Case Study

Efficient, automated data processing for a novel MS-based high-throughput screening platform

Dr Jonathan Wingfield, Mechanistic Biology and Profiling, Discovery Sciences, AstraZeneca R&D, Cambridge, UK

About AstraZeneca

AstraZeneca is a global, science-led biopharmaceutical company whose innovative medicines are used by millions of patients worldwide.

Industry

Biopharmaceuticals

Website

www.astrazeneca.com

Key Challenges

Automate data processing and increase quality of results

Solution

Automated MS data processing using Genedata Expressionist®

Results

Seven-fold decrease in processing time with elimination of potentially error-inducing manual interactions

Background

For more than a decade, Acoustic Droplet Ejection (ADE) has been used in high-throughput screening (HTS) to transfer low volumes of fluid—typically nanoliters or picoliters—between multiwell plates.

The advantages of ADE include the elimination of carryover associated with needle-based transfer and high transfer speeds. More recently, ADE technology has been developed to introduce samples into a mass spectrometer (Figure 1). The increase in throughput that Acoustic Mist Ionization Mass Spectrometry (AMI-MS) technology[1] brings in comparison to conventional UHPLC-MS—from one sample per minute to three samples per second—brings MS analysis into the realm of true high-throughput screening of 100K samples per 24 h day (Table 1).

Genedata worked together with scientists at AstraZeneca to develop a flexible, automated data processing workflow for efficiently processing the large amounts of data generated by this new screening platform.

Technique	Single sample	384-well plate	50,000-compound screen
HPLC-MS	5 minutes	1.33 days	175 days
UHPLC-MS	1 minute	6.4 hours	35 days
Auto SPE-MS	10 seconds	64 minutes	6 days
AMI-MS	0.3 seconds	< 3 minutes	4.6 hours

Table 1: Typical sampling times required by current MS techniques. AMI-MS screening is >120 times faster than UHPLC-MS and >20 times faster than Automatic Solid-Phase Extraction-MS (Auto SPE-MS).
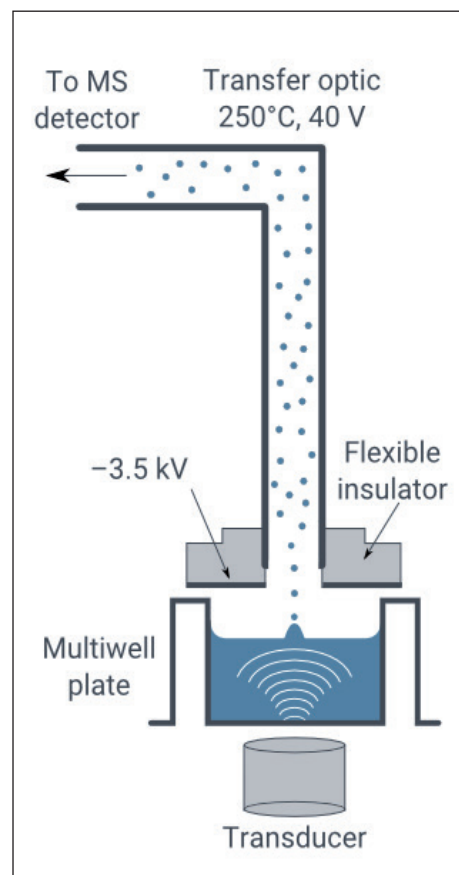

Figure 1: AMI-MS principle

Main Challenges

Eliminating laborious and potentially error-inducing manual interventions

AMI-MS data from multiple 384-well plates is recorded as a single raw data file. The first step in data processing is to extract the chromatograms from each well on each plate, after which, the spectra are assigned to the wells. While this process can be carried out by the MS instrument's software, it can be a slow process. The final step of identifying target peaks and calculating their areas requires manual interventions including scanning plate barcodes by hand: a laborious and error-prone process.

Guaranteeing data quality and traceability

Peak ratios are obtained by integration of peaks using manually entered target masses and predefined *m/z* windows. Using this approach, the values obtained are dependent on peak shape and multiplicity and the accuracy of the manually entered target mass. Variations and inconsistencies in these factors can negatively impact data quality. In addition, screening data may need to be interrogated in the future, and therefore a transparent, comprehensive data audit is required.

Adapting data processing to different targets and screening strategies

The richness of MS data in comparison to photometric assay data enables novel screening strategies in which, for example, multiple molecular species are measured in a single sample. In addition, the technique is applicable to any biological matrix. Ideally, the platform used to analyze AMI-MS data would be flexible enough to implement any assay strategy.

Obtaining insightful visualization to speed assay development

Examining spectra can greatly facilitate assay development and triaging of false-positives. The ideal platform would offer the opportunity to inspect spectra at every stage of processing.

Integration of MS data into screening data analysis platforms

For most compound screens, interaction with an MS-data processing software may not be required or desired. Ideally, in such cases, results would be automatically passed to a data analysis platform—such as the Genedata Screener platform in use at AstraZeneca—which enables quality control (QC) through statistical tools and direct visualization for detailed analysis of potential hits.

The Solution

Intelligent automated data processing

Genedata Expressionist data processing is based on automated, customizable workflows, in which data is subjected to sequential processing steps. Figure 2 shows an exemplary workflow for analysis of AMI-MS data. When raw AMI-MS data is loaded into Genedata Expressionist, it is automatically parsed and spectra are assigned to their respective 384-well plate well, making the manual scanning of barcodes unnecessary. In the next step, the data set is examined and spectra that do not meet user-definable quality criteria are reported and excluded from further processing.

Advanced algorithms ensuring highest data quality

Peaks of interest (for example, substrate and product peaks) are automatically quantitated using dynamic peak detection and clustering algorithms; thereby eliminating the variability introduced by manual entry of target masses and fixed *m/z* windows. To simplify data analysis, all peaks except those of interest are filtered out before the data is exported.
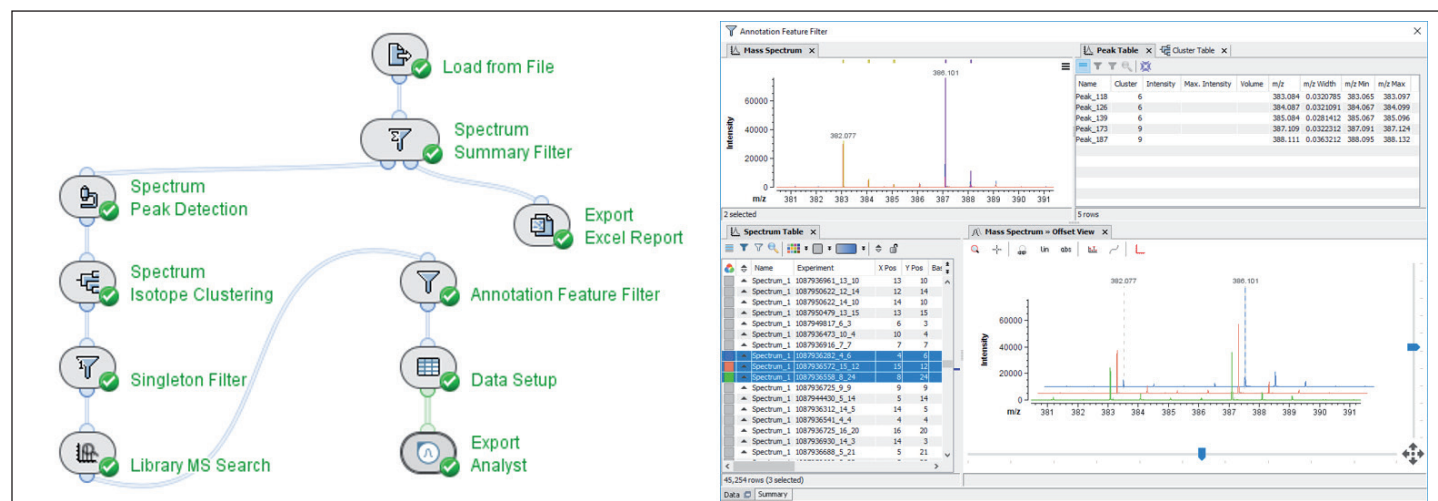

Figure 2: Genedata Expressionist workflow for processing AMI-MS data (left panel) and data inspection view in Genedata Expressionist software (right panel)

Lockable workflows ensuring consistent and reproducible processing

Each step of Genedata Expressionist workflows can be controlled by user-definable settings, offering an unparalleled level of control during MS data processing. When processing parameters have been optimized, workflows can be locked; ensuring reliable, reproducible, and consistent analysis across data sets.

Configurable workflows adaptable to any screening strategy

Typically, a screening assay tests the effect of adding sample compounds to a target molecule. In the case of enzymatic assays, these effects are usually assessed by measuring the concentration of the target enzyme's substrate and/or product. The exemplary workflow in Figure 2 implements exactly this strategy by measuring two peaks of interest, namely the substrate and product of the assay reaction. However, there is no limit to the number of features that could be investigated within the content-rich AMI-MS data sets. Such features can be imported from user-defined libraries, and using metadata-driven automated processing, the search library can be individually defined for each sample.

Integration into HTS data analysis platforms

At AstraZeneca, each new AMI-MS file is automatically uploaded into Genedata Expressionist and processed using a predefined workflow. The results are seamlessly exported to Genedata Screener, where plate-wise visualization of results as a heat map (Figure 3) enables routine QC and fast identification of potential hits.

Detailed visualizations adaptable to users' needs

Using Genedata Expressionist, data and spectra can be viewed at every stage of the workflow, which greatly facilitates quality control, assay development and optimization, and troubleshooting. Personalizable user interface layouts can be saved and applied to any data set, enabling users to configure data displays according to their individual needs.

Benefits

Significant time savings

Using Genedata Expressionist, the data processing workflow requires a single click and takes around 25 minutes for a data set obtained from 150 × 384-well plates. Previously, processing of such data sets—including the potential error-inducing manual steps—required 3 hours. Automated data processing and reporting significantly streamline the screening process and greatly increase productivity.

Higher-quality, more reliable data

By eliminating potential error-inducing manual steps, automated Genedata Expressionist data processing greatly increases the level of confidence in the quality of the data. Locking and sharing workflows ensures reliable, reproducible, and consistent analysis across data sets and between development labs in AstraZeneca.

Flexibility in assay design

Because it can be used for processing data from any MS instrument or application, Genedata Expressionist offers AstraZeneca's scientists the ability to employ any MS-based screening approach, and the opportunity to explore biological experimental spaces outside conventional substrate/product assay strategies.

Better understanding and decision-making

During assay development, detailed visualizations at every step of data processing facilitate in-depth analysis of results, enabling more convenient and faster data review and curation, faster triage of false-positives, and ultimately, better decision-making.
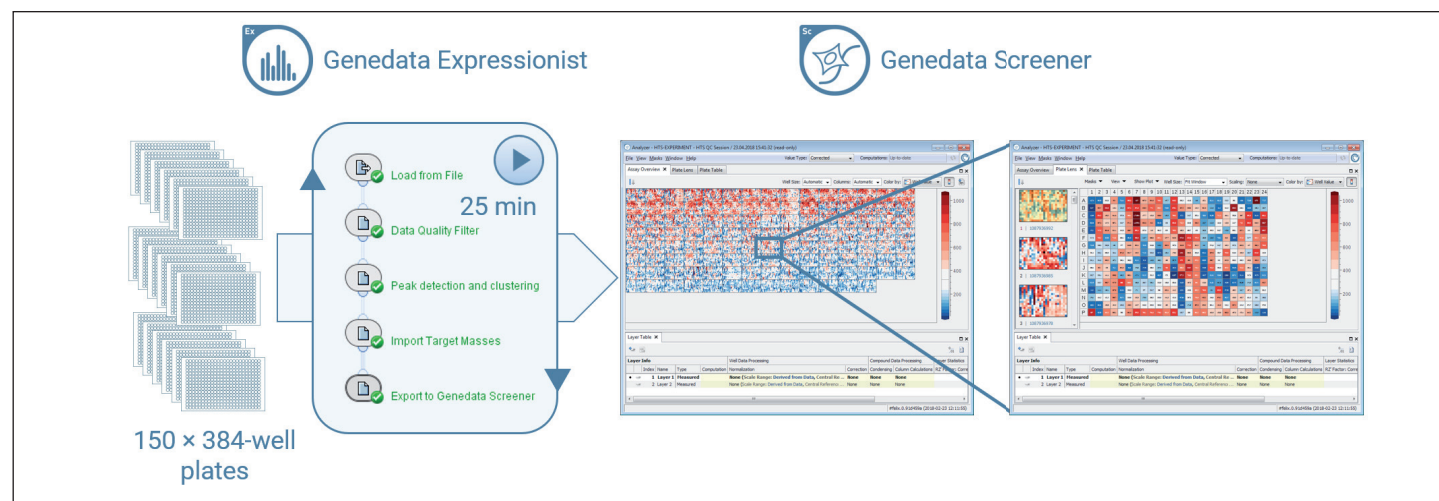

Figure 3: Automatic processing of AMI-MS data and transfer of results to Genedata Screener

Summary

The increase in sampling speeds provided by AMI-MS brings MS analysis into the realm of true high-throughput screening. The ability of AMI-MS to analyze any biological matrix and the availability of content-rich MS data opens previously unexplored spaces to HTS. Processing AMI-MS data using Genedata Expressionist ensures high-quality results that can be easily shared across organizations. This, and its inherent flexibility, make Genedata the software of choice for developing, performing, and benefitting from MS-based high-throughput screening.

Outlook

Genedata will continue to work closely with AstraZeneca to help leverage the enormous potential of AMI-MS in HTS applications. The AMI-MS project is designed to utilize MS—which offers a higher-quality, higher-throughput and lower-cost option—as an endpoint for biochemical high-throughput screening, thereby potentially generating better starting points for drug discovery projects at AstraZeneca.

"We are excited to be working with our long-time partner Genedata, who share our vision of automating HTS analysis workflows. The easy integration of high-quality, information-rich MS data from Genedata Expressionist into our existing Genedata Screener data analysis platform means that we can take full advantage of the increase in speed and data quality that AMI-MS provides." Jonathan Wingfield

References

1. Sinclair I, Stearns R, Pringle S, Wingfield J, et al. Novel Acoustic Loading of a Mass Spectrometer: Toward Next-Generation High-Throughput MS Screening. J Lab Autom. 2016;21(1):19-26.

About Genedata Expressionist
As a comprehensive software solution for transforming raw data from any mass spectrometry instrument into insights, Genedata Expressionist offers built-in functionalities that serve all MS applications for biopharmaceutical characterization, proteomics, and metabolomics in a single enterprise platform. Custom-built workflows address specific data processing, analysis, and reporting needs and enable harmonization and standardization between and across organizations. Complete automation, unbiased data interrogation with best-in-class algorithms and cutting-edge statistics, and intuitive visualizations deliver high-quality results with significant time and cost savings. Integrated data and project management enables organizations to streamline methods and efficiently manage data, results, and reports.

Genedata Expressionist is part of the Genedata Biopharma Platform for capturing, organizing, integrating, processing, and analyzing data to increase enterprise-wide productivity and R&D process efficiency.

For more information about Genedata Expressionist and other customer case studies, visit www.genedata.com/expressionist

Acknowledgement: Header image reproduced by kind permission of AstraZeneca.

Genedata transforms data into insights with innovative software solutions that support large-scale, experimental processes in life science research, and delivers enterprise solutions that streamline R&D workflows and improve research productivity.

Basel | Boston | London | Munich | San Francisco | Singapore | Tokyo
www.genedata.com | expressionist@genedata.com