An Automated MS Data Workflow Enabling Targeted, Site-Specific N-Glycosylation Monitoring of Biopharmaceuticals

Diego Bertaccini, PhD
Senior Scientist, Biotech Process Sciences Analytics
Merck Serono SA, Switzerland

Pauline Ulrich, PharmD
Associate Scientist, Biotech Process Sciences Analytics
Merck Serono SA, Switzerland

Sponsored by

Genedata
solutions in silico

BACKGROUND

With sales of antibody-based therapeutics totaling around $100 billion in the U.S. alone, monoclonal antibodies (mAbs) are currently the largest and fastest-growing class of biopharmaceutical. All therapeutic mAbs expressed in mammalian expression systems undergo post-translational modifications (PTMs), which may include the addition or replacement of functional groups, or higher-level structural changes like folding, cleavage, or racemization.

N-GLYCOSYLATION

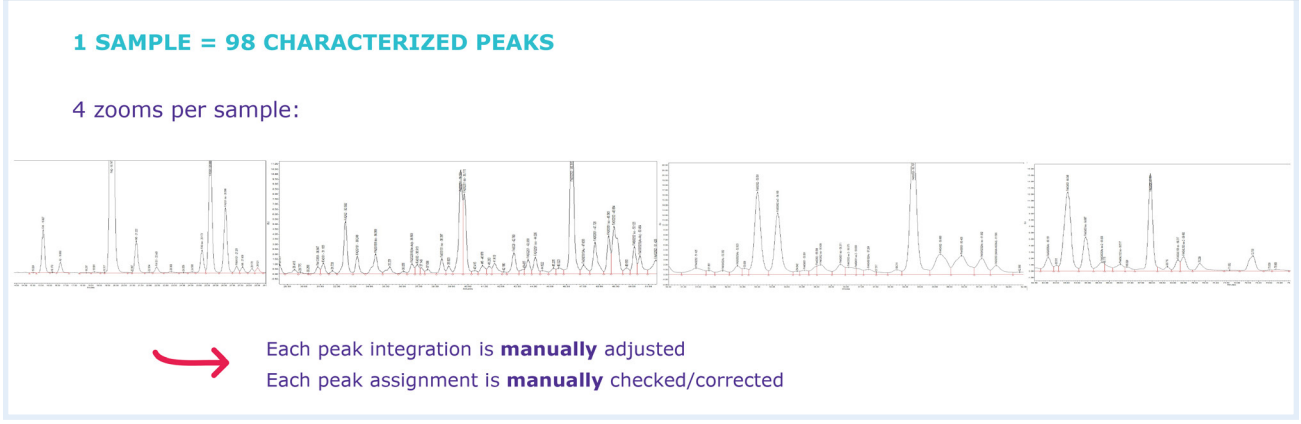
N-linked glycosylation is a form of PTM that describes the attachment of N-acetylglucosamine (GlcNAc) to the nitrogen atom of an asparagine side chain via a β-1N linkage. N-glycosylation modulates antibody activity by enhancing antagonist (blocking, inhibiting) or agonist (activation) functions. Glycosylation may affect a therapeutic antibody's immunogenicity, potentially increasing risks for patients. Pharmacokinetics, especially clearance and circulating half-life, are also affected by N-glycosylation.

In biopharmaceutical development, there is a need to understand how cellular conditions and biotechnical process parameters modulate N-glycosylation. Understanding these effects facilitates establishing robust manufacturing processes and ultimately safer therapeutics. To acquire such knowledge, hundreds of upstream conditions need to be tested for each cell line—a large undertaking—therefore robust and high-throughput glycan analysis methodologies are required.

N-GLYCAN ANALYSIS

The current gold standard for N-glycan mapping is ultrahigh-

Figure 1: Glycosylation pattern analysis: N-glycosylation monitoring by 2-AB UHPLC.



The limitations presented by the UHPLC method led us to examine general glycan-characterization methods that were more tolerant to co-elution issues.

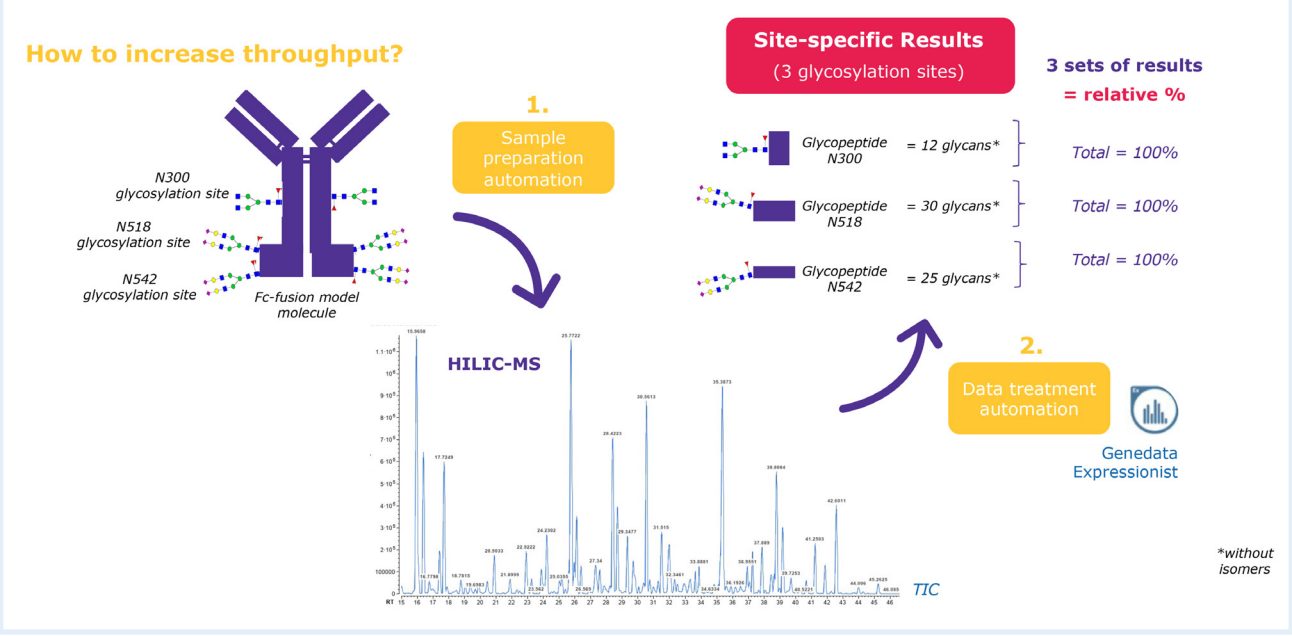
pressure liquid chromatography (UHPLC) with a fluorescence-based detection method that informs about a mAb's global or total glycosylation status. Merck KGaA employs this method of analysis, which first involves release of N-glycans using an enzyme (N-glycanase), followed by labeling of the released N-glycans with a reagent (2-aminobenzamide or 2-AB), followed by UHPLC analysis. Although this method is highly reproducible, it presents several limitations that are undesirable in our laboratory.

Firstly, because all N-glycans are released indiscriminately, site-specific information for molecules with multiple glycosylation sites is entirely lost through this method. Secondly—and even with extended gradients—coelution and new peak appearance, when present, provides identification complexity. Finally, data processing, analysis, and reporting are non-trivial and can be a laborious undertaking.

FIGURE 1 shows a typical analysis of an N-glycosylated Fc fusion molecule by this 2-AB UHPLC method, which uncovered 98 chromatographic peaks. Integrating peaks required that we "cut" the chromatograph into four sections and manually examine and assign each peak manually, a process that takes several days. And, since this study was supporting development work, we needed to analyze for glyco-variants—further increasing the analyst's task. We found that retention time alone is insufficient for properly characterizing known species, let alone for identifying new peaks.

The limitations presented by the UHPLC method led us to examine general glycan-characterization methods that were more tolerant to co-elution issues, which allowed direct identification of new peaks, and that provided the speed and throughput required

Figure 2: LC-MS glycopeptide mapping: method overview.



for modern glycoproteomic workflows. We sought an analysis that gave site-specific versus global glycosylation information, and the ability to individually monitor glycan repartition. That method is glycopeptide mapping by LC-MS.

FIGURE 2 illustrates LC-MS glycopeptide mapping. The molecule under investigation has three N-glycosylation sites N300, N518, and N582, conjugated to 12, 30, and 25 glycans, respectively. After sample preparation and analysis by HILIC liquid chromatography-mass spectrometry, the relative abundance of the different glycans at each glycosylation site is calculated in percentages such that each site adds up to 100%.

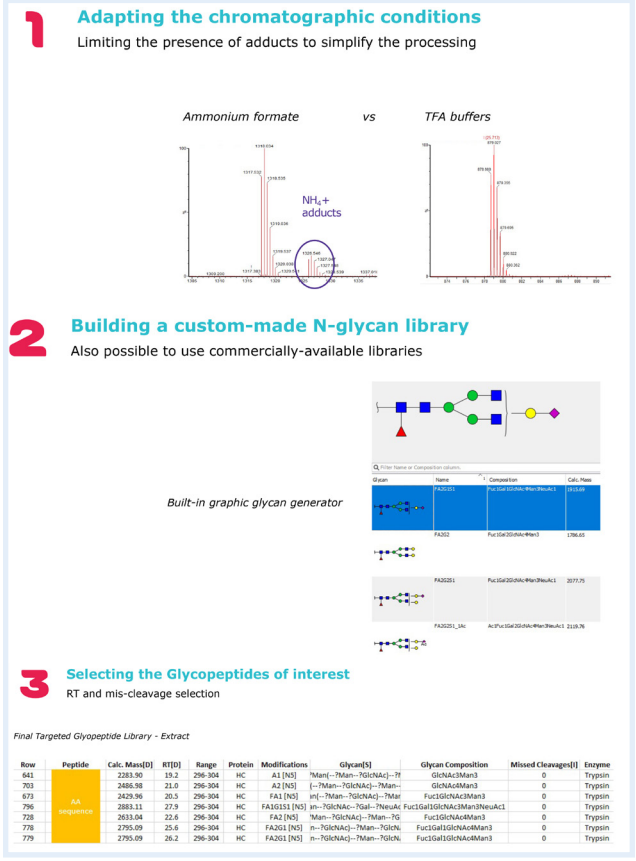
Now, with mass-to-charge information supplementing retention time, new peaks can be identified directly, and co-elution becomes a non-issue. The method is also fully compatible with complex glycoproteins that do not fully de-glycosylate.

As mentioned, processing the data from such an experiment typically requires manual processing, which is time-consuming and uses a workflow that was created for just a few samples and requires MS expertise. On that basis, we decided to streamline and, it was hoped, make the method more robust and accessible.

The strategy we identified for streamlining this data processing workflow involved automating the processing steps using Genedata Expressionist.

We know that MS experiments of this kind generate a lot of complex multidimensional data from multiple isomers and charge states. The presence of adducts will only add to the complexity. Data inspection using Genedata Expressionist allowed us to detect the presence of ammonium adducts as a result of the addition of ammonium formate in the mobile phase. Consequently, we decided to use trifluoroacetic

Figure 3: LC-MS glycopeptide mapping - Data treatment automation.



acid (TFA) as the mobile phase modifier which led to slight signal suppression but eliminated adducts and simplified MS data processing (FIGURE 3).

The next step involved populating the MS detection library with glycans of interest. Commercially available libraries suffice for the analysis of most simple IgG antibodies. Custom library entries will be needed if detection involves more-complex species such as acetylated sialic acids. Genedata Expressionist features built-in functionality for entering custom glycan structures and customizing your own glycan libraries. Once characterization work is done on the molecule and the glycopeptides of interest are identified, the analyst is ready to select

Genedata Expressionist has built-in functionality for adding custom glycan structures and customizing glycan libraries.

glycopeptides to use as the reference library for subsequent upstream experiments on the same cell line, provided identical analysis parameters are followed. When new peaks are of interest, a Genedata Expressionist workflow can be created to identify new peaks above a certain intensity threshold compared to a reference sample. In the presence of MS/MS data, new peaks can be investigated using a peptide mapping search algorithm as a first step followed by a Wildcard search to look for unexpected modifications.

Genedata Expressionist workflows are a succession of data processing steps, or "activities," which users can add, remove, optimize, or re-order as needed. The software allows parallel evaluation of multiple detection settings, for example varying levels of noise reduction, which aids in the optimization of the workflow. Once analysis parameters are optimized and library entries are set, fully automated batch processing is possible. FIGURE 4 illustrates the journey from the acquisition of relatively uninformative raw data to species detection, which is the point of the study. Genedata Expressionist selects each targeted glycopeptide, automatically reports its mass-to-charge ratio and retention time, directly groups isomers (including newly discovered ones), and can visualize data at

Figure 4: Fully customizable workflow.

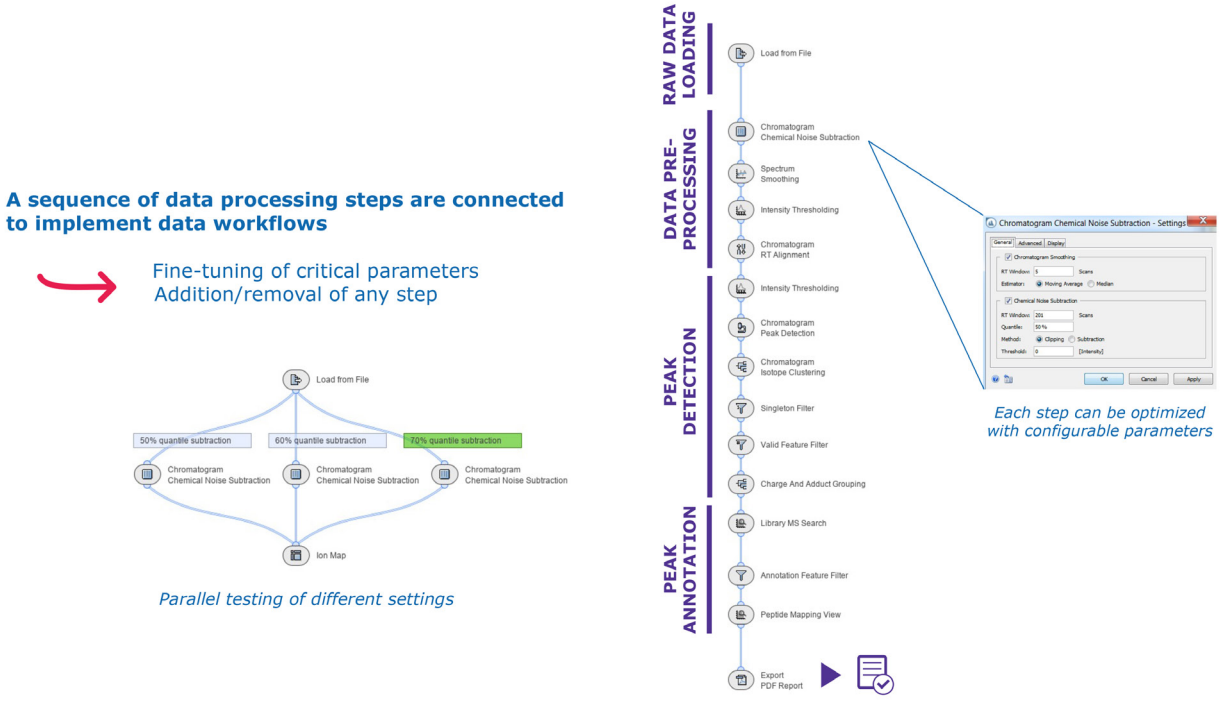


Figure 5: Automated workflows for routine use.

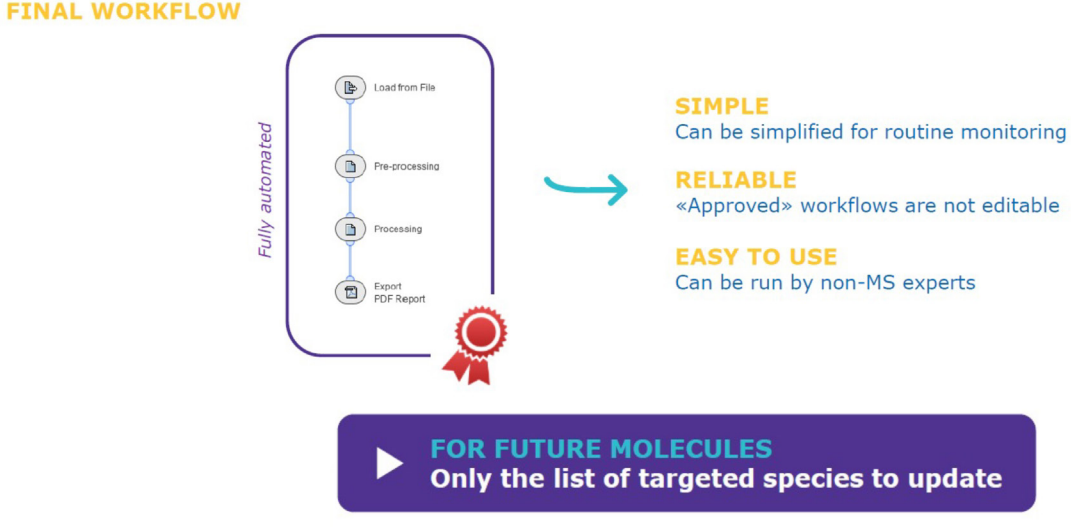
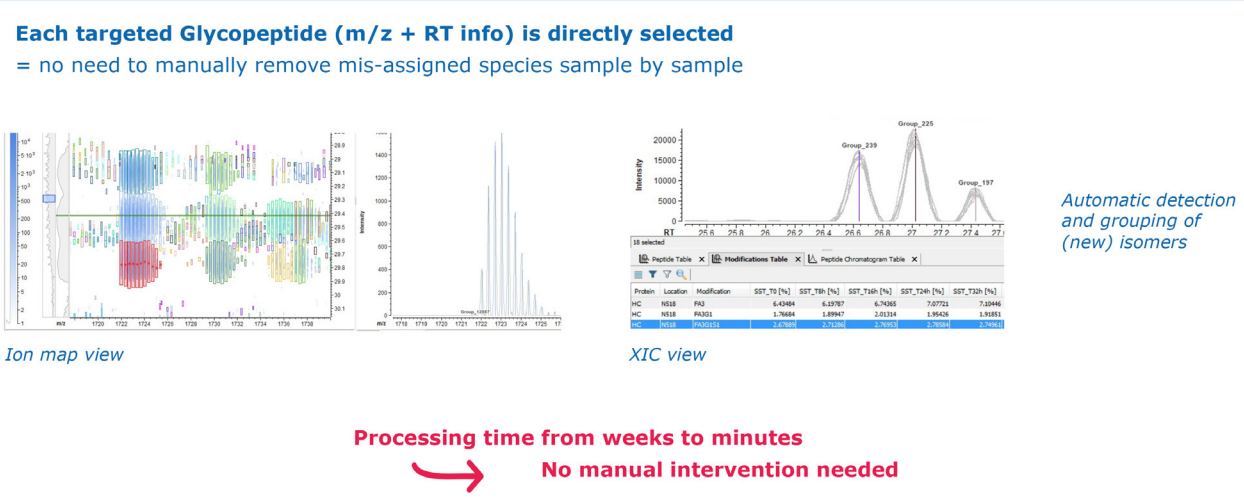


Figure 6: Reduced processing time with automated batch processing.



any point within the workflow. Users may also create "approved" workflows (FIGURE 5) by blocking edits to selected parameters or the entire workflow—simplifying usage in routine monitoring and allowing the workflows to be used by non-experts.

Together, these improvements cut data processing time for a typical N-glycan experiment from weeks to minutes, in a fully automated manner completely free of any risks due to human input (FIGURE 6).

HOW LC-MS COMPARES

FIGURE 7 reports on results from a preliminary experiment in which we compared the 2-AB UHPLC method and our revised LC-MS glycopeptide mapping method for glycosylation at asparagine 300. Because the UHPLC method quantifies the totality of glycans whereas mapping separates these moieties into three sets of results, we can only investigate trending. As we can see in the figure, glycopeptide N300 (light blue) follows the same trend as 2-AB UHPLC results while glycopeptide N518 remains stable, providing site-specific information.

FIGURE 8 compares results obtained at our characterization lab (in red) with the high-throughput workflow described earlier (blue bars). This experiment looks at various glycans (monosialylated, fucosylated, etc.) attached to N518. Differences between the heights of

Figure 7: LC-MS Glycopeptide mapping vs 2-AB UHPLC.

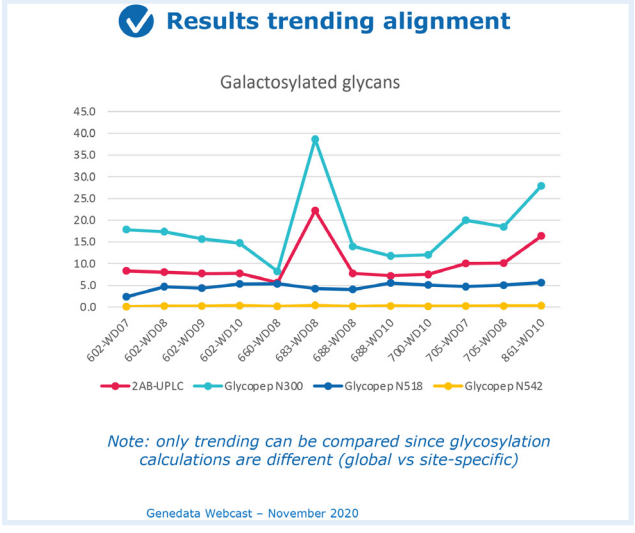
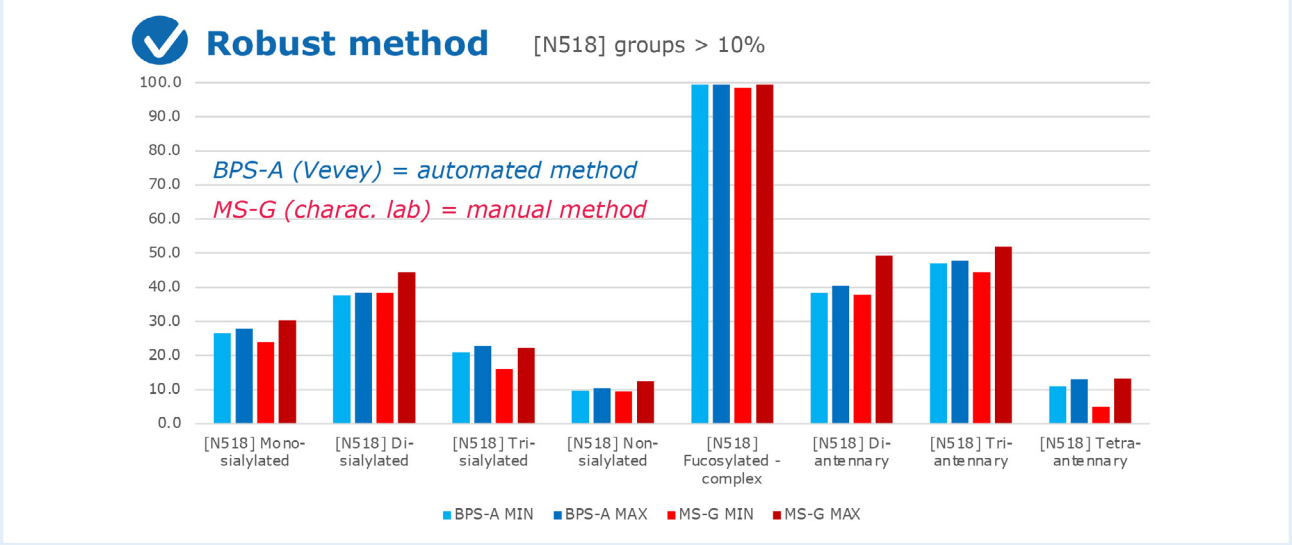


Figure 8: Glycopeptide mapping vs reference characterization method.



light and dark colored bars represent readout variability. As we can see, agreement is very high between the standard and high-throughput method, but the lower variability for the latter suggests greater reproducibility of results.

CONCLUSION

Biopharmaceutical developers are constantly searching for strategies that improve the quality and effectiveness of their products. Characterizing glycosylation, and ultimately controlling glycosylation patterns, is one way to achieve that goal.

The Genedata Expressionist automated workflow we describe for mapping N-glycosylation sites provides site-specific monitoring with no co-elution issues using LC-MS. Suitable for complex molecules, this powerful analytic platform identifies new peaks directly, and is also suitable for monitoring O-glycosylation. Moreover, this accessible, nonexpert method provides increased throughput from 20 to 80 samples per week with minimal operator input.

This accessible, non-expert method provided increased throughput from 20 to 80 samples per week with minimal operator input.

Genedata Expressionist is comprehensive enterprise software for processing, analyzing, and reporting mass spectrometry data in application areas such as biotherapeutics characterization, proteomics, and metabolomics. Built on open, flexible client-server architecture, Genedata Expressionist accepts and processes large, complex experimental data sets from all major vendors and technology platforms, and easily integrates into existing R&D information processing environments.