



**BRUKER**

## Rapid Discrimination Analysis of Biological Samples by MALDI-TOF MS Using Clover Biosoft MSDA Software

Clover Biosoft MSDA software provides a comprehensive statistical toolbox for streamlined analysis of MALDI-TOF MS profiling data from biological samples featuring tailored biomarker analysis, classification, and identification workflows.

### Abstract

MALDI-TOF, due to its unparalleled analysis speed, is an ideal tool for rapid mass spectrometric profiling of a wide range of biological samples. Analysis of the resulting data by means of advanced statistical methods allows for discrimination, classification, or identification of samples based on either discrete  $m/z$  features, commonly referred to as biomarkers, or diagnostic fingerprints.

Here, we demonstrate the use of the Clover Biosoft Mass Spectrometry Data Analysis (MSDA) software for statistical analysis of MALDI-TOF MS profiling data in various applications. Clover MSDA provides a powerful statistical toolbox accessible via an intuitive, workflow-oriented user interface. As demonstrated here for MALDI-TOF datasets originating from various biological sample matrices, Clover MSDA software is capable of efficiently analyzing large cohorts of raw MS spectra with minimal effort, providing tremendous benefit to a wide range of biomarker discovery, classification, and identification tasks.

Keywords:  
Clover Biosoft MSDA software, data analysis, MALDI-TOF, mass spectrometry, profiling, classification, discrimination, identification, statistics

## Introduction

MALDI-TOF, as a direct mass spectrometry method, offers fast analysis speed, high sample throughput, and the capability to detect a wide range of biomolecules. MALDI-TOF, therefore, is an ideal tool for simple and fast profiling analysis, allowing for discrimination, classification, or identification of biological samples based on either individual diagnostic  $m/z$  features, commonly referred to as biomarkers, or characteristic mass spectral fingerprints.

However, unlocking the discriminative power of MALDI-TOF data, in particular when dealing with large sample cohorts, requires advanced software featuring streamlined data analysis workflows based on sophisticated statistical algorithms.

Clover MSDA, a server-client-based software, represents a powerful toolbox for statistical analysis of MALDI-TOF MS profile data accessible via an intuitive, workflow-oriented graphical user interface [1].

In this application note, we provide an overview of the workflows and statistical methods supported by Clover MSDA software and illustrate its key capabilities by using various types of example datasets for demonstration. These workflow examples include:

- biomarker analysis (i.e., search for discriminative  $m/z$  features) of human blood serum protein profiles,
- classification by means of hierarchical clustering and principal component analysis (PCA) of therapeutic antibody samples based on their N-glycan profiles,
- fish species verification from protein profiles via a trained support vector machine (SVM) prediction model.

## Statistical toolbox accessible via intuitive workflow GUI

- Clover MSDA software runs on a server accessible via an intuitive web browser interface.
- A user management system protects all data and other information stored on the server from unauthorized access.
- Structured data analysis projects allow for contextualized data management.
- Access rights can be shared between users for individual studies to enable collaborative data analysis work.
- Analysis results can be downloaded in various formats and at various levels of detail.

The following data analysis workflows are supported by the software:

### Biomarker Analysis

- This workflow analyzes MALDI-TOF data from samples of known categories with the aim of detecting discriminative  $m/z$  features that optimally separate these categories. The workflow uses univariate statistical methods, such as the T-test, Mann-Whitney U test, and ROC.
- Optionally, datasets can be subjected to outlier detection and reproducibility testing before the actual biomarker analysis.

### Classification

- This workflow allows for classification of MALDI-TOF spectra employing various multivariate statistical algorithms, such as Hierarchical Clustering, K-Means Clustering, Partial Least Squares Discriminant Analysis (PLS-DA), SVM, Random Forest (RF), K Nearest Neighbors (KNN), and Light Gradient-Booster Machine (LightGBM).
- Classification results obtained from PLS-DA, SVM, RF, KNN and LightGBM can be saved as prediction models for category assignment of spectra acquired from samples of unknown origin (see *Identification* workflow below).

## Identification

- This workflow is intended for category assignment of spectra acquired from unknown samples using one or multiple prediction models generated in the *Classification* workflow.

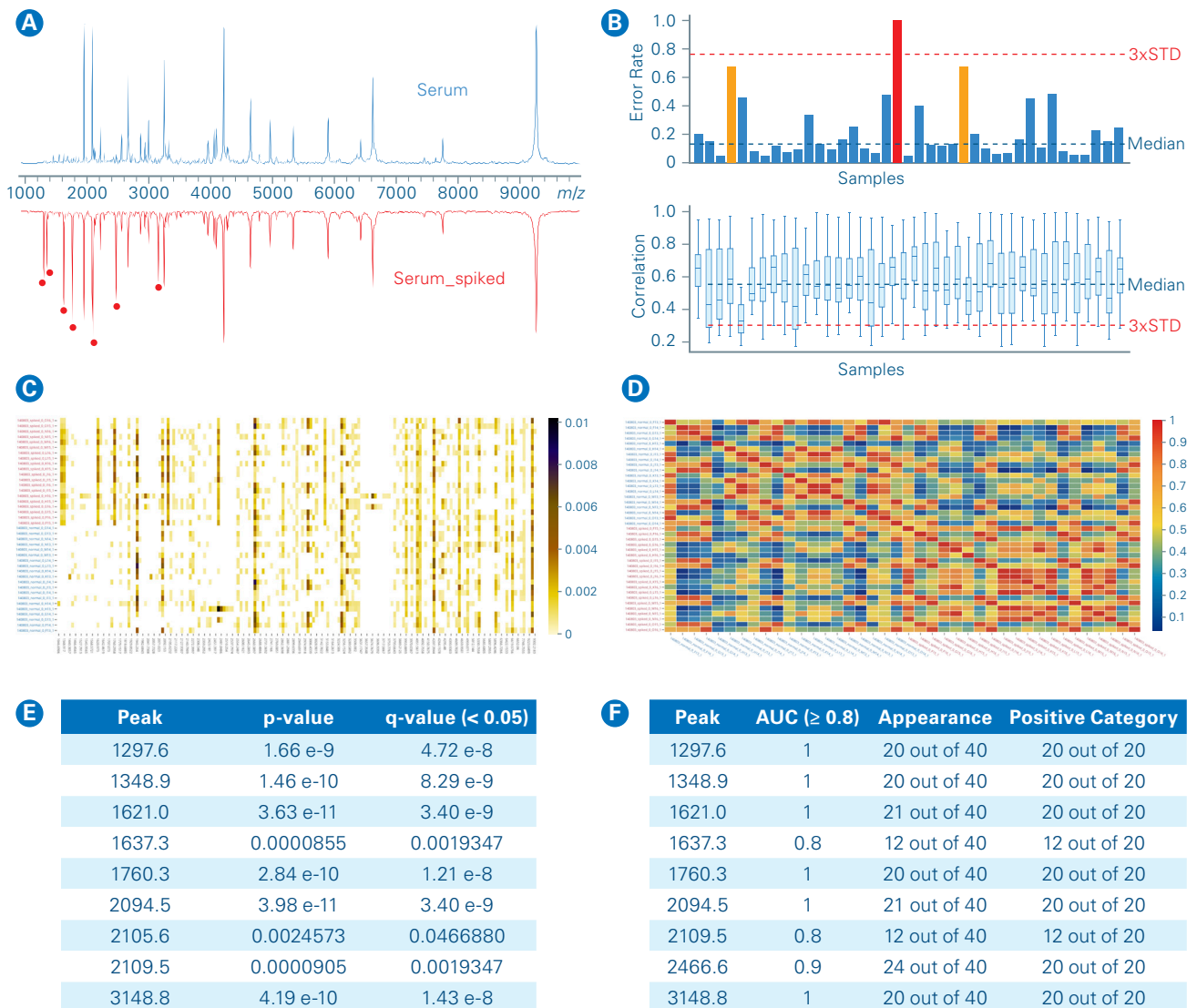
## Workflow example I: Biomarker Analysis

### Experiment

- Identify the most discriminative  $m/z$  features in MALDI-TOF protein profiles to differentiate between human blood serum samples spiked with artificial peptides and non-spiked controls.

### MALDI-TOF dataset

- Bruker autoflex<sup>®</sup> maX MALDI-TOF/TOF, positive linear mode,  $m/z$  range 1000-10,000, average-mass  $m/z$  features, 20 spectra per sample category (i.e., spiked, non-spiked).



**Figure 1. Biomarker analysis of spiked vs. non-spiked human blood serum.**

**(A)** Butterfly plot of representative spectra acquired from non-spiked and spiked human blood serum. Marked peaks represent spiked peptides. **(B)** Outlier detection result (PCA reconstruction and spectra correlation plots). **(C)** Peak heatmap. **(D)** Pearson correlation heatmap. **(E)** T-test result table listing  $m/z$  features with  $q < 0.05$ . **(F)** ROC result table providing  $m/z$  features with  $AUC \geq 0.8$ . All  $m/z$  features revealed by T-test and ROC analysis as discriminative between sample categories represent the spiked peptides and their oxidized versions.

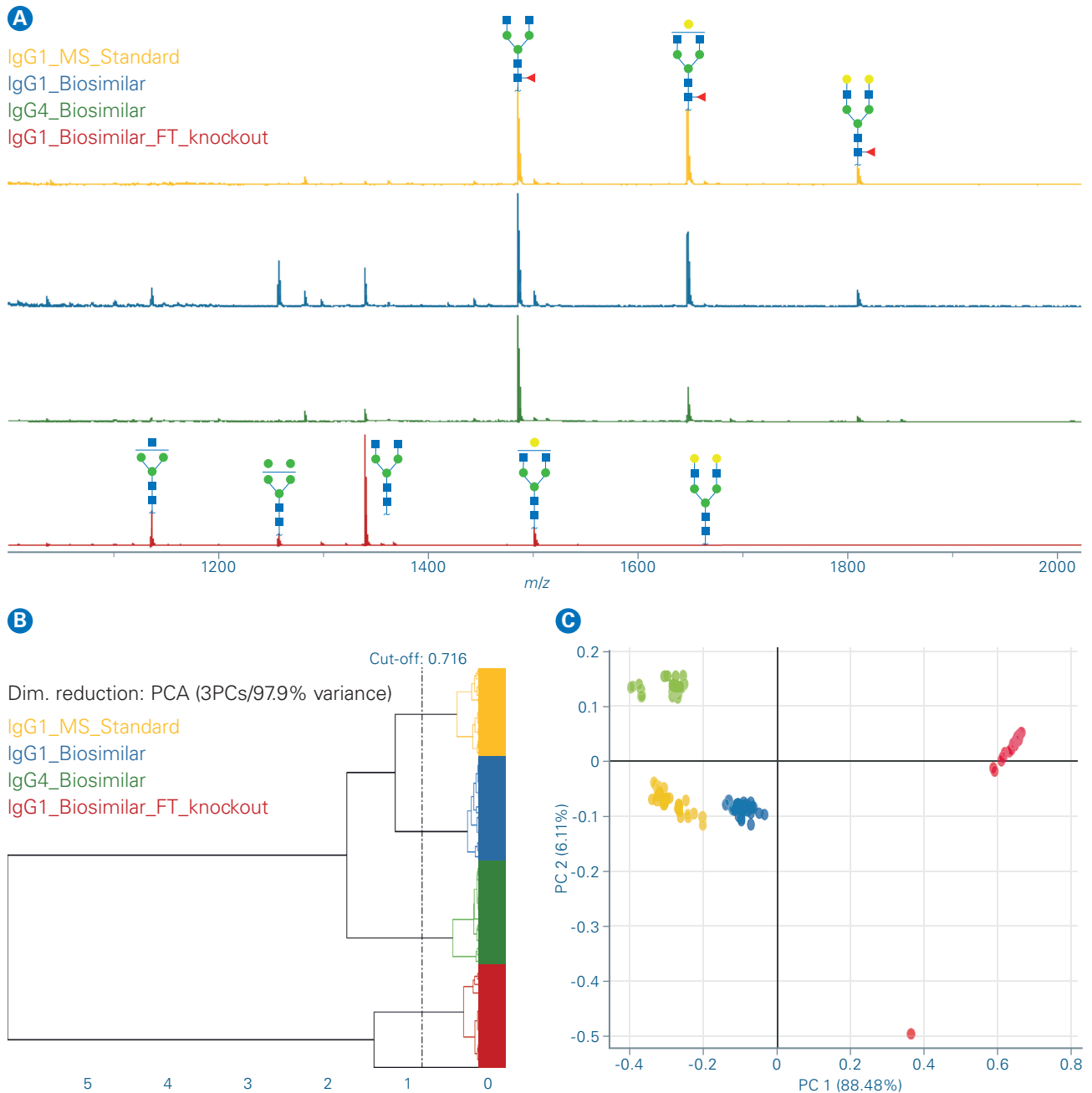
## Workflow example II: Classification

### Experiment

Classify various nivolumab MS reference standards and research-grade biosimilars based on their MALDI-TOF N-glycan profile spectra.

### MALDI-TOF dataset

Bruker neoflex™ benchtop MALDI-TOF/TOF, positive reflector mode,  $m/z$  range 1000-3500, monoisotopic  $m/z$  features, 30 spectra per sample.



**Figure 2. Classification of nivolumab standards and biosimilars based on their N-glycan MALDI-TOF profile spectra.**

**(A)** Representative N-glycan spectra obtained from nivolumab standards and biosimilars. **(B)** Hierarchical clustering result (dendrogram). **(C)** PCA result (score plot). Both hierarchical clustering and PCA successfully classified the spectra in the dataset by sample origin. Furthermore, both classification methods confirmed the fucosyltransferase (FT) knockout IgG1 biosimilar being particularly different in its N-glycan profile from the other three sample categories, which is explained by the lack of fucosylated N-glycan compositions in the FT knockout sample. Additionally, PCA successfully separated samples by IgG subtype (IgG1 vs. IgG4 Fc regions) along PC2.

## Workflow example III: Identification

### Experiment

Generation and validation of an SVM prediction model for verification of closely related *Salmonidae* family fish species: Arctic char (*Salvelinus alpinus*), Atlantic salmon (*Salmo salar*), and King salmon (*Oncorhynchus tshawytscha*) based on MALDI-TOF protein profile spectra generated from fish tissue extracts.

### MALDI-TOF dataset

Bruker microflex® LRF benchtop MALDI-TOF, positive linear mode,  $m/z$  range 2000-30,000, average-mass  $m/z$  features

Model generation: 15-16 spectra per fish species

Model validation: 6-8 spectra per fish species

Identification: 18 spectra



**Figure 3.** Generation of the SVM model for *Salmonidae* fish species verification.

**(A)** Representative MALDI-TOF protein profile spectra of Arctic char (blue), Atlantic salmon (orange), and King salmon (green). **(B)** 10-fold cross-validation matrix of the SVM model generated from spectra of Arctic char, Atlantic salmon, and King salmon samples. **(C)** SVM loadings used in the model generation.

Identification Results		
Low (0%-55%)	Medium (55%-75%)	High (75%-100%)
Sample Name	Predicted Category	Probability
atlantic-salmon1_1_H8_fresh_2010_0712_a	Atlantic Salmon	93,19%
atlantic-salmon1_4_B9_fresh_2010_0712_a	Atlantic Salmon	91,72%
Arctic_char_2_3_B9_2010_0715_fresh	Arctic Char	91,62%
Atlantic_salmon_1_6_G2_fresh_2010_0713_a	Atlantic Salmon	90,65%
Atlantic_salmon_1_7_H3_2010_0713_a	Atlantic Salmon	90,59%
King_AK_salmon_1_6_G5_fresh_2010_0713_a	King Salmon	89,97%
Arctic_char_auto_1_2010_0712_fresh_0_C5_1	Arctic Char	89,56%
Arctic_char_auto_1_2010_0712_fresh_0_G6_1	Arctic Char	88,56%
King-Salmon_1_4_G3_2010_0709_fresh_a	King Salmon	88,52%
Arctic-char_1_2_H6_fresh_2010_0712_a	Arctic Char	87,77%
Arctic_char_auto_1_2010_0712_fresh_0_H5_1	Arctic Char	87,34%
Arctic_char_auto_1_2010_0712_fresh_0_H6_1	Arctic Char	85,99%
Arctic-char_1_1_B5_fresh_2010_0712_a	Arctic Char	84,80%
Atlantic_salmon_1_6_B2_2010_0713_a	Atlantic Salmon	84,38%
Arctic_char_auto_1_2010_0712_fresh_0_B5_1	Arctic Char	83,30%
Arctic_char_auto_1_2010_0712_fresh_0_C6_1	Arctic Char	81,47%
Atlantic_salmon_2_7_C9_fresh_2010_0716	Atlantic Salmon	80,08%
King_salmon_2_6_G2_fresh_2010_0716	King Salmon	66,30%

**Table 1.** Identification results using the SVM model for 18 protein profile sample spectra from Arctic char, Atlantic salmon, and King salmon samples. 17 spectra were correctly verified with high probability (>75%), and one with medium probability (55%-75%).

## Conclusions

- Clover Biosoft MSDA software provides a powerful statistical toolbox to efficiently leverage the discriminative power of MALDI-TOF MS profiling data originating from biological samples.
- The software provides access to a wide range of statistical methods via an intuitive, workflow-oriented web interface. It features user management, project and experiment-driven data management, as well as detailed reporting of results.
- Clover Biosoft MSDA software supports tailored data analysis workflows, allowing for time-efficient analysis of large cohorts of MALDI-TOF spectra with minimal effort and providing tremendous benefit to biomarker discovery as well as classification and identification tasks across a broad range of life science applications.

## References

[1] <https://www.clovermsdataanalysis.com/>

Bruker Daltonics is continually improving its products and reserves the right to change specifications without notice. © Bruker Daltonics 11-2025, TN-63, 1924807

For Research Use Only. Not for use in clinical diagnostic procedures.

### **Bruker Switzerland AG**

Fällanden · Switzerland  
Phone +41 44 825 91 11

### **Bruker Scientific LLC**

Billerica, MA · USA  
Phone +1 (978) 663-3660

