thermoscientific

Detection and Identification of New Features as Part of Mass Spectrometry-based Quality Control

Michael Blank, Stephane Houel, and Jonathan Josephs, Thermo Fisher Scientific, 355 River Oaks Pkwy, San Jose, California 95134

ABSTRACT

Purpose: The GMP compliant detection and quantitation of low level proteins in a drug substance modeling host cell proteins

Methods: An integrated robust workflow utilizing high resolution accurate mass LCMS methods and automated processing and reporting.

Results: Quantitative limits down to 10 ppm for all of the spiked proteins with at least two peptides.

INTRODUCTION

In addition to the expression of the desired drug molecule, transgenic host cells such as the widely employed Chinese Hamster Ovary cell (CHO) manufacture a number of native proteins associated with cell life cycle and maintenance. While these proteins are often excluded from the final drug product through a series of purification and chromatographic polishing steps, some are retained and present immunogenic effects that can adversely affect the safety of the therapeutic drug. Consequently the detection and quantification of these proteins is of great interest, but is challenging as many of these host cell proteins (HCPs) are at extremely low stoichiometry (on the order of parts per million) compared to the drug molecule. This work incorporates targeted ultra low level protein quantitation and non-targeted detection into a GMP-compliant multi attribute (MAM) quality control method using mass spectrometry.

MATERIALS AND METHODS

Sample Preparation

The NIST mAb standard commercially available at 10 mg/mL was adjusted to 1 mg/mL using 90 uL of 7M guanidine HCl with TRIS buffer. To this solution, 500 mM DTT was added as a reducing agent and reacted for 30 minutes at RT. Iodoacetic acid at 500 mM was added to alkylate the sample for 20 minutes. The sample was thoroughly buffer exchanged to remove the guanidine, following digestion with Pierce™ LCMS Grade trypsin in a 1:10 ratio for 30 minutes at 37 C. Digestion was quenched using 10% formic acid. The Pierce™ 6 Protein digest was spiked in at a calculated 1% molar ratio to the NIST mAb. A serial dilution was performed to reduce the concentration of the 6 proteins to 0.1% and ultimately to 0.001% (10 ppm concentration)[1].

Test Method(s)

Proteins and peptides present in the product and spiked samples were detected using data dependent acquisition (DDA). On the Thermo Scientific™ Q Exactive™ HF mass spectrometer under control of Thermo Scientific™ Chromeleon™ CDS software , a 120,000 FWHM @ 200 m/z MS scan was followed by the HCD fragmentation of the top 5 most abundant precursors and detection of fragment ions at 30,000 resolution. For the MAM method, the data were acquired in triplicate using MS1 only from 300 to 1800 m/z at 120,000 resolution at an AGC target of 3e6.

Chromatographic separation and delivery to the mass spectrometer was performed with a Thermo Scientific™ Dionex™ Vanquish™ Horizon uHPLC system coupled to a Thermo Scientific™ Accucore™ C18 uHPLC, 1.5 µm, 2.1 x 150 mm column maintained at 60 C. The binary solvent system consisted of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) flowed at a fixed rate of 0.20 mL/min. A multistage gradient was employed beginning with an equilibration phase for 5 minutes at 1% B before increasing abruptly to 10% at 6 minutes and eventually attaining 35% after 70 minutes. A column regeneration cycle at 90% B followed from 72 to 77 minutes before returning to 1% B at 79 minutes. A second “blanking gradient” see below brought the total run time to 115 min.

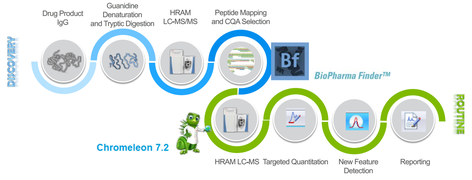
3 uL containing approximately 2.5 ug (15 pmol) of NISTmAb and as little as 0.15 fmol (10 ppm) of Pierce 6 Protein digest were loaded onto column for each injection which was performed in triplicate for each concentration down to 10 ppm. The eluent from the LC was diverted to waste for 2 minutes prior to being switched in line with the mass spectrometer. It was again diverted to waste at 70 minutes concurrent with the beginning of the column regeneration step.

Data Analysis

The MS/MS data of the neat injection of the 6 protein digest were searched against a combined FASTA file of their respective sequences in Thermo Scientific™ BioPharma Finder™ software with carboxymethylation (+58.005 Da) as a fixed modification and without dynamic modifications, as these were expected to be of extremely low abundance and unsuitable for protein quantitation. The resultant list of peptides with a confidence of greater than 0.90 were exported as a target list for monitoring in a soon to be released version of Chromeleon 7 software based on 7.2 SR5.

The MS spectral raw data acquired by Chromeleon software were extracted using 5 ppm extraction tolerance for the targeted analysis with confirmation of peptides based upon their isotopic correlation being greater than 0.90 and their mass accuracy all confirming isotopes being less than 5 ppm. A weak moving average smoothing was applied.

Figure 1. Multi Attribute Method (MAM) workflow summary. A peptide mapping experiment followed by analysis in BioPharma Finder comprises the discovery phase to identify salient peptides. A targeted peptide list is then handed off to Chromeleon to routinely and GMP compliantly monitor these targets, detect new features, and report out.



RESULTS

Peptide Mapping

A search of the peptide mapping data for the neat injection of the Pierce 6 Protein Digest revealed a total 384 components identified from their MS/MS spectra. Identification took into account not only mass accuracies of both the precursor and the fragment ions but also the correlation between the theoretically predicted fragmentation spectra and the observed spectra. For many of the detected peptides, multiple charge states were observed. This list was filtered only for those of the very highest confidence and then exported into Chromeleon software for method optimization.

The imported list was substantially larger than any that would be typically used for an MAM method but was able to be processed handily by this version of Chromeleon software without issue. This enabled an opportunity to execute further method development that sought to eliminate peptides with poor signal to noise, interference from the NIST mAb, low ionization efficiency, or unfavorable chromatographic behavior leaving a refined optimized method. Additionally, in the cases of multiple charge states usually at least one would be eliminated. While the inclusion of multiple charges is highly beneficial for drug product attribute monitoring as they can serve as additional confirmatory evidence and be used as part of the scoring algorithm, for the detection and quantitation of low level foreign proteins, some charge states were simply unsuitable as they did not have sufficient signal to noise at lower concentrations.

Figure 2: BioPharma Finder results showing the identification of a peptide based on the correlation between the observed MS/MS spectra and the theoretically predicated fragmentation pattern. High correlation and mass accuracies less than 3 ppm lead to a confident assignment. This peptide and many others can then be directly exported to Chromeleon for monitoring as part of MAM.

Targeted Quantification

As part of a system suitability evaluation, several critical criteria were evaluated prior to commencing primary data analysis, including retention time, peak area reproducibility, and mass accuracy. Several representative peptides from the NIST mAb sequence were used to evaluate this over three replicate runs. Average retention time variation was 0.1%, average peak area variation was 1.5%, and the median mass accuracy was 1 ppm, which was deemed acceptable for analysis to continue.

With the optimized list, it was possible to detect at least two peptides from each of the six proteins spiked into the NIST mAb digest to quantitatively model the behavior of HCPs (Table1). In many cases all four confirming isotopes could clearly be observed at 5 ppm mass accuracy (Figures 3, 4, and 5) and could be automatically detected and quantified, especially when matched against the imported retention time values, lending a high degree of confidence to the assignment of these peptides and their resulting integration. This is coupled together with the fact that these peptides have already been identified previously using MS/MS.

Not unexpectedly, the two smallest proteins (Bovine Cytochrome C and Chicken Lysozyme) were the most difficult to detect as the pool of tryptic peptides to monitor was much smaller than the remaining four, not to mention that by mass these were by far the lowest in concentration (around 1 ppm). Overall, each the peptides in Table 1 could be detected in each of the replicates, even at 10 ppm concentration, but not every targeted peptide was able to be effectively quantified according to ICH guidelines.

Table 1. Proteins and their peptides detected and quantified at 10 ppm concentration relative to the NIST mAb drug substance. At least two peptides for each protein could be detected and many quantified based on their linearity and coefficients of variation between technical replicates.

Protein Description	Peptide Name (Charge)	Peak Areas (counts*secs) 10ppm_ 1	10ppm_ 2	10ppm_ 3	Coefficient of Variation (%) 0.1% 1	0.1% 2	1.0% 1	1.0% 2	10 ppm	0.1%	1%
Bovine Cytochrome C	(+3) TGPNLHGLFGR	6.4E+03	9.5E+03	8.7E+03	5.2E+05	5.2E+05	5.2E+05	4.7E+06	4.8E+06	4.6E+06	22.5	0.3	1.6
Bovine Cytochrome C	(+3) GITWGEETLMEYLENPK	4.5E+03	3.9E+03	3.1E+03	2.5E+05	2.6E+05	2.5E+05	1.9E+06	1.8E+06	1.8E+06	18.8	1.2	1.7
Bovine Serotransferrin	(+3) WCJ+58]TSTHEANK	3.4E+03	6.7E+03	3.9E+03	3.8E+05	3.7E+05	3.8E+05	3.2E+06	3.3E+06	3.3E+06	38.6	1.3	2.6
Bovine Serotransferrin	(+2) LC[+58]QLC[+58]AGK (+3)	2.9E+03	3.5E+03	2.4E+03	3.0E+05	3.1E+05	3.1E+05	2.8E+06	2.7E+06	2.7E+06	17.8	1.6	2.3
Bovine Serotransferrin	SVDDYQ[C[+58]YLAMVPSHAVVAR	6.3E+03	6.5E+03	6.4E+03	4.7E+05	4.6E+05	4.6E+05	3.9E+06	4.0E+06	4.0E+06	1.2	1.3	1.5
Bovine Serum Alb.	(+2) C[+58]C[+58]TESLVNR	2.7E+03	2.8E+03	3.9E+03	3.1E+05	3.0E+05	2.8E+05	2.8E+06	3.0E+06	2.9E+06	19.8	5.7	2.9
Bovine Serum Alb.	(+2) LVTDLTK	1.0E+04	1.1E+04	9.0E+03	6.2E+05	6.5E+05	6.4E+05	5.0E+06	5.9E+06	5.8E+06	8.3	2.5	1.3
Bovine Serum Alb.	(+2) YLYEIAR	7.7E+03	7.6E+03	6.8E+03	6.6E+05	6.5E+05	6.5E+05	5.6E+06	5.5E+06	5.5E+06	6.8	0.7	0.7
Bovine Serum Alb.	(+2) HLVDEPQNLIK	4.6E+03	5.5E+03	5.9E+03	4.6E+05	4.6E+05	4.6E+05	3.7E+06	3.7E+06	3.5E+06	12.5	0.5	1.6
Bovine Serum Alb.	(+3) RHPEYAVSVLLR	2.3E+03	3.3E+03	4.3E+03	1.0E+06	1.0E+06	9.9E+05	7.6E+06	7.6E+06	7.4E+06	30.8	1.3	1.2
Bovine Serum Alb.	(+2) LVVSTQTALA	1.8E+04	1.8E+04	1.8E+04	5.0E+06	4.8E+06	4.9E+06	4.7E+07	4.5E+07	4.3E+07	1.9	2.7	2.3
Bovine Serum Alb.	(+3) KVPQVSTPTLVEVSR	9.7E+03	1.6E+04	1.6E+04	9.0E+05	9.0E+05	8.7E+05	8.1E+06	8.3E+06	8.1E+06	7.9E+06	25.6	1.9	1.4
Bovine Serum Alb.	(+2) LVNELTEFAK	4.1E+03	3.3E+03	5.0E+03	7.1E+05	7.1E+05	6.9E+05	6.7E+06	6.8E+06	6.6E+06	21.3	1.9	0.4
Bovine Serum Alb.	(+2) LGEYGFQNALIVR	1.7E+04	1.4E+04	1.8E+04	7.8E+05	8.3E+05	7.2E+05	5.9E+06	5.8E+06	5.0E+06	11.2	1.9	0.5
Chicken Lysozyme	(+2) GTDVQAWIR	3.7E+03	3.0E+03	1.3E+03	2.6E+05	2.7E+05	2.6E+05	2.3E+06	2.3E+06	2.3E+06	46.4	1.3	0.7
Chicken Lysozyme	(+2) NTDGSTDYGILQINSR	2.6E+03	2.5E+03	1.7E+03	1.8E+05	1.7E+05	1.8E+05	1.6E+06	1.6E+06	1.6E+06	22.9	3.6	0.9
E.coli Beta-Galactosidase	(+2) FNDDFSR	5.8E+03	2.9E+03	4.9E+03	3.6E+05	3.6E+05	3.6E+05	3.3E+06	3.1E+06	3.1E+06	32.5	0.7	6.4
E.coli Beta-Galactosidase	(+2) WVGYGQDSR	3.2E+03	3.7E+03	3.0E+03	2.9E+05	2.9E+05	2.8E+05	2.7E+06	2.6E+06	2.5E+06	11.1	2.2	2.3
E.coli Beta-Galactosidase	(+2) IDPNAWVER	4.9E+03	1.8E+03	3.1E+03	4.2E+05	4.3E+05	4.1E+05	3.7E+06	3.7E+06	3.6E+06	48.6	2.6	1.6
E.coli Beta-Galactosidase	(+2) VDEDQPFPAVVK	3.5E+03	5.1E+03	1.2E+03	4.6E+05	4.8E+05	4.6E+05	4.2E+06	4.1E+06	4.3E+06	58.9	1.9	2.1
E.coli Beta-Galactosidase	(+2) WLPAMSER	3.3E+04	3.7E+04	3.3E+04	3.8E+05	3.7E+05	3.8E+05	3.1E+06	3.1E+06	3.1E+06	5.6	1.6	0.6
E.coli Beta-Galactosidase	(+3) LAAHPPASWIR	4.2E+03	4.1E+03	4.5E+03	9.0E+05	9.0E+05	8.5E+05	8.4E+06	8.4E+06	8.4E+06	5.2	1.0	0.1
E.coli Beta-Galactosidase	(+3) YHYQLVWC[+58]QK	1.2E+04	5.6E+03	1.2E+04	5.3E+05	5.4E+05	5.4E+05	4.8E+06	4.9E+06	4.7E+06	37.1	0.4	1.6
E.coli Beta-Galactosidase	VVQPNATAWSEAGHISAWQQWR (+3)	2.3E+04	2.5E+04	2.6E+04	9.9E+05	1.0E+06	9.8E+05	8.8E+06	8.9E+06	8.7E+06	7.7	2.7	1.6
E.coli Beta-Galactosidase	AVVELHTADGTLIEAEACJ+58]DVGFR	4.1E+04	4.9E+04	3.8E+04	1.0E+06	1.0E+06	1.0E+06	9.9E+06	9.6E+06	9.4E+06	12.8	0.9	1.1
E.coli Beta-Galactosidase	(+3) LSGQTIEVTSEYLFR	8.9E+03	7.9E+03	9.7E+03	4.4E+05	4.4E+05	4.2E+05	3.3E+06	3.2E+06	3.2E+06	9.9	1.9	1.8
E.coli Beta-Galactosidase	(+2) LSGQTIEVTSEYLFR	7.2E+03	1.0E+04	1.3E+04	5.2E+05	5.2E+05	5.3E+05	4.0E+06	4.1E+06	4.0E+06	27.1	1.3	0.7
E.coli Beta-Galactosidase	(+2) LPSFIETDLSAFLR	2.2E+04	2.6E+04	2.5E+04	1.2E+06	1.2E+06	1.2E+06	1.0E+07	1.0E+07	9.8E+06	12.8	1.0	0.9
Yeast Alc. Dehydrogenase	(+3) GVIFYESHGK	6.5E+03	6.3E+03	6.9E+03	5.7E+05	5.9E+05	5.9E+05	5.1E+06	5.2E+06	5.1E+06	5.0	1.6	1.9
Yeast Alc. Dehydrogenase	(+3) ATDGGAHGVINVSVSEAAIEASTR	4.2E+04	3.4E+04	4.0E+04	1.1E+06	1.1E+06	1.1E+06	1.0E+07	1.0E+07	9.7E+06	11.0	0.8	1.4
Yeast Alc. Dehydrogenase	(+2) EALDFFAR	2.5E+03	6.0E+03	4.1E+03	5.7E+05	5.6E+05	5.4E+05	5.0E+06	5.1E+06	5.0E+06	41.7	2.4	0.4
Yeast Alc. Dehydrogenase	(+3) LPLVGGHEGAGVVVGMGENVK	2.2E+04	3.3E+04	2.1E+04	1.1E+06	1.1E+06	1.1E+06	8.0E+06	7.9E+06	7.8E+06	25.5	3.6	1.1
Yeast Alc. Dehydrogenase	(+2) VVGSTLPEYTEK	8.0E+03	7.7E+03	1.1E+04	9.0E+05	8.8E+05	8.9E+05	7.7E+06	7.9E+06	7.7E+06	22.6	1.1	1.7
Yeast Alc. Dehydrogenase	(+2) SIGGEVFIQFTK	3.0E+04	2.9E+04	3.3E+04	6.9E+05	6.9E+05	6.9E+05	5.8E+06	5.8E+06	5.7E+06	7.1	0.5	0.7

Figure 3: XICs for *LPLVGGHEGAGVVVGMGENVK* a yeast alcohol dehydrogenase 1 peptide at 10 ppm concentration vs. NIST mAb Digest

Figure 4: XICs for *LPLVGGHEGAGVVVGMGENVK* a yeast alcohol dehydrogenase 1 peptide at 0.1% concentration vs. NIST mAb Digest

Figure 5: XICs for *LPLVGGHEGAGVVVGMGENVK* a yeast alcohol dehydrogenase 1 peptide at 1.0% concentration vs. NIST mAb Digest

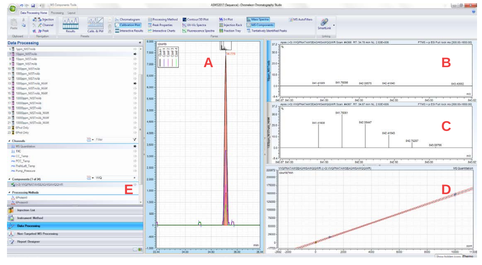
As can be discerned from Table 1, coefficients of variation between technical replicates for the 6 Protein Digest peptides were extremely good at 1% and 0.1% concentration versus the NIST mAb and ranged between 0.1 and 5.7%. Many of the peptides had CVs of less than 20% even at the 10 ppm concentration condition and there were more than a dozen peptides with single digest coefficients of variation at this spike-in level.

However, CVs are only one aspect of quantitation. By coupling the peak integration results with an evaluation of the quantitation linearity, as shown in Figure 7, it was possible to demonstrate that many were not only reproducible at 10 ppm concentration but could also be linearly quantified thanks to the high mass accuracy and resolution that defeated any interferences that were present, allowing good selectivity and sensitivity when monitoring such low level features.

While there was no need to further complicate what is a robust and easy GMP method by the use of MS/MS, as the identity of these peptides had already been confirmed prior to import, it was nevertheless essential to have some confirmation strategy. A peptide scoring algorithm was applied to automatically score the detection based on the criteria of mass accuracy being less than 5 ppm for all isotopes and that all isotopes (down to either M+2 or M+3) were present. An example is seen in Figure 7 where a green check mark as automatically been applied to indicate confident detection even at 10 ppm concentration.

This work also helps to establish an approximate detection limit for the targeted method aspect of MAM which does appear to be between 0.1 and 1 fmol on column depending of the specific physiochemical properties of the peptide. Factors like chromatographic resolution, ionization efficiency, and analysis time will all play a role in what can be detected and to what limit. Nevertheless, 10 ppm relative abundance is clearly and confidently within reach.

Figure 7: Chromeleon Studio dashboard showing quantitation of VVQPNATAWSEAGHISAWQQWR, an *E. coli* β-galactosidase peptide. A) The XIC of the 10 ppm concentration condition extracted at 5 ppm mass tolerance including the top four isotopes. B) The raw MS isotopic pattern for the 10 ppm concentration condition. C) the raw MS isotopic pattern for the 0.1% concentration condition for comparison. D) The computed calibration curve exhibiting very good linearity of quantitative measurements with 99.5% confidence limits shown in red E) the score result for this peptide indicating successful detection and confirmation.

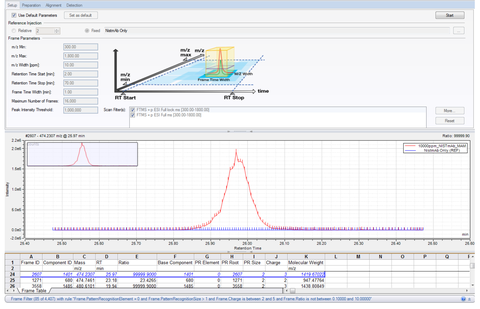


New Feature Detection

Non-targeted MS processing evaluation serves as an important purity check as part of the MAM method. The test verifies that no new components are present at a given threshold relative to the drug substance as well as that there are no significant changes within the peptide map itself. This is analogous to the conventional method of a visual inspection of different UV-chromatograms for a peptide map , but with the tremendously enhanced specificity and sensitivity of additional dimension of high resolution accurate mass spectrometry.

The chosen threshold of 1E6 corresponds to about 0.1% of the TIC and is well suited to the detection of impurities, but not those at the level of host cell proteins (less than 100 ppm). While choosing a lower threshold is certainly possible, the function of the non-targeted MS processing is not to detect host cell proteins (this is better left to the targeted workflow discussed previously), but rather to pick up large impurities or substantial unexpected changes in the product, a final quality check if you will.

Figure 7: Non-targeted MS processing studio panel showing the detection of a new feature at 474.2307 @ 25.97 minutes above 0.5% of the TIC. This impurity was found in the 1% concentration condition for the 6 Protein Mix spiked into the NISTmAb digest and so is to be expected.



The 1% 6 Protein Digest spike-in condition is an ideal positive control for the evaluation of Non-Targeted-MS processing. While no new features were seen comparing replicate injections, as anticipated, a large number of new features were detected when comparing the NIST mAb digests containing the six protein mix to the NIST mAb digest alone. The features detected (more than 4000 above the threshold) were filtered down to only those containing a clear monoisotopic peak, having an envelope with at least two isotopes, being between charge state 2 and 5, and showing a fold change of more than 10 (i.e. truly new). Only 85 features remained after this filtering, one of which, m/z value 474.2307 is shown in Figure 7.

While this peptide is expected to be one of the peptides of the 6 Protein Mix, its actual identity can be easily confirmed by returning to the extensive peptide maps within BioPharma Finder. Reviewing the results, it is the same peptide shown in Figure 2, belonging to Bovine Serum Albumin (SLHTLFGDELCK) and can quickly be added to the targeted processing method for further automated monitoring.

CONCLUSIONS

- The ability to directly export BioPharma Finder’s comprehensive peptide maps into Chromeleon software provided a seamless link between discovery of significant attributes and associated peptides and their 21 CFR Part 11 compliant monitoring in Chromeleon.
- The improved performance of Chromeleon software enabled the detection and quantitation of hundreds of peptides as part of method optimization and refinement.
- The standard MAM method outlined herein is suitable to routinely monitor and quantitate low level external proteins on the order of 10 ppm concentration (1 ppm by mass) in a GMP compliant manner.

REFERENCES (if necessary)

1. Ren, Da; Pipes, Gary D; Liu, Dingjiang; Shih, Liang-Yu; Nichols, Andrew C.; Treuheit, Michael J.; Brems, David N.; Bondarenko, Pavel V., An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Analytical Biochemistry*, **2009**, 392, 12–21.

2. R. Rogers, et al. Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics *Mabs*, **2015**, 7, 881-890

TRADEMARKS/LICENSING