



*Genedata Expressionist provides high-quality qualitative and quantitative results and its flexible workflow-based approach makes it the ideal platform for processing and analyzing mass spectrometry data generated using our innovative methods.*

Lucio Manzi, Ph.D., Associate Researcher — Merck Group, Guidonia, Italy

#### AUTHORS

L. Manzi, M. Sassi,  
M. Melchiorre, A. Palmese,  
and L. Colarusso  
Pharmaceutical &  
Analytical Development  
Biotech Products, Merck  
Serono, Guidonia, Italy

#### INDUSTRY

Biopharmaceuticals

#### CUSTOMER SINCE

2016

#### ABOUT MERCK

Merck is a leading science and technology company that operates across healthcare, life science, and performance materials.

#### GENEDATA SOLUTION



**EXPRESSIONIST**

# Enabling Novel MS-Based Methods for Characterization of Biopharmaceuticals

## Background

The Merck Serono structural characterization laboratory in Guidonia is part of the Characterization and Innovative Analytics Unit of the Analytical Development Biotech Department. Our role is to advise and support Merck worldwide in the development and production of novel biopharmaceuticals by providing a comprehensive structural overview of drug molecules and supporting process development and validation through advanced characterization approaches. The ability to deliver information-rich data at a molecular level makes mass spectrometry (MS) the technology of choice for applications such as (glyco)peptide mapping, disulfide-bridge and intact-mass analysis, top-down sequencing, and released glycan analysis. Our MS facility uses instruments from several different manufacturers, which are used on a case-by-case basis to perform a wide range of MS experiments. This case study presents the development and optimization of MS-based methods for charge-variant analysis (CVA) and glycosylation profiling and subsequent data processing using Genedata Expressionist, and describes the advantages over assays that had previously been performed using other technologies and data-processing approaches.



## Main Challenges

### Developing novel and efficient MS-based characterization methods

Optimization of sample prep and separation procedures was the first challenge we addressed when developing novel MS-based analytical methods. However, to bring the overall efficiency of the end-to-end analytical process to the level required at Merck, we also required a platform that streamlined MS data processing, analysis, and management.

### Harmonizing analysis and preventing bottlenecks

Acquiring data with instruments from different manufacturers means dealing with multiple software packages that create data silos and hinder comparison of results. In addition, the analysis of large volumes of complex data typically generated during MS experiments represents a serious bottleneck when using non-scalable desktop applications.

### Increasing data quality and confidence in results

The large volumes of information-rich data generated during MS experiments require expert interpretation and dedicated application-specific algorithms to extract the maximum amount of information and deliver the most comprehensive characterization results possible. Where possible, error-prone manual evaluation (e.g., peak boundary determination) should be replaced by unbiased algorithms to improve result accuracy.

### Breaking down data barriers and improving knowledge sharing

After data has been analyzed, the knowledge obtained should be freely accessible to all stakeholders to improve decision-making.

Presenting complex information in a form that is quickly and readily understood is vital to the dissemination of knowledge within and across labs.

## Solution

### Optimized sample preparation and flexible, automated data processing

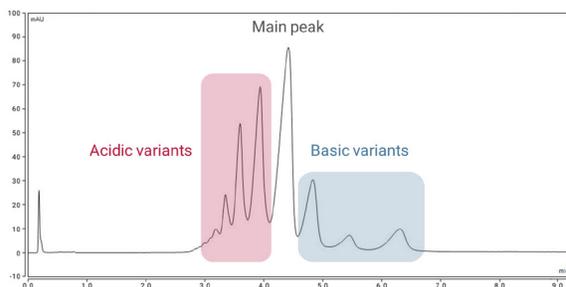
We developed novel sample preparation methods that used alternative chromatographic matrices, MS-compatible buffers, and reagents that generate analytes that are amenable to efficient MS analysis. During method development, the flexibility of Genedata Expressionist enabled us to comprehensively analyze results and iteratively optimize procedures. After method optimization was complete, the workflow-based system enabled us to automate data analysis, further streamlining the overall analytical process.

### Vendor-neutral, highly scalable data analytical system

The ability of Genedata Expressionist to load data from any MS instrument enables us to process data from all our instruments using a single platform. As a centralized server-based enterprise solution, Genedata Expressionist scales to the needs of large organizations and allows us to quickly and efficiently analyze the large amounts of MS data that we generate.

### Streamlined MS analysis using custom libraries

In CVA, we use the flexibility of Genedata Expressionist to perform two approaches within a single data processing workflow. In a targeted approach, custom molecular libraries containing entries corresponding to commonly encountered



Mass	RT	Modifications	Glycosylation	Mass (1)	RT (1)	Mass Delta [ppm]	SCX_Erbitux_25ug
153100.6	3.063	2*Gln->pyro-Glu + 2*Lys-loss	FA2 + FA2G1 + 2*FA2G2Sg2	153095.3	3.06	34.68	28.31
152950.8	3.408	2*Gln->pyro-Glu + 2*Lys-loss	FA2 + FA2G1 + FA2G2Ga1Sg1 + FA2G2Sg2	152950.2	3.41	3.97	19.83
152807.0	3.650	2*Gln->pyro-Glu + 2*Lys-loss	FA2 + FA2G1 + 2*FA2G2Ga1Sg1 : FA2 + FA2G1 + FA2G2Ga2 + FA2G2Sg2	152805.1	3.65	12.33	10.57
152659.6	3.974	2*Gln->pyro-Glu + 2*Lys-loss	FA2 + FA2G1 + FA2G2Ga1Sg1 + FA2G2Ga2	152660.0	3.97	-2.72	9.59
152517.6	4.446	2*Gln->pyro-Glu + 2*Lys-loss	FA2 + FA2G1 + 2*FA2G2Ga2	152514.8	4.45	18.33	17.48
152642.2	4.888	2*Gln->pyro-Glu + Lys-loss	FA2 + FA2G1 + 2*FA2G2Ga2	152643.0	4.89	-5.27	7.48
152911.3	5.525	2*Gln->pyro-Glu	FA2 + FA2G1 + FA2G2Ga1Sg1 + FA2G2Ga2	152916.3	5.53	-32.63	6.55
152770.7	6.398	2*Gln->pyro-Glu	FA2 + FA2G1 + 2*FA2G2Ga2	152771.2	6.40	-3.54	0.15

species are used to provide an overview of highly abundant, expected variants across the entire elution profile (Figure 1). In an untargeted approach, more extensive libraries are used to fully characterize timeresolved elution ranges. Together with intuitive and insightful visualizations, the ability to search sample-specific custom libraries and perform advanced iterative review streamlines the analysis of released glycan and glycopeptide mapping results.

### Enterprise software with advanced reporting

We can fully integrate Genedata Expressionist into our existing data systems to create a knowledge database that allows existing product knowledge (for example, the nature of variants found in specific clusters) to be leveraged in subsequent analyses. Structured, highly configurable reports and visualizations provide at-a-glance interpretation that speeds up analysis and aids decision-making. For example, an XIC offset view of Fab subunit glycoforms demonstrates that they elute in order of increasing complexity, which aids their identification (Figure 2).

## Benefits

### Deeper characterization

Pairing the novel methods with an automated data analysis platform enabled us to maximize our labs' productivity and provided key insights into charge-variant and glycan analysis of biopharmaceuticals. For example, in middle-down glycosylation profiling, the novel sample preparation delivered a significant 46% increase in the number of identified glycoforms.

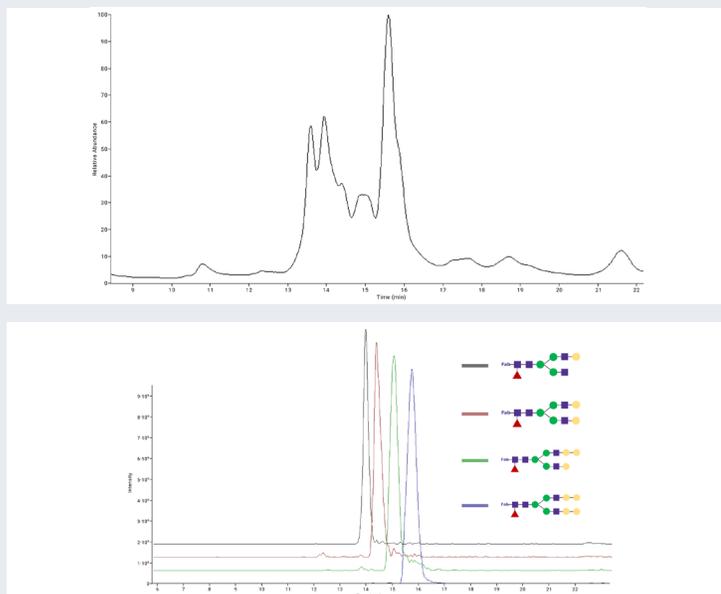
Using Genedata Expressionist for data analysis further boosted the total number of detected glycoforms with a striking 2.5-fold increase with respect to the conventional sample preparation and data processing methods—and minimized false-positives by using a custom-built glycan library.

### Harmonized processes and significant time savings

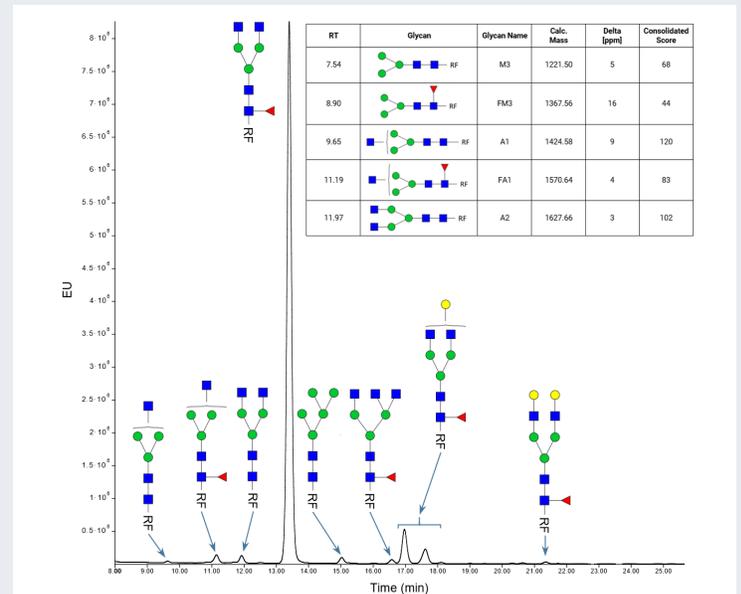
As a single-software solution for all MS-based applications, Genedata Expressionist breaks down data barriers and enables us to compare results from all our instruments using harmonized procedures. Because Genedata works closely together with leading instrument manufacturers, we can be sure that we can fully benefit from our investments in the latest MS technologies. Cutting-edge algorithms and the highly configurable workflow nature of Genedata Expressionist—which enables automated data processing and reporting—reduce the time required for data analysis by up to 80% compared to previously used methods (Figure 4). Automatic, unbiased integration and quantification eliminates the need for manual evaluation of signals.

### Higher-quality results

Using Genedata Expressionist not only provided significant time-savings but also increased the quality of results in all our analyses. Targeted MS analysis using custom libraries and consolidated scores provided high-confidence results and significantly reduced the number of false-positive identifications in CVA and glycosylation profiling following HILIC-MS analysis. In released glycan analysis—after their structure was assigned and confirmed by MS/MS analysis—glycans were automatically quantified based on their fluorescence detector signal, increasing the overall specificity of the method.



2 Upper panel: HILIC-MS TIC profile of WS 2014/01 digested with FabALACTICA; Lower panel: Extracted ion chromatogram profiles of Fab subunit glycoforms of increasing complexity



3 Annotated FLR trace; Inset: Typical report table listing identified released glycans with consolidated MS/MS score.

## Efficient dissemination and leverage of knowledge

Reports generated using Genedata Expressionist can be configured by adding or removing individual elements to provide as much or as little detail as required. An overview report can be generated for quick comparison of samples within a batch, or a complete report detailing all results and data processing parameters can be created for analyses in regulated environments.

Curated results can be quickly leveraged to optimize subsequent analyses. For example, the custom-built library that reduced falsepositives in middle-down glycosylation profiling was derived from released glycans identified in previous experiments.

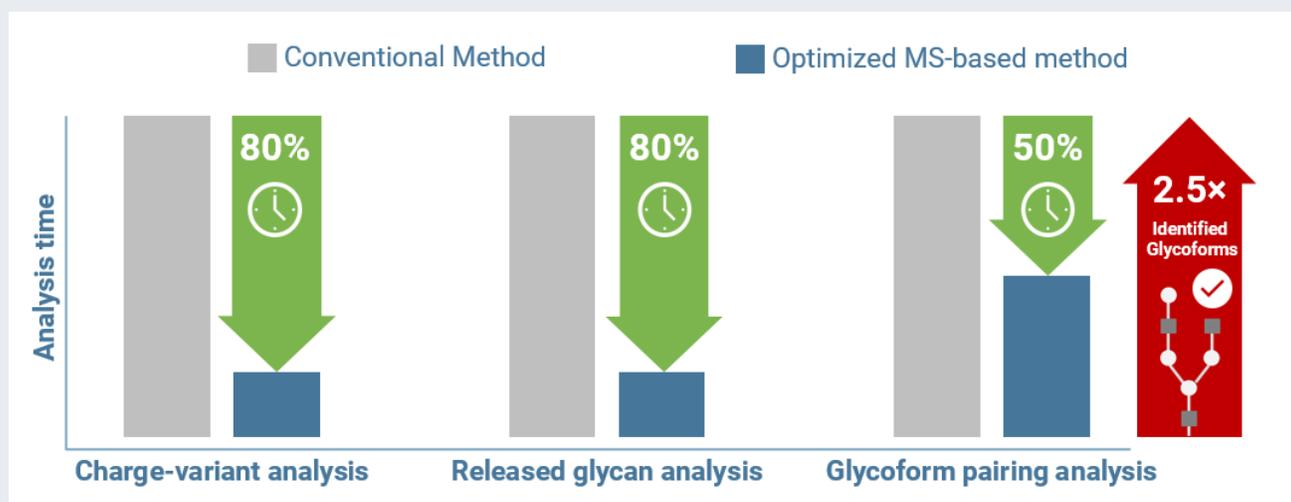
In a similar fashion, a subset of modifications found in a charge variant cluster were saved as a custom library to reduce combinatorial complexity and data processing time when analyzing corresponding clusters in subsequent experiments.

## Summary

Genedata Expressionist enabled us to rapidly develop novel methods and automated data workflows that were specifically customized and optimized for our novel MS-based biopharmaceutical characterization approaches. Utilizing a scalable platform allows us to rapidly process the large volumes of complex data that our labs generate—thereby increasing our overall productivity—and automated reporting facilitates collaboration and knowledge sharing across our organization.

## Outlook

We are in the process of establishing a Good Manufacturing Practice (GMP) environment at our site. The compliance capabilities of Genedata Expressionist—which include detailed client and server logs, result review and approval, stringently defined projectspecific user/role profiles, and the ability to “lock down” optimized Genedata Expressionist workflows—will facilitate harmonized and reproducible MS data processing.



<sup>4</sup> Time savings and information gains provided by novel protocols and Genedata Expressionist data processing.

**Automated data processing and report generation provide remarkable time-savings in all our MS-based biopharmaceutical characterization applications.**

Mauro Sassi, Ph.D., Associate Researcher — Merck Group, Guidonia, Italy

### GENEDATA SOLUTION



Genedata Expressionist® is part of the Genedata portfolio of advanced software solutions that serve the evolving needs of drug discovery, industrial biotechnology, and other life sciences.

© 2022 Genedata AG. All rights reserved. Genedata Expressionist is a registered trademark of Genedata AG. All other product and service names mentioned are the trademarks of their respective companies. 22E02