potential interferences that often contribute to the sides of the mass peaks, the deconvolution algorithm centroided only the mass peak apexes. Table 1 lists the results. For these test proteins, the TOF Protein Confirmation software delivered deconvoluted average masses that were accurate to between 1 and 20 ppm.

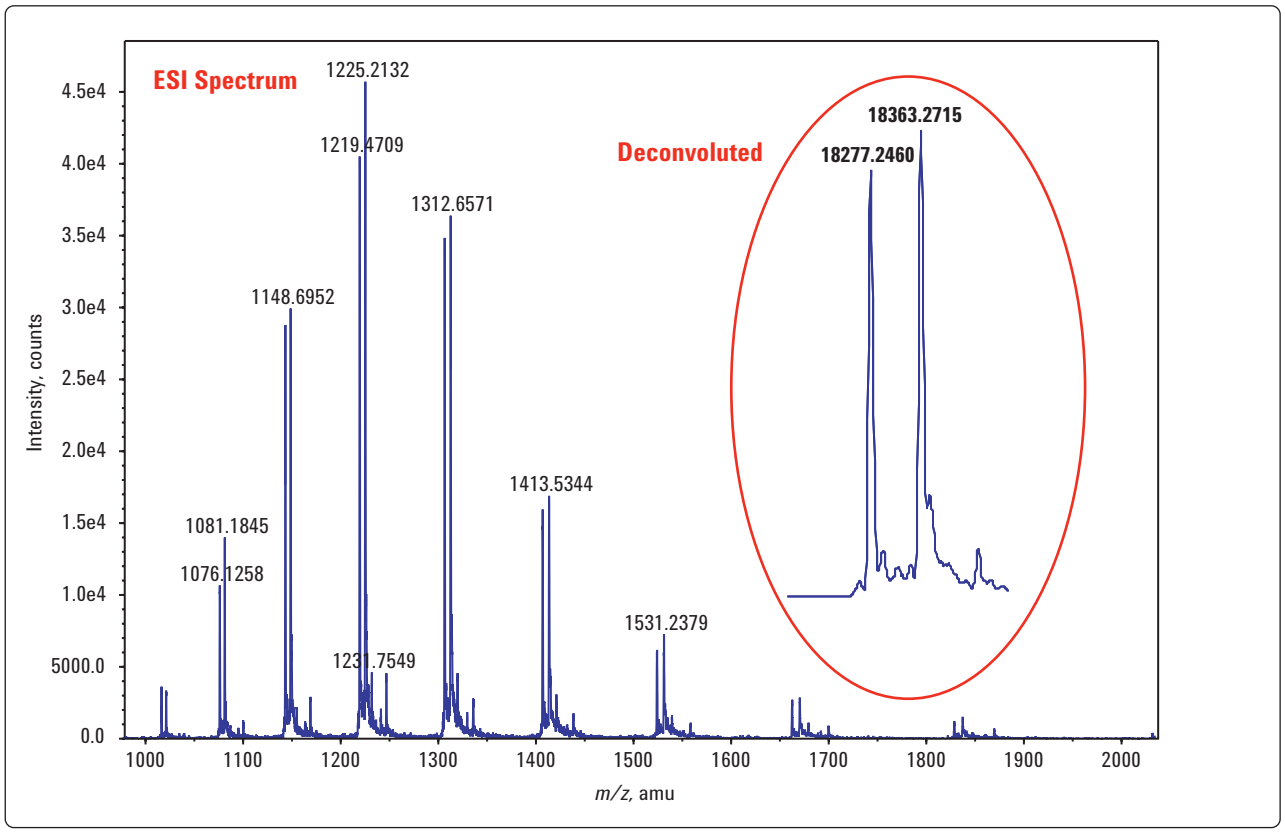
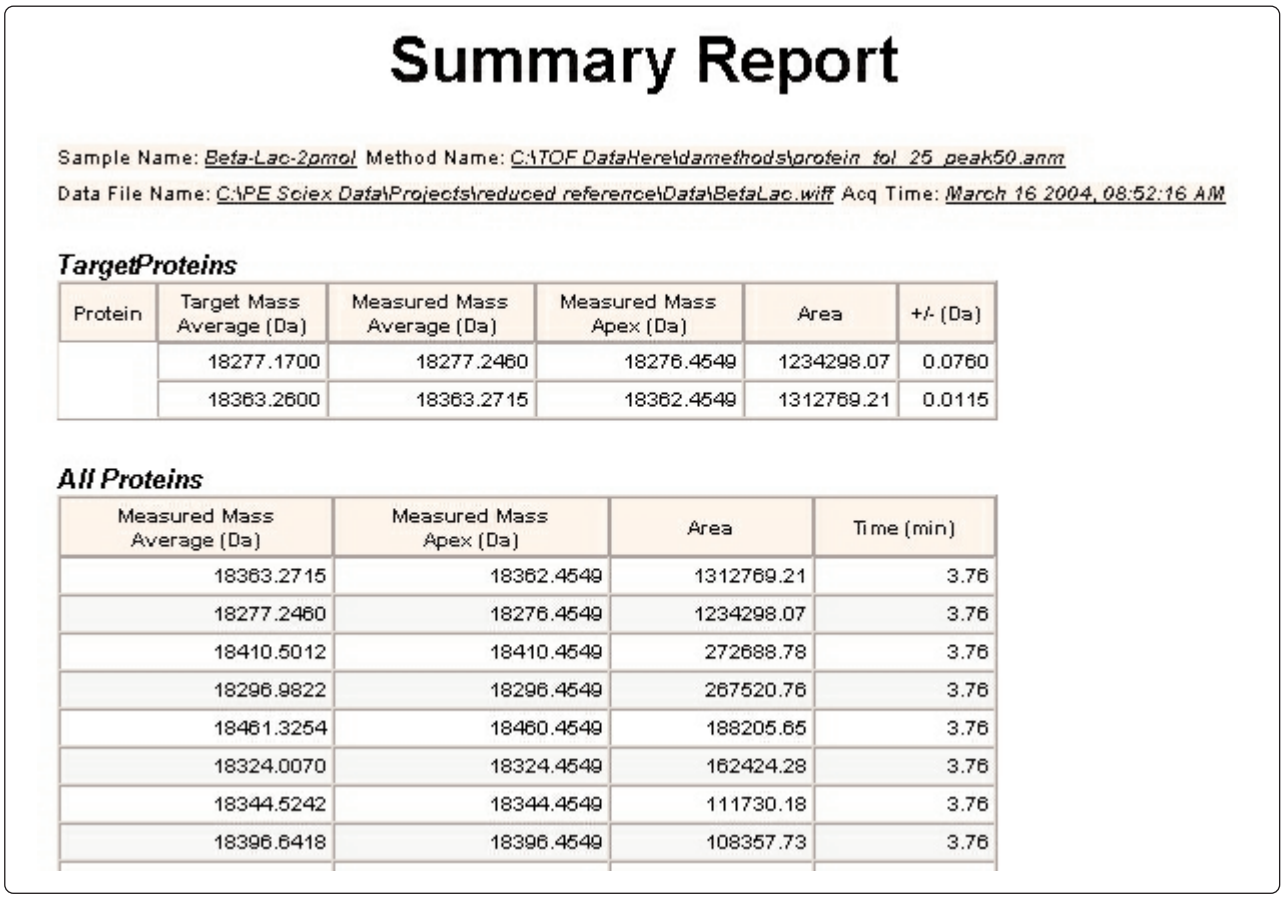


Figure 5. LC/MSD TOF analysis of β-lactoglobulin



Summary Report

Sample Name: Beta-Lac-2pmol Method Name: C:\TOF DataHere\damethods\protein_tol_25_peak50.anm

Data File Name: C:\PE Sciex Data\Projects\reduced reference\Data\BetaLac.wiff Acq Time: March 16 2004, 08:52:16 AM

TargetProteins

Protein	Target Mass Average (Da)	Measured Mass Average (Da)	Measured Mass Apex (Da)	Area	+/- (Da)
	18277.1700	18277.2460	18276.4549	1234298.07	0.0760
	18363.2600	18363.2715	18362.4549	1312769.21	0.0115

All Proteins

Measured Mass Average (Da)	Measured Mass Apex (Da)	Area	Time (min)
18363.2715	18362.4549	1312769.21	3.76
18277.2460	18276.4549	1234298.07	3.76
18410.5012	18410.4549	272688.78	3.76
18296.9822	18296.4549	267520.76	3.76
18461.3254	18460.4549	188205.65	3.76
18324.0070	18324.4549	162424.28	3.76
18344.5242	18344.4549	111730.18	3.76
18396.6418	18396.4549	108357.73	3.76

Figure 6. A portion of the automated report from the β-lactoglobulin analysis, showing target proteins confirmed with low-ppm mass errors

Table 1. Mass accuracy for protein molecular weight analysis

	Average		Delta	
Protein	**Theoretical**	**Observed**	**Da**	**ppm**
Ubiqutin	8564.83	8564.99	0.16	19
Cytochrome c bovine #1	12229.86	12229.98	0.12	10
Cytochrome c bovine #2	12229.86	12229.98	0.12	10
Cytochrome c bovine #3	12229.86	12229.79	–0.07	–5
Cytochrome c equine	12359.06	12358.98	–0.08	–6
Apomyoglobin	16951.44	16951.40	–0.04	–2
β-Lactoglobulin	18277.17	18277.14	–0.03	–2
β-Galactosidase	116352.00	116353.66	1.66	14

Conclusions

The Agilent TOF Protein Confirmation software provides extremely accurate molecular weights and offers flexible processing options. The software includes modules for calculation of molecular weights from amino acid sequences, interactive deconvolution, and automated deconvolution and protein confirmation. The combination of the Agilent LC/MSD TOF and the Agilent TOF Protein Confirmation software enables low-ppm mass accuracy for protein molecular weight analysis on an electrospray ionization time-of-flight mass spectrometer.

Authors

Donghui Yi, Irene Lee, and ***John Fjeldsted*** are scientists at Agilent Technologies in Santa Clara, California U.S.A.

www.agilent.com/chem

Printed in the U.S.A. July 16, 2004
5989-1342EN

Agilent Technologies