Insights
Genedata Expressionist

FEATURED ARTICLE

David Bush, Ph.D., Scientific Account Manager, Genedata Inc, Lexington, MA, USA

Host Cell Protein Monitoring by Mass Spectrometry: From Raw Data to Final Report

Mass spectrometry (MS) is emerging as the most promising technique to supplement immunoassays for the analysis of host cell proteins (HCPs). Genedata Expressionist® provides the biopharma industry with an instrument- and vendor-independent software platform allowing innovative analytical approaches that address the major challenges related to the identification, quantification, and routine monitoring of HCPs by MS.

Introduction

Biotherapeutics based on recombinant proteins are produced at the industrial scale using genetically engineered expression systems (e.g. bacteria, yeasts, plants). In particular, mammalian cells, such as CHO cells, have become the preferred hosts for manufacturing biopharmaceuticals due to their capacity to provide appropriate protein folding, assembly, and post-translational modifications. Besides producing the biotherapeutics protein, host cells also encode their own natural proteins that can contaminate the final product. In fact, low amounts of residual HCPs surviving the purification steps can cause product degradation (e.g. residual proteases) or even lead to immune response in patients. For these reasons, regulatory authorities have defined guidelines for the monitoring of residual HCPs in biotherapeutics.

Immunoassays, e.g. ELISA, are the current industry standard for HCP measurement, yet they suffer from several limitations: (a) assay development is expensive and time-consuming; (b) HCP detection relies on the antibody serum used, which does not comprehensively recognize all HCPs (e.g. non-immunoreactive proteins); and (c) immunoassays typically do not provide any information on the identity or quantity of individual proteins, hampering the adoption of efficient purification schemes that could target specific components of the host cell's proteome.

MS-based methods orthogonal to immunoassays are being developed to address these limitations. Allowing unbiased identification and high throughput quantification of multiple low-abundant proteins, MS enables a thorough risk-based assessment of HCPs in biotherapeutics and is proving to be the ideal analytical platform to work alongside immunoassays. However, the use of MS for HCP analysis also poses some challenges:

1) HCP contaminations can span a wide range of concentrations, with low abundant species present at the ppm level;
2) a number of analytical approaches are currently being explored, however no standardized methods are available yet;
3) robustness and reproducibility issues with complex MS instrumentation and methods delay deployment in regulated environments such as production and quality control (QC).

Thanks to best-in-class algorithms for data pre-processing, versatility to adapt to any lab workflow, and advanced automation capabilities, Genedata Expressionist has proved to be instrumental in addressing these challenges. The software is currently the platform of choice of many enterprises that are implementing HCP analysis and monitoring by MS.

1) Data pre-processing facilitates low abundant HCP detection and quantification

Raw data from any MS instrument can be loaded directly into Genedata Expressionist. The software utilizes highly efficient algorithms for mass recalibration, noise removal, and retention time alignment to obtain optimal peak detection even for low abundant species. The implemented peak detection method is sensitive over a wide range of intensities, since it

Genedata
solutions in silico

is designed to add up evidence for peaks collected in different samples and at the same time uses consistency checks to distinguish real biological signals from remnant noise. Moreover, isotope clusters are built taking into account the known intensity of peptides' isotopic patterns. In this way, monoisotopic peaks can be reconstructed even for low abundant peptides with peak intensities below the detection threshold. The refined algorithms and advanced data visualization pro-

Figure 1: *Genedata Expressionist workflow for a two-stage HCP identification procedure.*

vided by Genedata Expressionist allow optimization of the data pre-treatment to facilitate identification and quantification of HCPs.

2) Workflow-based software enables diverse HCP analytical strategies

In the context of evolving regulatory guidelines, the proliferation of different MS approaches for the analysis of HCPs requires a software solution adaptable to different data analysis and reporting needs. Genedata Expressionist ensures unparalleled flexibility thanks to the concept of automated, yet customizable workflows. The software allows setting up a sequence of activities, optimizing them, and saving them in a workflow for later use. These operations enable customization of data analysis to specific methods and instrumentation, and they are normally performed just once by an expert at the beginning of a project. The two specific applications of Genedata Expressionist workflows described below demonstrate the flexibility of the software in supporting different analytical paradigms for HCP analysis.

Example 1: Monitoring low abundance HCPs using a two-stage identification procedure. In order to separate the signals deriving from HCPs from the dominant peaks of the biotherapeutic, the workflow reported in Figure 1 was applied.
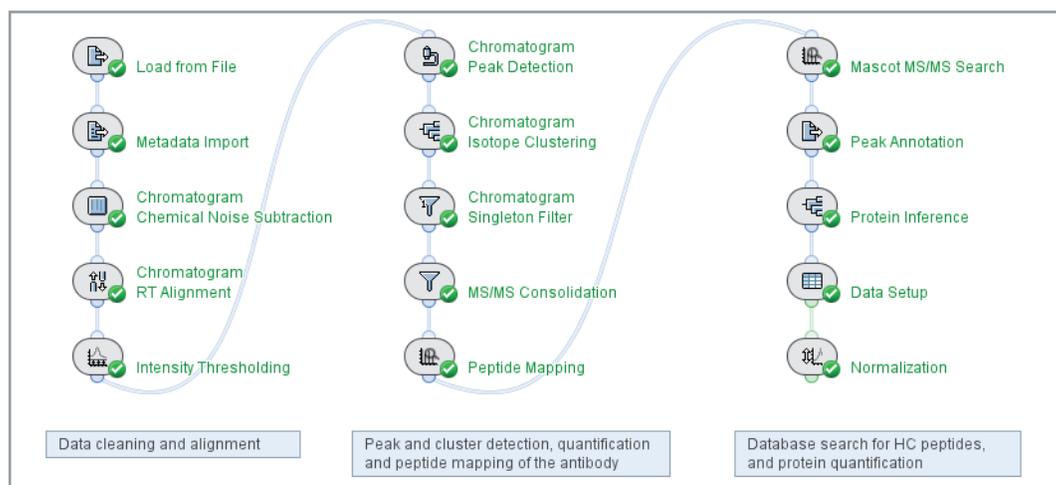
The strategy was to identify the whole collection of signals belonging to the protein biotherapeutics (e.g. peptides, glycopeptides, modifications, truncations, disulfide bonded peptides, etc.) before submitting the data to conventional peptide-spectrum match (PSM) searches. This not only allowed the researcher to monitor and quantify the expected peptides, but also to reduce the chance of low abundance features from the biotherapeutics to be falsely identified as HCP signals (false positives). However, for a more in-depth analysis, it is also possible to submit to search engines the peaks and clusters already annotated in the first step. In case of conflicting annotations, the scientist can decide if a peak of interest should be investigated as a potential HCP.

In a similar approach, HCP analysis is performed by combining two-dimensional liquid chromatography (2D-LC) with MS detection by data-independent acquisition. Sample fractionation can provide the increased sensitivity and wide dynamic range required for monitoring HCP impurities in biopharmaceutical drugs. With a simple modification of the workflow, Genedata Expressionist offers to scientists the possibility to apply the two-staged identification procedure described above to the 2D-LC-MS approach. Data analysis and result reporting can be automatically applied across the fractions as shown in Figure 2.
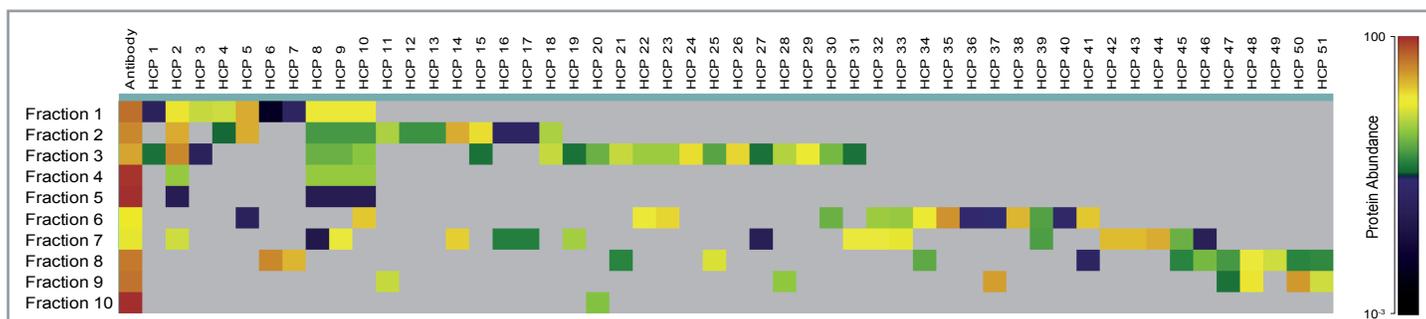

Figure 2: *Heat map of percent abundances of an antibody and enriched HCPs using sample fractionation in conjunction with the workflow shown in Figure 1. Rows correspond to different fractions.*

Example 2: Developing and applying corporate knowledge base of HCPs to downstream samples. Submitting mass spectra from low-abundant signals to PSM algorithms also poses the risk of missing the identification of contaminant proteins (false negatives). One reason is that high quality fragmentation data needed for MS/MS library searches is typically difficult to obtain for low abundance peaks, in particular when using data-dependent acquisition (DDA). Genedata Expressionist users can overcome this issue in the following way: potential HCPs can be identified by PSM search on samples with enriched HCP content, and the information pertaining to the respective peptides can then be used by the software as identification criteria for any other samples.

For example, m/z and retention time coordinates of known impurities can be stored in a corporate knowledge base and used for matching of the respective signals in single stage MS data (Figure 3). Also, fragmentation spectra of peptides can be stored to create user-defined spectral libraries to identify HCPs by tandem MS. These libraries can then be used in combination to annotate signals in downstream samples and allow for proper identification even in cases where no fragmentation data is available.

3) Automation enables MS-based HCP monitoring

Industry is moving toward using MS to perform routine HCP monitoring of large sample batches in production and QC. For example, high-throughput assays based on multiple reaction monitoring (MRM) are often applied as a fast, quantitative testing method. Genedata Expressionist offers high quality automated data analysis, applying settings optimized throughout the development process of the biotherapeutic.
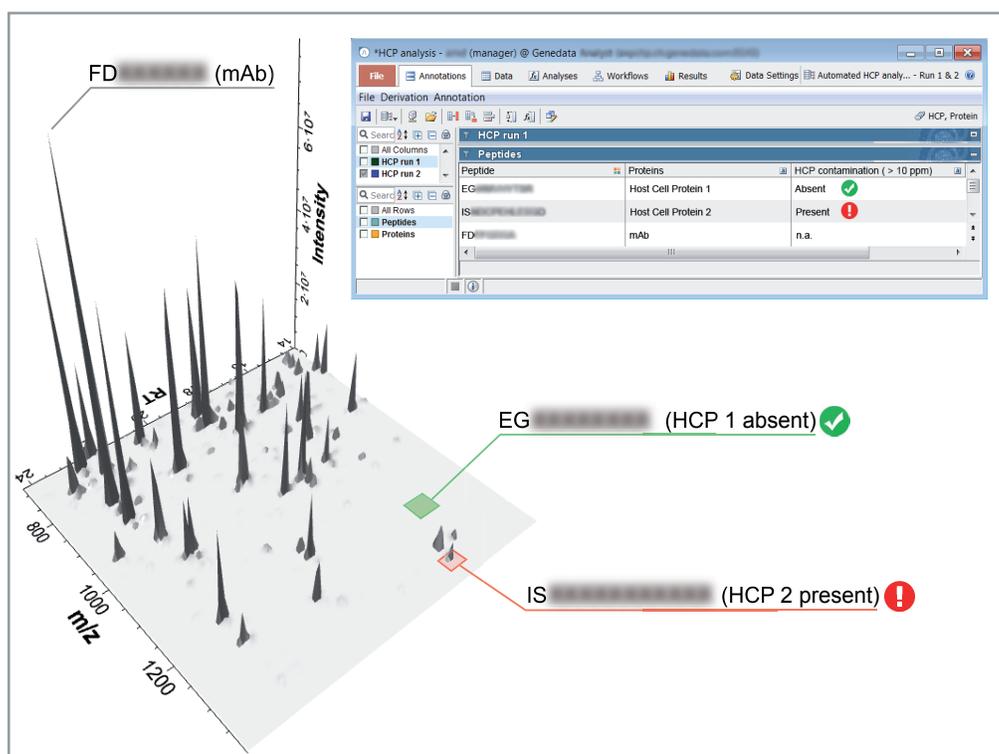

Figure 3: *Zoomed-in view of peaks contributing to the antibody and several HCPs. Indicated are regions where known impurities are expected. Table of peptides of the antibody and some HCPs together with summarized information about their abundance is also shown.*

Methods developed and validated previously can easily be shared across departments and functions. Tedious procedures that normally consume significant time and effort of experienced personnel, such as spectra analysis, peak annotation, sample comparison, and results reporting are seamlessly executed and summarized in electronically signed reports. Users granted manager roles can lock down parameter settings in approved workflows that can be executed in a reproducible manner with a simple one-click operation. Instrument suitability tests can be added to these workflows to ensure compliant, high-quality data, even from complex MS instrumentation. These capabilities provide considerable time savings in routine tests and enable streamlined data analysis in laboratories working in GxP environments.

Summary

From discovery to production, Genedata Expressionist offers an instrument- and vendor-independent software platform for the identification, quantification, and monitoring of host cell proteins. Refined signal pre-processing enhances protein quantitation of low abundance proteins where robust detection is critical. Workflow-based data analysis uniquely supports a wide range of HCP analytical strategies and enables reliable batch comparisons. Further, reproducible results combined with flexible reporting support the submission of regulatory dossiers.

Genedata
solutions in silico