# Identification and Quantification of Low-Abundance Proteins in Biotherapeutics by a Sensitive and Universal LC High-Resolution MS-based Assay

*Hongxia (Jessica) Wang, Zhiqi Hao, Yi Zhang, David Horn and Patrick Bennett;*
*Thermo Fisher Scientific, San Jose, CA, USA*

Thermo SCIENTIFIC

# Overview

**Purpose:** To develop a simple, sensitive and universal platform for identification and quantification of trace-level impurities in biotherapeutics.

**Methods:** IgG mAb samples with different concentrations of low-abundance proteins were denatured, reduced and alkylated, then digested using trypsin. Samples were analyzed by online nano-flow LC and high-resolution MS. Identification and quantification of the impurity proteins were performed using software designed for qualitative and quantitative proteomics.

**Results:** Due to the difference in ionization efficiency of peptides, three out of four trace-level protein impurities were identified at 5ppm using database searching in through the software. The observed dynamic range was 5 orders of magnitude. Good linearity was observed over the range of 10-1000 ppm for Lysozyme and carbonic anhydrase in the presence of IgG, with correlation coefficients ($R^2$) greater than 0.99. Lower Limits of quantitation (LLOQs) were detected at 10 ppm in 1.0mg/mL IgG.

# Introduction

Host cell proteins (HCPs) accompanied with recombinant biotherapeutics can significantly affect drug efficacy and cause immunogenicity. Detection and quantification of residual HCPs as potential process-related impurities is critical for biopharmaceutical companies in accordance with regulatory agency guidelines. Current analytical methods suffer from long method development times, require prior knowledge of contaminant proteins, and lack the ability to detect a wide range of protein concentrations. LC-MS has been used for HCPs analysis, but qualitative and quantitative studies are traditionally performed using two different MS platforms due to the sensitivity limitations of high-resolution MS. Here, a sensitive and universal assay employing nano-flow LC coupled with a Thermo Scientific™ Q Exactive™ hybrid quadrupole-Orbitrap mass spectrometer is evaluated for simultaneous identification and quantification of low abundant proteins in biotherapeutic products.

# Methods

### Sample Preparation

Four protein standards, lysozyme (14k Da), carbonic anhydrase (29k Da), cytochrome C (12k Da) and bovine serum albumin (67k Da) were spiked in 1mg/mL of IgG mAb to final relative concentrations of 5-1000 ppm. The mixture was denatured, reduced and alkylated, then equal amount of stable isotope labeled peptides were spiked in each level. After trypsin digestion (1/20=protein/enzyme, 37 C overnight), samples were desalted by Thermo Scientific™ Pierce™ C18 Spin Tips.

### Liquid Chromatography

Trypsin digested mAb samples were analyzed on Thermo Scientific™ EASY-nLC 1000™ Liquid chromatograph coupled with a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer. Peptides were trapped using a Thermo Scientific™ Acclaim™ PepMap™ C18 column (75$\mu$m x 2cm) and separated on a Thermo Scientific™ Easy-Spray™ PepMap column (75$\mu$m x 50cm) with integrated emitter heated at 40 C at a flow rate of 250 nL/min with solvent A (0.1% formic acid in $H_2O$) and solvent B (0.1% formic acid in acetonitrile).

Injection volume: 4 $\mu$L; Loading volume: 12 $\mu$L; Max. pressure: 45 bar with Solvent A

| Time [min] | Flow [nL/min] | Mixture [%B] |
|---|---|---|
| 0 | 250 | 2 |
| 5 | 250 | 5 |
| 130 | 250 | 25 |
| 160 | 250 | 40 |
| 165 | 250 | 95 |
| 180 | 250 | 95 |

### Mass Spectrometry

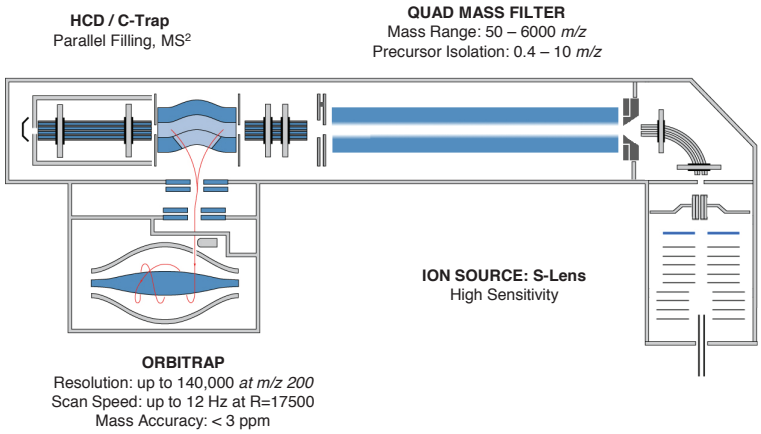
The Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Figure 1) was equipped with a Thermo Scientific™ EASY-Spray™ source was employed for MS analysis. Methods were set with full scan (140,000 resolution) and top 10 data dependant MS/MS (17,500 resolution) in positive mode. The MS parameters are shown in Table 1.

**TABLE 1. MS parameters**

| | |
|---|---|
| === Easy-Spray Source: === | |
| • Spray Voltage (+) | 1800 V |
| • Capillary Temperature (+) | 250 °C |
| • Sheath Gas (+) | 0 |
| • Aux Gas (+) | 0 |
| • Sweep Gas (+) | 0 |
| • Heater Temperature (+) | 0 °C |
| • Isolation Window(+) | 2 m/z |
| • NCE(HCD) | 27 |
| • S-lens | 50 |

• Positive MS Scan: 1 microscan; Full Scan R=140,000; AGC=3e6;
IT=120 ms; Scan range: *m/z* 400-1800; Lock Mass: off
Top 10 data-dependant MS/MS: R=17,500; AGC=1e5; IT=250ms; First mass: 130 *m/z*

**FIGURE 1. Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer**




### Data Analysis

The identification of protein impurities was performed using Thermo Scientific™ Proteome Discoverer™ Software revision 1.4. The quantitative analysis of those proteins was performed using Thermo Scientific™ Pinpoint™ Software revision 1.3.

## Results

### Identification of Low Abundance Protein Impurities

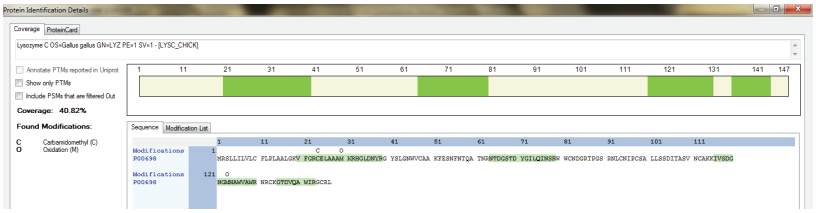
Four proteins with molecular weights from 10k to 70k Da were spiked in 1 mg/mL of mAb (150k Da) to demonstrate the sensitivity and dynamic range of the nano LC-MS based assay. In the presence of large amount of monoclonal antibody and high background noise, selectivity was greatly increased by using high resolving power at 140,000 (FWHM) in full-scan and data-dependant MS/MS at 17,500 (FWHM).

Data was searched using the SEQUEST® algorithm in Proteome Discover Software with Percolator peptide validation (1% FDR) against a customized database with targeted mAb and possible protein impurities.

Searching parameters were set as: Precursor mass tolerance: 15 ppm; Fragment mass tolerance: 20mmu; Enzyme: Trypsin; Missed cleavage: 2; Dynamic modification: Methionine oxidation; N-terminal: Gln to Pyro-Glu; Static modification: Carbamidomethylation on Cysteine. The searching results were filtered with high peptide confidence and peptide rank #1.

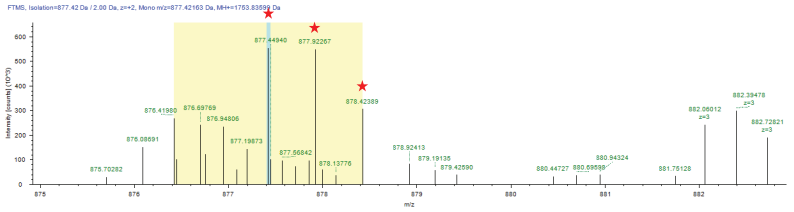
Lysozyme, bovine serum albumin and carbonic anhydrase were identified at 10 ppm with the sequence coverage of 34.92%,18.8% and 5.1% due to the ionization efficiency difference of the enzymatically cleaved peptides. Figure 2 shows the sequence coverage of Lysozyme. The concentration is 10 ppm relative to 1 mg/mL IgG. Amino acids highlighted in green are the identified peptides based on fragment ions matched with theoretical ones. The modified amino acids (AA) were also indicated above the AA in Figure 2, i.e. Cysteine 24 was carbamidomethylated, and Methionine 30 was oxidized.

**FIGURE 2. Sequence coverage of spiked 10 ppm of Lysozyme relative to 1 mg/mL IgG.**




Among four identified peptides for lysozyme, NTDGSTDYGILQINSR eluted at 92 min. Figure 3 shows the isotope distribution of the peptide at charge 2+, and the three major isotopes are indicated with a red star. The sequence was confirmed by a very good match of the experimental MS/MS spectrum from the precursor ion 877.4216 *m/z* to the theoretical one (Figure 4).

**FIGURE 3. Mass spectrum of the isotope distribution of $[M+2H]^{2+}$ peptide NTDGSTDYGILQINSR for Lysozyme.**

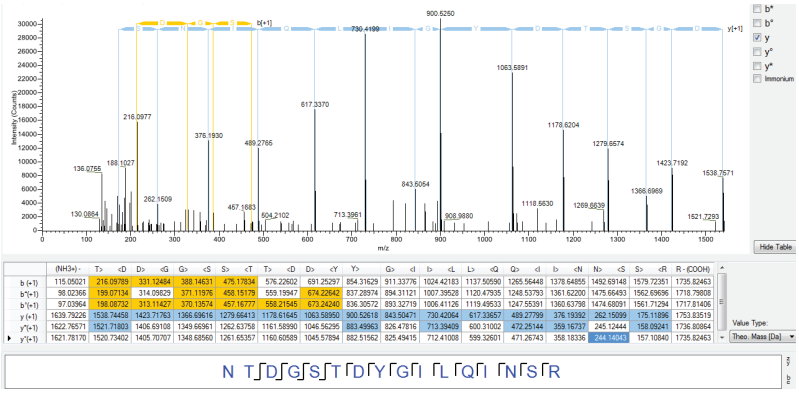



**Quantification of Protein Impurities**

Targeted protein quantification on the Q Exactive hybrid quadrupole-Orbitrap mass spectrometer can be performed by (1) Full Scan, (2) Selected Ion Monitoring (SIM), and (3) Targeted MS/MS or Parallel Reaction Monitoring (PRM). The second and third methods provide better sensitivity and selectivity compared to first. However, the first method is ideal for simultaneous quan/qual applications in discovery research.

With a full scan method, the targeted peptide is quantified using the integrated area of the extracted ion chromatogram of the monoisotopic mass or sum of the different isotopes. The post translation modifications (PTMs) of peptides can also be identified.

**FIGURE 4. Annotated MS/MS spectrum of peptide NTDGSTDYGILQINSR from [M+2H]$^{2+}$ precursor ion 877.4216 *m/z* for Lysozyme.**




Proteome Discoverer Software search results (msf files) were imported to Pinpoint Software to automatically generate the accurate mass of isotopes or MS/MS fragment ions based on the identified peptide sequence. The signature peptides from different proteins were selected for the quantitation.

**Quantification of Lysozyme and Carbonic Anhydrase in Human IgG Sample**

Lysozyme was quantified using the signature peptide NTDGSTDYGILQINSR. The signals of three most abundant isotopes at charge 2+, 877.4212, 877.9226 and 878.4240 *m/z* (in green, purple and blue respectively in Figure 5b) were summed for quantification. Integrated peak area in Figure 6 shows the concentration dependent increase. The calibration curve was generated with duplicate sample injections. Good linearity was observed from 10 ppm (10 pg/$\mu$L in IgG sample) to 1000 ppm with $R^2$=0.991 (Figure 5a).

SEQUEST search results demonstrated the sequence coverage of carbonic anhydrase was 42.31% in the IgG spiked at 1000 ppm. Three isotopes at charge 2+ of peptide QSPVDIDTK , 501.7588 (green in Figure 7b), 502.2602 (purple) and 502.761 5(blue) *m/z* were integrated to generate calibration curve. Figure 7a shows good linearity with $R^2$=0.996, 1/X, dynamic range: 10-1000ppm (pg/$\mu$L).

**FIGURE 5. Calibration curve (a) of peptide NTDGSTDYGILQINSR for Lysozyme and XICs of three most abundant isotopes (b).**

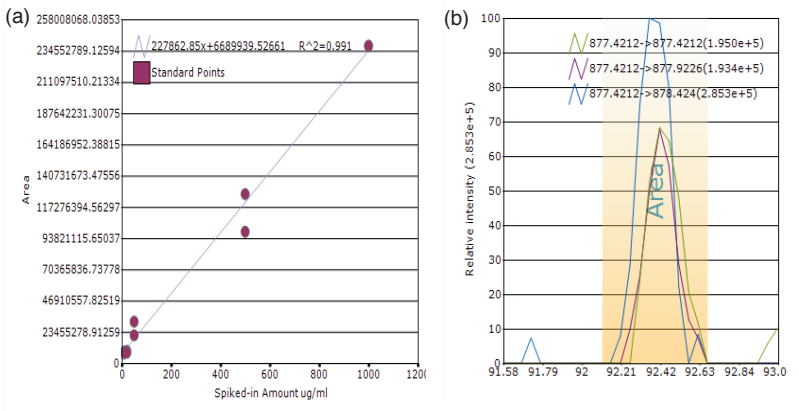


**FIGURE 6. XICs of peptide NTDGSTDYGILQINSR with isotopes 877.4212, 877.9226 and 878.4240 *m/z* within 5 ppm mass tolerance window.**

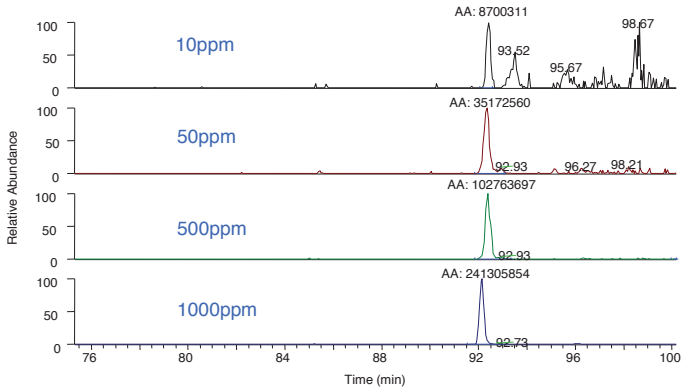
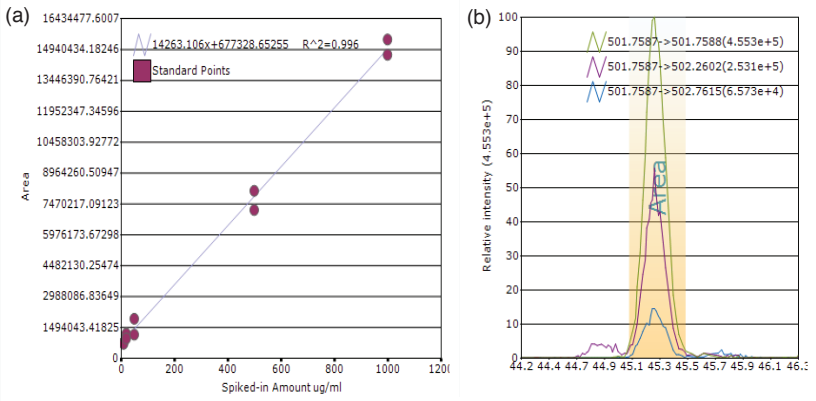



**FIGURE 7. Calibration curve (a) of peptide QSPVDIDTK for Carbonic Anhydrase and XICs of the three most abundant isotopes (b).**




# Conclusion

A simple, sensitive and universal platform was developed for identification and quantification of trace-level protein impurities in biotherapeutics. IgG monoclonal antibody spiked with low abundant proteins was used as the model to test the assay sensitivity and dynamic range.

- The observed dynamic range was 5 orders of magnitude. Lysozyme, carbonic anhydrase and bovine serum albumin were detected at 5 ppm level relative to 1 mg/mL of IgG antibody.
- Good linearity was observed over the range of 10-1000 ppm for Lysozyme and carbonic anhydrase in the presence of IgG, with correlation coefficients ($R^2$) of 0.991 and 0.996 (n=2), respectively.
- The simultaneous qualitative and quantitative analysis of low abundance proteins was achieved with single injection on the Q Exactive hybrid quadrupole-Orbitrap mass spectrometer. This LC-MS based approach can be applied for host cell proteins generated during drug process development by recombinant DNA technology.
- The combination of the EASY-Spray source and EASY-Spray column with integrated emitter and column makes trace-level protein impurity nano-flow analysis more close to routine.

**www.thermoscientific.com**




*Thermo Fisher Scientific, San Jose, CA USA is ISO 9001:2008 Certified.*

**Africa-Other** +27 11 570 1840
**Australia** +61 3 9757 4300
**Austria** +43 1 333 50 34 0
**Belgium** +32 53 73 42 41
**Canada** +1 800 530 8447
**China** +86 10 8419 3588
**Denmark** +45 70 23 62 60
**Europe-Other** +43 1 333 50 34 0
**Finland/Norway/Sweden** +46 8 556 468 00
**France** +33 1 60 92 48 00
**Germany** +49 6103 408 1014
**India** +91 22 6742 9434
**Italy** +39 02 950 591
**Japan** +81 45 453 9100
**Latin America** +1 561 688 8700
**Middle East** +43 1 333 50 34 0
**Netherlands** +31 76 579 55 55
**New Zealand** +64 9 980 6700
**Russia/CIS** +43 1 333 50 34 0
**South Africa** +27 11 570 1840
**Spain** +34 914 845 965
**Switzerland** +41 61 716 77 00
**UK** +44 1442 233555
**USA** +1 800 532 4752

Thermo SCIENTIFIC

Part of Thermo Fisher Scientific

ASMS13_M502_JessicaWang_E 06/13S