TP 575 Automated Data Analysis Workflow Leveraging PASEF Data Accelerates Confident Detection of Low-Abundance HCPs

Amy Claydon[1]; Guillaume Tremintin[2]; Juergen Kastler[3]; Stuart Pengelley[4]; Jonathan Krieger[2]; Arnd Brandenburg[3]

Ltd, Cambridge, United Kingdom; [2]Bruker, San Jose, CA; [3]Genedata AG, Basel, Switzerland; [4]Bruker Daltonics, Bremen, Germany

Introduction

Host cell proteins (HCPs) present a potential safety risk in biopharmaceutical products. In HCP analysis, LC-MS/MS is often used as an orthogonal method to provide information complementary to conventional assays, and advances in MS instrumentation have led to increasingly sensitive detection. For example, trapped ion mobility spectrometry (TIMS) instruments leverage the additional dimension of ion mobility to optimize duty cycle and acquisition with the PASEF scan mode, resulting in improved identification rates in complex proteomics samples, and increased sensitivity for detection of low-abundance HCPs. However, due to the complexity of the generated data, labs require an efficient data processing, analysis, and reporting platform to keep pace with throughput demands and prevent data handling from becoming an analytical bottleneck.

Methods

We present an automated MS data processing workflow specifically configured to streamline analysis of the large volumes of data from TIMS-TOF-PASEF experiments and increase confidence in HCP identifications. Samples of NIST antibody were prepared following the native digest protocol described by Huang *et al*. Digests were separated by reverse phase on a nanoElute LC system (Bruker, Bremen) and data was acquired using a timsTOF Pro 2 mass spectrometer (Bruker) using PASEF with a 1.1 second duty cycle. Raw MS data was processed and analyzed using an automated Genedata Expressionist® software workflow (Genedata AG, Switzerland) consisting of sequential processing nodes (“activities”), which were configured to efficiently process and analyze batches of TIMS-TOF-PASEF data and deliver high-confidence HCP identifications.

Preliminary Data

The automated data processing workflow can be broken down into four sections: i) preprocessing, ii) processing of MS data, iii) processing of MS/MS data, and iv) reporting.After raw timsTOF data was loaded without conversion into the Expressionist software, data preprocessing activities were used to perform smoothing, chemical noise (background) subtraction, RT structure removal, and *m/z* structure removal. Optimizing these steps enabled efficient removal of noise and artefactual signals and a dramatic reduction (> 95%) in the size of MS data files, while preserving all relevant signals and information.For MS data processing, workflow activities included automatic peak detection and isotope clustering. Any singleton peaks that could not be isotopically clustered were removed, as it could be assumed that they did

not arise from genuine molecular signals. This filtering also reduced the overall amount of data, further accelerating and facilitating downstream processing.MS/MS data was processed to remove spectra that did not originate from genuine precursor MS signals, reducing the risk of false-positive identifications. A subsequent deisotoping step condensed MS/MS spectra to singly charged monoisotopic peaks.In the final step, the selected MS/MS spectra were submitted to the MASCOT search engine for protein identification. The results of the database search are automatically passed directly back to Genedata Expressionist to group associated proteins and filter shared peptides for final protein inference. In total, over 180 NISTmAb HCPs — including around 90 with two or more significant peptides and fifteen previously reported to be present at levels less than 1 ppm — were identified. The combination of highly sensitive MS acquisition with robust, automated data analysis delivers increased confidence in identification of low-abundance HCPs and significantly reduces the time needed for the critical data processing stage during biopharmaceutical product development and manufacture.

Novel Aspect

Automated MS data workflow for highly sensitive detection and confident identification of low-abundance HCPs with minimal false discovery rates.

Conflict of Interest Disclosure

The authors declare no competing financial interest.