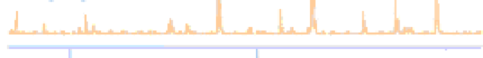
Compass

Application Tutorial



BRUKER

Data Processing of ESI LC-MS/MS Data for Protein Identification Tutorial

Contents

1 Starting Point: HyStar Sample Table 5

2 Processing Data using DataAnalysis 4.0 6

2.1 Loading a File 6
2.2 Find Compounds and Create Mass List 7
2.3 Deconvolution 15
2.4 Peaklist Export 19
2.5 ProteinAnalysis for Analyzing Large Data Sets 24
2.6 Automatic Processing 25

3 Database Search using ProteinScape 26

Revision 1.5 (October 2009)

Bruker Daltonics

Legal and Regulatory Notices

Copyright 2009 Bruker Daltonik GmbH

All Rights Reserved

Reproduction, adaptation, or translation without prior written permission is prohibited, except as allowed under the copyright laws.

Document History

Data Processing of ESI LC-MS/MS Data for Protein Identification Tutorial, Revision 1.5 (October 2009). Printed in Germany

Warranty

The information contained in this document is subject to change without notice.
Bruker Daltonik GmbH makes no warranty of any kind with regard to this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Bruker Daltonik GmbH shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance or use of this material. Bruker Daltonik GmbH assumes no responsibility for the use or reliability of its software on equipment that is not furnished by Bruker Daltonik GmbH.

Bruker Daltonik GmbH
Fahrenheitstrasse 4
28359 Bremen
Germany

Telephone: +49 (0) 421 2205 200

FAX: +49 (0) 421 2205 103

E-mail: sales@bdal.de
Internet: www.bdal.de

Limitations on Use

For Research Use Only (RUO). Not for use in diagnostic procedures.

Hyperlink Disclaimer

Bruker Daltonik GmbH makes no express warranty, neither written nor oral, and is neither responsible nor liable for data or content from the linked internet resources presented in this document.

Preliminary remarks

Related software:

Compass 1.3 and Compass 1.5 for solariX:

- HyStar 3.2 SR2 (HyStar 3.4 for solariX)
- trapControl 7.0
- micrOTOF Control 3.0
- solariXcontrol 1.5
- and DataAnalysis 4.0 SP2

BioTools 3.2
ProteinScape 2.0
Mascot 2.2.04

This document focuses on data processing of ESI LC-MS/MS data of digested protein samples using DataAnalysis. The demo data `BSA Standard 50 fmol_D3_01_356.d` are installed per default in the folder `D:\Data\Demo` (if demo data installation is active). The steps performed and parameters required are explained and the entire automatic workflow is outlined. This includes sample acquisition and peak list transfer to ProteinScape and BioTools for data archiving, detailed analysis, SILE quantification, etc.

Table of Contents

1 Starting Point: HyStar Sample Table ... 5

2 Processing Data using DataAnalysis 4.0 ... 6

2.1 Loading a File ... 6
2.2 Find Compounds and Create Mass List ... 7
2.2.1 Display Parameters ... 8
2.2.2 Parameters for FindCompounds AutoMS(n) ... 9
2.2.3 Parameters for annotating MS and MS(n) Signals ... 10
2.2.4 Find Compounds ... 13
2.3 Deconvolution ... 15
2.3.1 Deconvolution Parameters ... 15
2.3.2 Perform Deconvolution ... 18
2.4 Peaklist Export ... 19
2.4.1 Export Parameters ... 19
2.4.1.1 Use of Global Charge Limitation ... 22
2.4.1.2 MS(n) Charge Limitation for Database Search ... 23
2.5 ProteinAnalysis for Analyzing Large Data Sets ... 24
2.6 Automatic Processing ... 25

3 Database Search using ProteinScape ... 26

1 Starting Point: HyStar Sample Table

Parameters for the following steps in protein identification are set in the HyStar Sample table:

- Peptide separation — Nano LC systems of different vendors are supported. For details please refer to the respective manual.
- Data acquisition — amaZon series instruments, micrOTOF-Q and Bruker FTMS series instruments are supported. For details please refer to the respective manual.
- Recalibration — For a description of recalibration (for micrOTOF-Q) please refer to the relevant tutorial (in preparation).
- Data processing including export — This is described in the present tutorial.

The HyStar Sample Table also contains the target in the ProteinScape navigation tree and a BioTools database search method (optional). For details please refer to the respective manual.

2 Processing Data using DataAnalysis 4.0

This chapter describes how to set the parameters for interactive data processing in DataAnalysis. When the parameters have been optimized, the workflow can be set up for automatic runs. This is recommended for routine measurements.

`BSA Standard 50 fmol_D3_01_356.d` is a micrOTOF-Q dataset that is suitable for optimizing the system. The dataset has been installed in `D:\Data\Demo` (default installation).

2.1 Loading a File

Select **File > Open** to load the LC-MS/MS analysis of the digested protein. The navigation tree shows the file name, and the chromatograms available within the folder `Chromatograms`. When opening a file for the first time, by default the Total Ion Chromatogram (TIC) for All MS and All MSn are shown (see Figure 2-1).

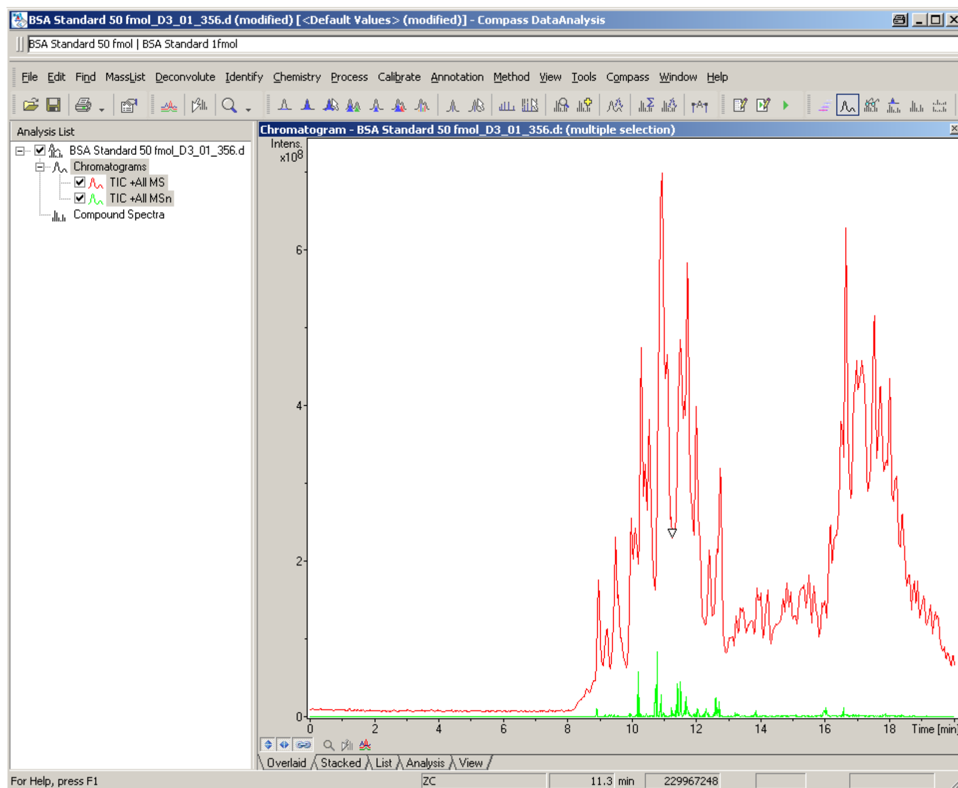


Figure 2-1 Default file view

The Base Peak Chromatogram better clarifies the course of the separation. To display it, activate **Edit > Chromatograms…** . Suggested parameters are shown in Figure 2-2.

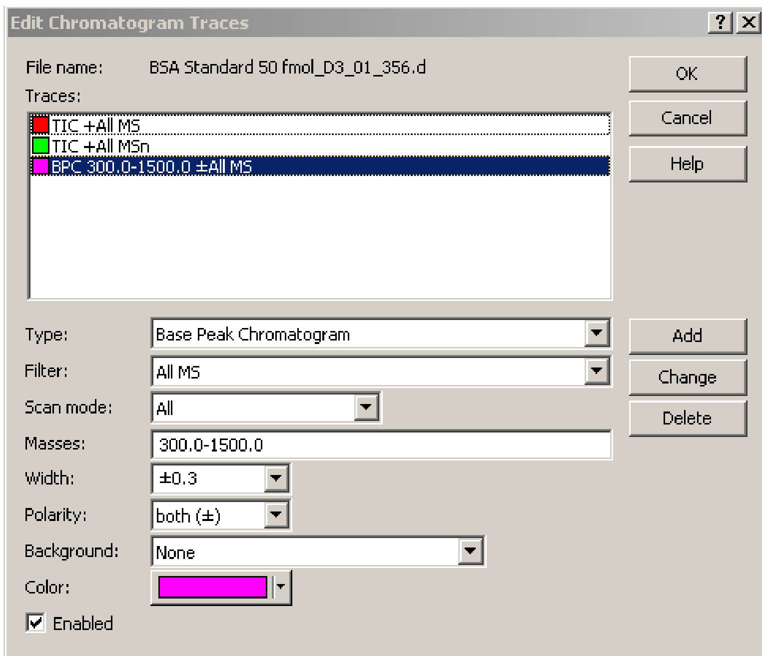


Figure 2-2 Dialogue for defining a base peak chromatogram

2.2 Find Compounds and Create Mass List

Use **Find > Compounds - AutoMS(n)** to find compounds in the acquired Auto-LC/MS(n) data. Compounds are composed of:

- an MS (single or averaged)
- an MS/MS spectrum (single or averaged). The MS/MS spectra of one compound contain the identical parent ion within a narrow m/z and time range. The latter can be set in the Find AutoMS(n) parameters (chapter 0).
- A compound can contain more than one MS/MS spectrum e.g. for alternating CID&ETD data and for MS(n) with n>2.

The compound MS spectrum is not necessarily the spectrum directly preceding the MS(n) spectra. If another MS spectrum is better in respect of intensity and resolution, it will be chosen as precursor MS spectrum instead.

2.2.1 Display Parameters

Mass precision is set in **Method > Parameters > Display**. This value depends on the mass precision of the MS instrument. For example, a setting of 2 is recommended for ion traps, 4 is recommended for micrOTOF-Q, and 5 is recommended for Bruker FTMS series instruments (see Figure 2-3).

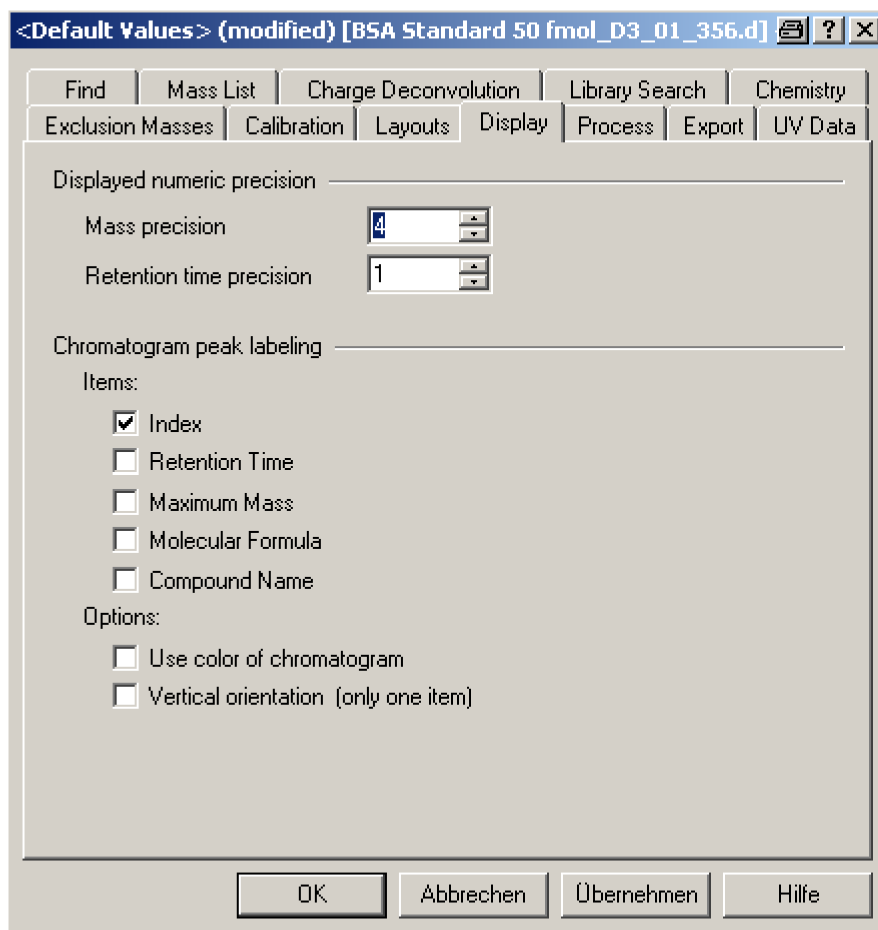


Figure 2-3 Display Parameter dialogue with settings suitable for micrOTOF-Q data

2.2.2 Parameters for FindCompounds AutoMS(n)

The parameters used for the **FindCompounds** function are set in **Methods > Parameters > Find > AutoMS(n)** (see Figure 2-4).

Parameter	Explanation / Recommendation
Compound detection	
Intensity threshold (TIC AllMSn	Minimum intensity of the "TIC, All MS2". Only MS/MS spectra with a "TIC, All MS2" above this value will be considered for compound detection.
Maximum number of compounds	Limits the total number of compounds detected (= n). The setting of this parameter strongly depends on the complexity of the sample. Compounds are ranked according to 'TIC, All MS/MS' signal intensity and the n highest intensity compounds are detected.
Retention time window [min]	Within the given retention time window, all MS/MS spectra of the same parent ion will be averaged. It is recommended that this parameter is set to the base width of the chromatographic.
Merge positive and negative compounds	This parameter can be applied for alternating mode measurements. If it is set, it results in the compound containing positive and negative spectra. If this option is not set $[M+H]^+$ and $[M-H]^-$ will be individual compounds.
Chemistry and filters	
Fragments qualified by	Select **Amino acids** to reduce the number of "noise compounds" (not quantitative). Only MS/MS spectra showing fragments containing at least one amino acid will be used for compound extraction.
Filter for detected neutral losses only	Only those compounds will be filtered out which have a fragment spectrum that was acquired due to a "detected neutral loss" (in trapControl). Not important for standard protein ID workflow.
Mass spectrum calculation	
Spectrum type (Line/Profile)	Use **Auto**. Profile spectra will be used if available, otherwise Line spectra will be used. For Bruker FTMS instruments line spectra should be used.
Background subtraction (MS only)	Select the type of background subtraction to be performed on averaged compound spectra. The optimal type of subtraction depends on the quality of the data. Applies to the MS spectra only.
Add UV spectrum if available	Should be **OFF**

Parameter	Explanation / Recommendation
Delete previously found compounds of same type in selected range …(recommended)	Before starting the selected compound finding procedure, check this option to delete the compounds of the same integration type previously found in the selected Rt range. Otherwise, the new compounds are added to the previous ones.

Table 2-1 FindCompounds parameters

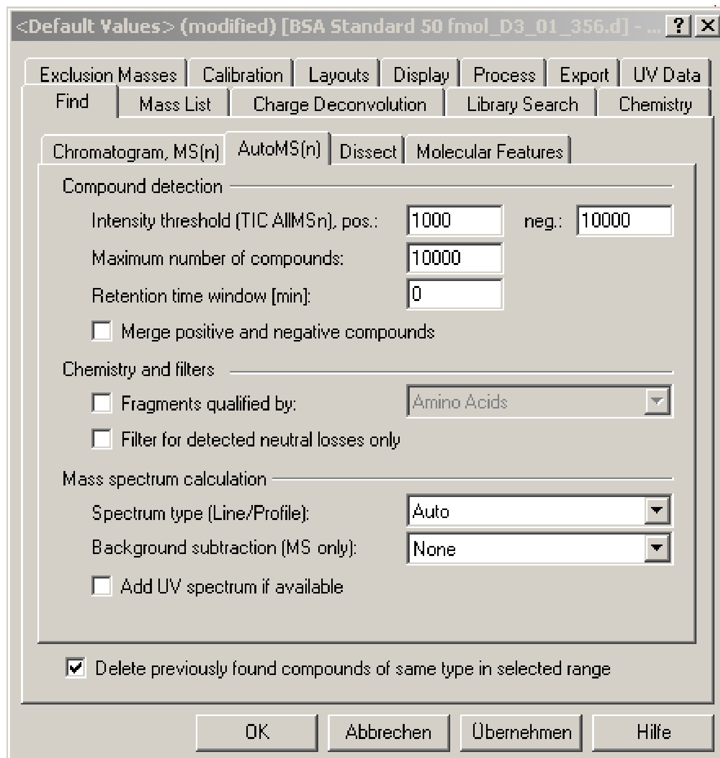


Figure 2-4 Find > AutoMS(n) parameters

2.2.3 Parameters for annotating MS and MS(n) Signals

The parameters for Mass Annotation are set in **Methods > Parameters > MassList** (see Figure 2-5). This should be done before starting **FindCompounds**, since MS and MS(n) signals are automatically labeled with the currently set parameters. The table below describes all peak picking algorithms available in DataAnalysis and focuses on micrOTOF-Q parameters in detail.

Parameter	Explanation / Recommendation
Instrument type	Select the appropriate type of instrument from the drop-down list
Peak finder	
Apex	Based on the first derivative of the signal intensity, the **Apex** algorithm calculates the position of the peak maximum by an interpolation between the neighboring data points (fitted maximum). Recommended for peak annotation of ion trap data.
Centroid	The **Centroid** algorithm uses data points that are larger than 0.01% height x maximum peak intensity to calculate the centroid (mean of the intensity weighted masses) of the peak. Of minor importance for protein ID.
SNAP	The **SNAP II** algorithm (Sophisticated Numerical Annotation Procedure) calculates the isotopic distribution of the intensities for a given monoisotopic mass of a known substance class and chemical derivations. This simulated curve is fitted to the peak in the spectrum. Both overlapping patterns and differently charged ions are considered. Probably too slow for complete LC-MS/MS data. Mainly used for MALDI-TOF and FTMS analyses.
FTMS	Specific FTMS peak finder for the Bruker FTMS series.
Sum Peak	Very fast peak detection algorithm which uses sliding sums. Recommended for micrOTOF-Q.
Use same width as used in acquisition (recommended)	Check this option to use the same peak width for processing as for data acquisition. This is currently only supported for micrOTOF and micrOTOF-Q analyses.
Resolving power (m/dm)	Check this option to apply the resolving power (m/Δm) for peak detection, and enter the measured resolving power value of the mass spectrum in the entry field to the right. Recommended for Time-Of-Flight analyses.
Peak width (FWHM)	Check this option to apply the peak width at half height (Full Width at Half Maximum) for peak detection, and enter the measured minimum peak width value either in [m/z] or [pnts.] of the spectrum in the entry field to the right. To toggle between [m/z] and [pnts.], click **Points**.

Parameter	Explanation / Recommendation
S/N threshold	The S/N is defined as the height of the mass peak above its baseline, relative to the noise. The noise level is defined as 5 x sigma (sigma = standard deviation), and it is calculated on the full mass spectrum except the regions of peaks. Since only one S/N value is calculated per spectrum, it is recommended to set a low value.
Relative intensity threshold (base peak)	Intensity threshold relative to the most intense (base) peak in the mass spectrum or in the selected mass range.
Absolute intensity threshold	Absolute intensity threshold which a peak must exceed. Depends on peptide amount used for the measurement.
Filter exclusion masses (minor relevance)	This will exclude the masses currently defined in the Exclusion Masses tab of the Method Parameters dialog from being included in the Mass List.
Interactive Mass List editing	
Use peak finder to calculate peak position	Only relevant for interactive mass list editing.

Table 2-2 Peak picking algorithms available in DataAnalysis and micrOTOF-Q parameters on the Mass List page

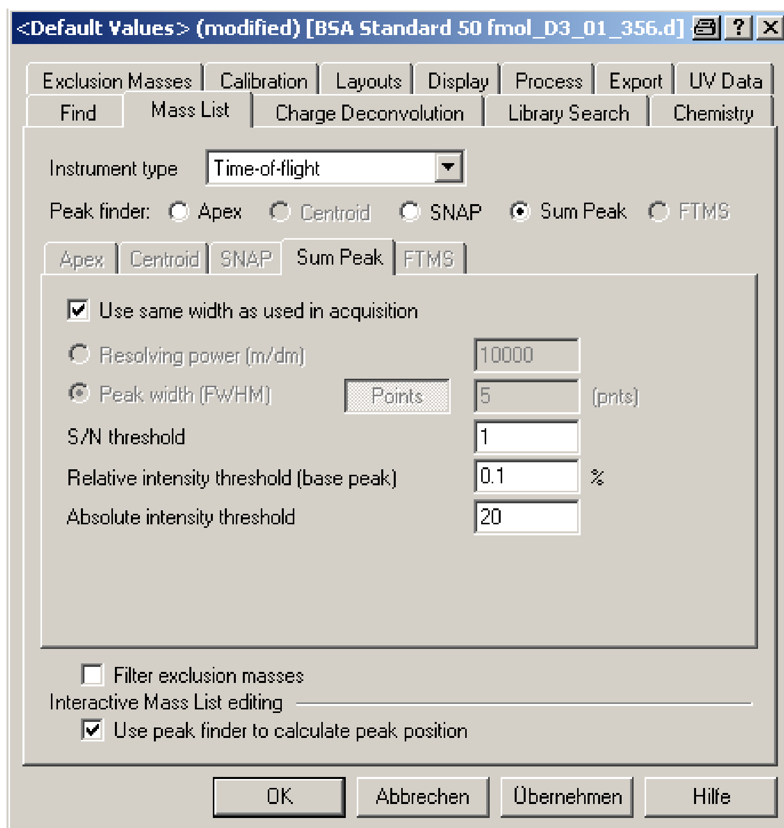


Figure 2-5 Mass Find parameters (micrOTOF-Q).

MS(n) data should be used to check the suitability of the Mass Annotation parameters because good fragmentation coverage requires very well annotated MS(n) signals.

2.2.4 Find Compounds

Left-click on the file in the tree view to activate it. For interactive processing, the time range can be restricted by deactivating **Edit > SelectRange/ViewSpectra** and selecting the appropriate time range in the Chromatogram window by clicking and dragging over the desired region with the left mouse button depressed. Select **Find > Compounds-AutoMS(n)** to create compounds. Once these steps have been peformed, three changes occur (see Figure 2-6):
The Compound Spectra folder in the tree view contains a number of compounds.
In the Chromatogram window the TIC is labeled with compound numbers, and the AutoMS(n) trace connects all MS-TIC data points for each compound.

- The Compound Spectra window shows compound spectra.

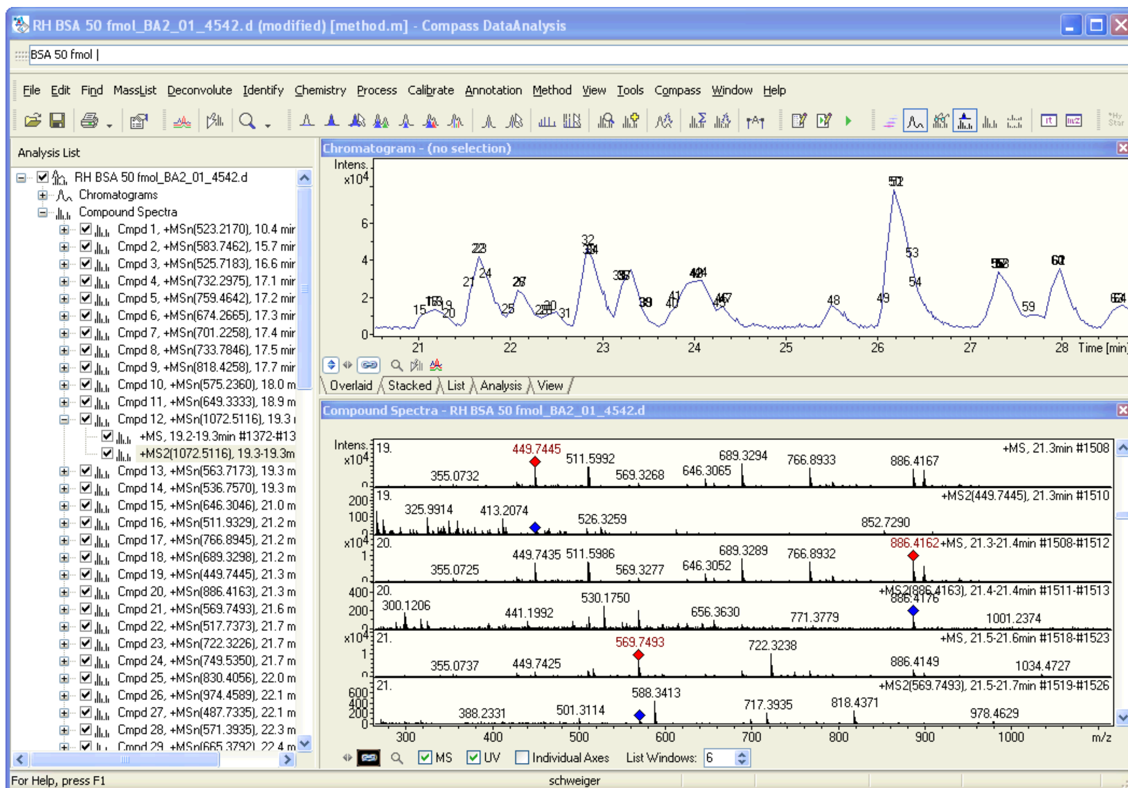


Figure 2-6 DataAnalysis view after Find Compounds calculation

2.3 Deconvolution

Deconvoluting the data before performing a database search is strongly recommended. Deconvolution is performed to determine the molecular mass MH^+ of a peptide. Two types of deconvolution are supported by DataAnalysis:

- In Related-ion deconvolution the mass of a compound is determined from multiple signals representing different charge states.
- In Resolved-isotope deconvolution, the mass of a compound is determined from multiple signals representing different isotopes (see Figure 2-7).

Both types can be used for MS data. For MS(n) spectra, only Resolved-isotope deconvolution is used.

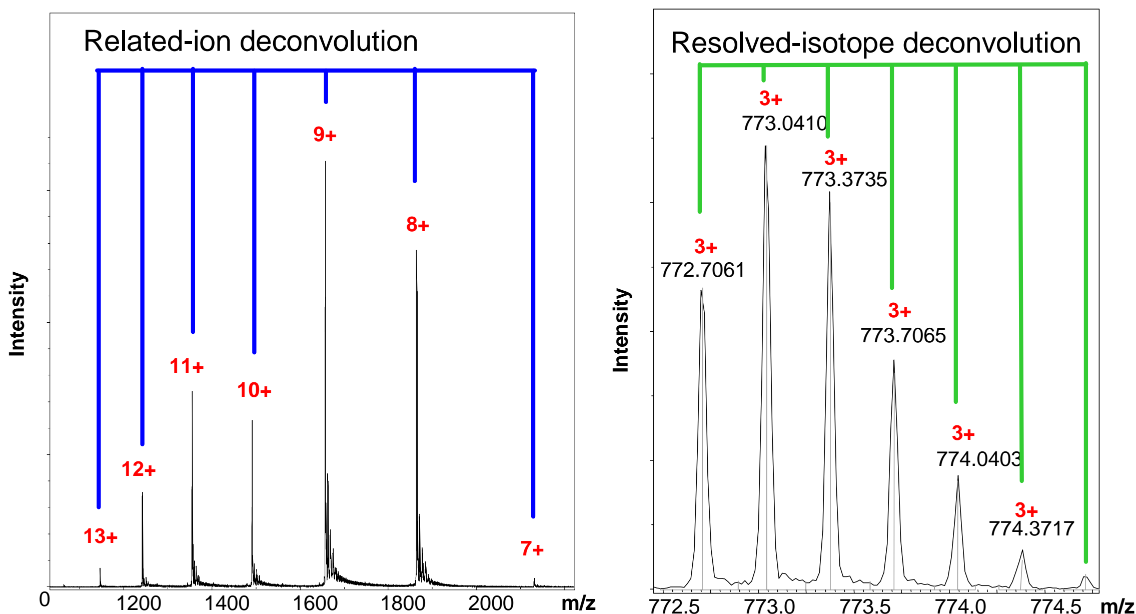


Figure 2-7 Diagrams demonstrating the difference between Related-ion (left) and Resolved-isotope deconvolution for micrOTOF-Q data.

2.3.1 Deconvolution Parameters

Some of the Deconvolution parameters are set in **Methods > Parameters > Deconvolution > General**.
Here, the option **Include shifted spectrum** should be **deselected** (see Figure 2-8).
More relevant for Protein ID are the parameters in **Methods > Parameters > Deconvolution > Peptides / Small Mol** (see Figure 2-9):

Parameter	Explanation / Recommendation
Adduct ions	The settings usually used for peptides are +H and –H (as shown in Figure 2-9)
Deconvolute	
MS Full scan	Should be turned **ON**.
MS(n)	Should be turned **ON**.
Low mass / High mass	This is the range of the resulting deconvoluted masses (MH^+). It depends on the m/z range set in data acquisition and the observed charges. It is set automatically when mass range is unchecked. The MaxResScan settings are of minor relevance
Abundance cutoff (%) [MS Full Scan]	Minimum percentage abundance of a signal compared to the abundance of the largest signal in the spectrum. 10 is recommended (2 for very low-abundance signals).
Abundance cutoff (%) [MS(n)]	2 is recommended (0.5 for very low-abundance signals).
Resolved-isotope deconvolution	
Maximum charge	Depends on scan type and instrument: For ion traps, 2–3 is recommended for MS and 2 for MS/MS. For micrOTOF-Q, 3–4 is recommended for MS and 2–3 for MS/MS.
Allow precursor deconvolution from fragment spectra (recommended)	The compound MS/MS spectrum is used to determine the parent charge.
Related-ion deconvolution (MS Full Scan only)	
Maximum charge; MW agreement (0.01%); Min. peaks in comp.	Usually not necessary for well resolved full-scan MS spectra. If selected, allows setting of maximum charge, molecular weight agreement, and minimum number of peaks in a set.
Clear previous results	Recommended
Create neutral spectrum (instead of singly charged spectrum)	Should be turned **OFF**.

Table 2-3 Deconvolution parameters

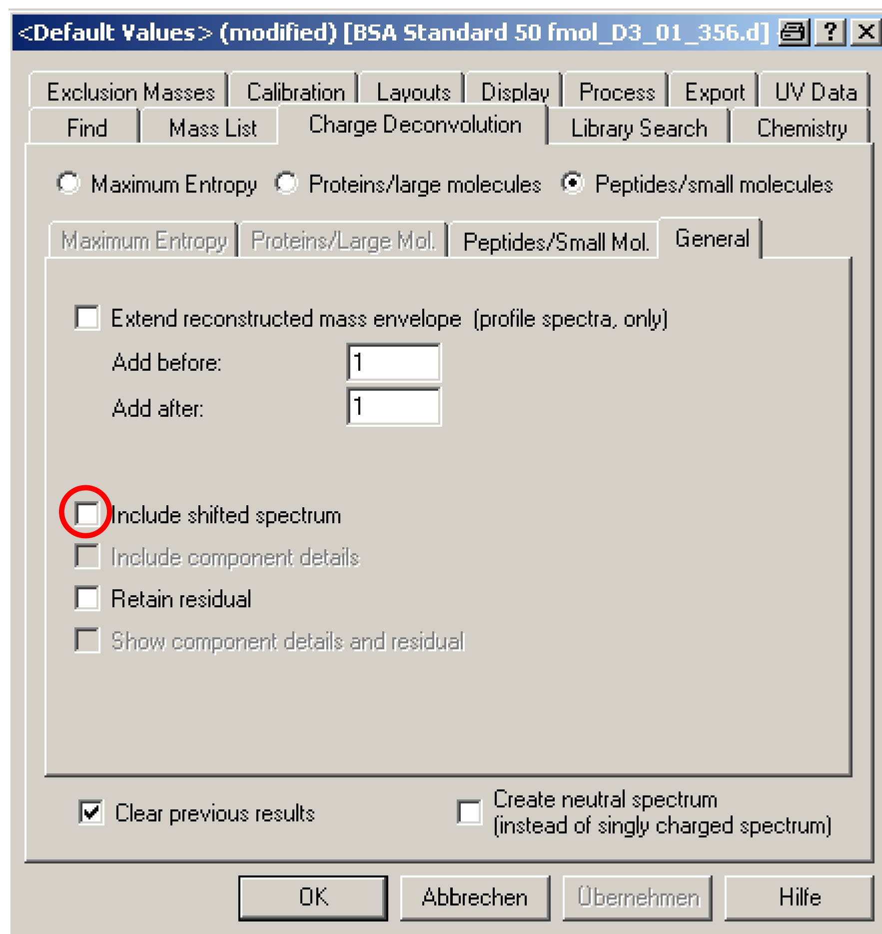


Figure 2-8 General Deconvolution parameters

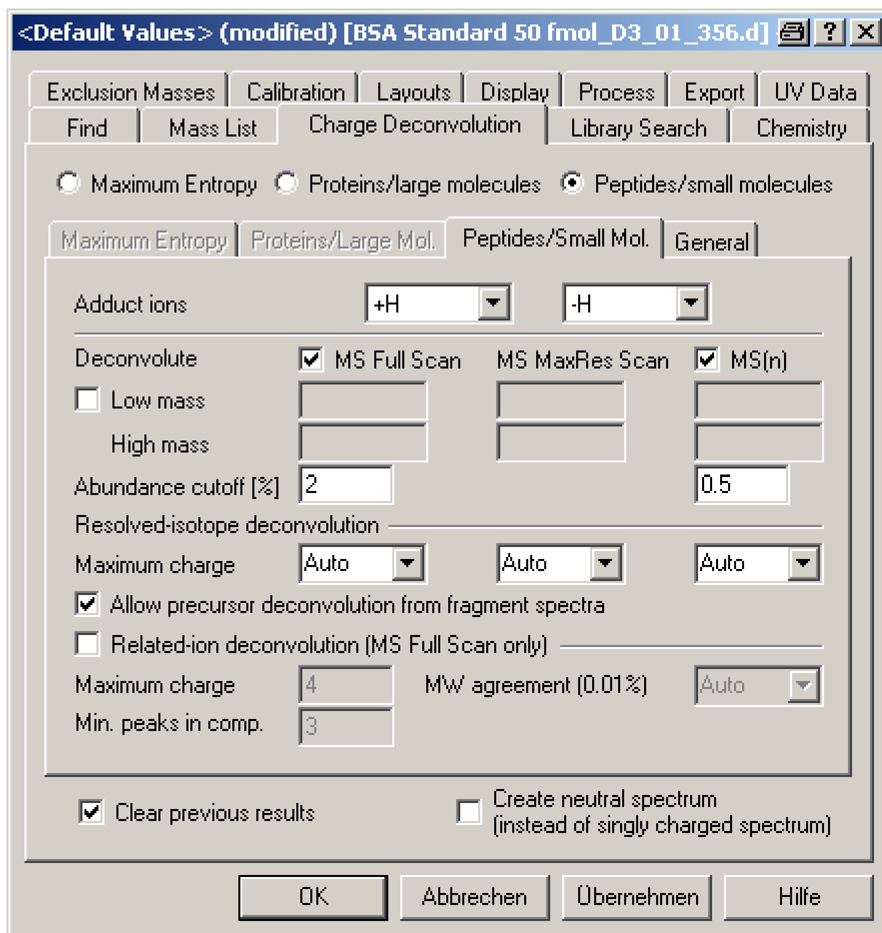


Figure 2-9 Deconvolution parameters for low-abundance peptides measured by micrOTOF-Q

2.3.2 Perform Deconvolution

Make sure that the desired compound mass spectra are activated in the navigation tree. Selecting **Deconvolute > MassSpectra** displays the deconvolution results of the MS and the MS(n) spectra in the Compound Mass Spectra window. The signals are additionally labeled with their corresponding charges (Figure 2-10).

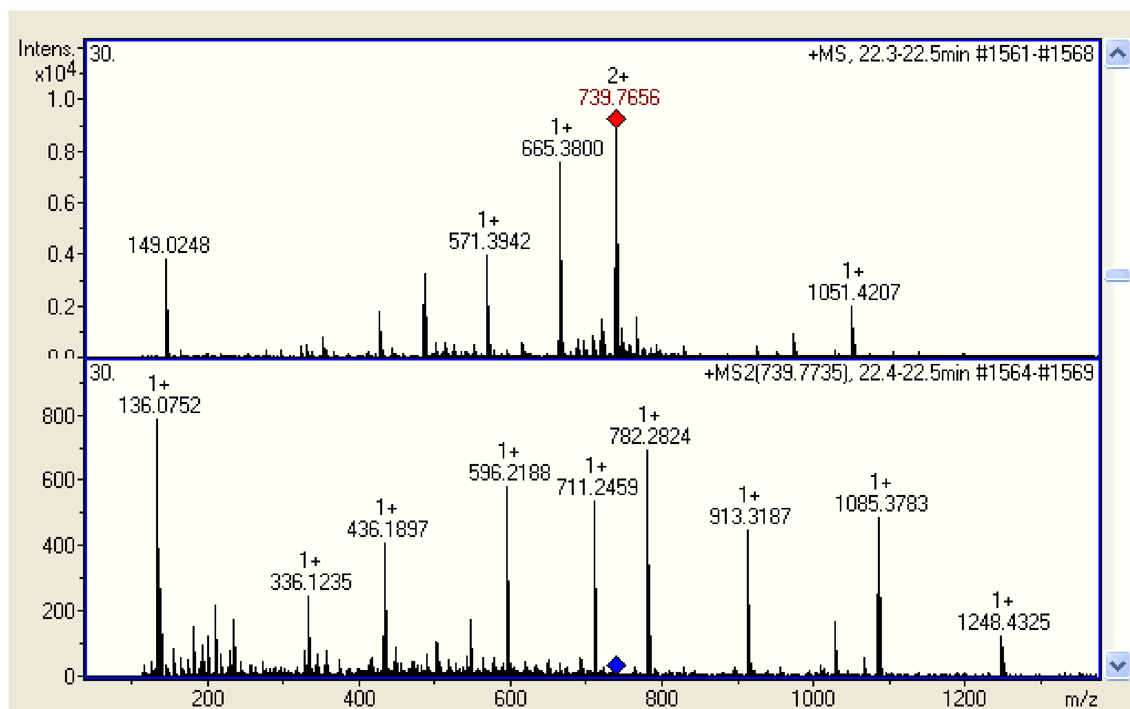


Figure 2-10 Compound Spectra view of one compound after deconvolution.

2.4 Peaklist Export

2.4.1 Export Parameters

General export parameters are set in **Methods > Parameters > MascotExportOptions > General** (see Figure 2-11). The more relevant peak list export parameters for peptide data are set in **Methods > Parameters > MascotExportOptions > PeptideDatabaseQuery** (see Figure 2-12).

Parameter	Explanation / Recommendation
General	
XML compound export	
Include detailed information	Should be **OFF**. Only useful for very special applications that require detailed information in the exported XML file.
Include peptide database query information	
Mascot compound export	
Include peptide database query information	
Peptide Database Query	
Global charge setting	
Set global charge limitation:	"1+, 2+ and 3+" is recommended since not always all parent peaks are deconvoluted. For a detailed explanation see below.
'AccuMASS' exported m/z values	Only useful for mass deviations between 0.05 and 0.5 Da. Do not use this option for micrOTOF or Bruker FTMS series instruments.
Export for 'MS/MS Search' and 'Peptide Mass Fingerprint'	
Mixed list (non-deconvoluted and deconvoluted)	The option of choice for protein ID. Here, all compounds are exported, independent of the deconvolution state of the parent mass. The peak list contains the parent mass (incl. charge info if available), all deconvoluted MS/MS signals and a limited number of non-deconvoluted MS/MS signals (see below).
Deconvoluted data only	Only those compounds are exported where the deconvolution of the MS signal and at least of one MS(n) signals for each compound has been successful.
Non-deconvoluted data only	Only compounds with non-deconvoluted parents together with a limited number (see below) of non-deconvoluted MS(n) signals are exported.

Parameter	Explanation / Recommendation
Include S/N ratio and FWHM	When turned **ON** S/N and FWHM are included in the exported peak list.
Normalize MS(n) data	This option refers to the intensity of signals in the MS(≥3) fragment spectra which are lower than the intensity of the fragmented peak. When **Normalize MS(n)** is turned **ON** the intensity of the highest signal in the MS(3) spectrum is set equal to the intensity of the fragmented peak in the MS(2) spectrum. This feature is useful for *De-novo* sequencing applications, but for protein identification it is recommended to turn **Normalize MS(n)** to **OFF**
Export deconvoluted peaks as single-charged ion	Only useful for detailed peptide analyses.
Export for 'MS/MS Search'	
Prefer Full Scan spectrum deconvolution results to MaxRes results	MaxRes results are of minor importance for routine protein ID.
Export N most abundant non-deconvoluted ions	The number of non-deconvoluted ions should be set to about 150.
Intensity threshold for non-deconvoluted ions	To avoid exporting noise peaks a reasonable threshold should be set for the intensity (100 for FTMS, 20 for micrOTOF-Q, 100 for ion trap).
Export for 'Peptide Mass Fingerprint' (MS/precursor)	
Export N most abundant deconvoluted ions **Export N most abundant non-deconvoluted ions**	In addition to the parent MS peak, further peaks from the MS spectrum can be exported (XML only). This is not required for the standard protein ID workflow.

Table 2-4 Export parameters

2.4.1.1 Use of Global Charge Limitation

The global charge limitation can be set using a number of different factors that are prioritized in the following way:

Charge Determination by Deconvolution
MS signal charges determined by deconvolution are used for database searches and are not affected by any charge limitation parameters. To use this type of global charge limitation:

- **MS Full Scan** deconvolution must be active.
- **Mixed list** should be selected in the **Export** parameters.

Setting a charge limitation for database search using ProteinScape 2.0, BioTools 3.1 SR, or Mascot 2.2
If activated, this global charge limitation is applied to non-deconvoluted parent MS signals. This means that charge permutation will be applied for the database search.

- **Mixed list** should be selected in the **Export** parameters.

Setting charge limitation in DataAnalysis
If set, this global charge limitation becomes active for non-deconvoluted MS signals, but only for older versions of the software listed above. In the **Export** options:

- **Mixed list** should be selected in the **Export** parameters.
- set a GLOBAL charge.

2.4.1.2 MS(n) Charge Limitation for Database Search

Non-deconvoluted MS signals are permuted within the **global charge limitation** range set in **DataAnalysis > Export > Peptide Database Query > Global charge setting.** The option **1+, 2+ and 3+** is recommended. Non-deconvoluted MS(n) signals are permutated from 1+ up to the globally set charge limitations, respectively.

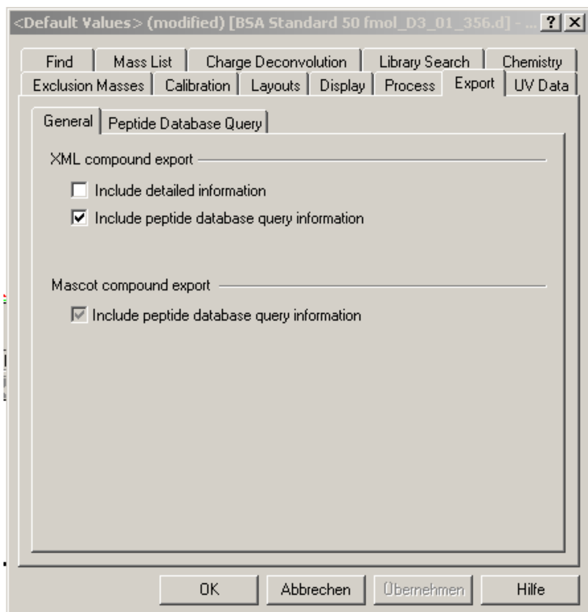


Figure 2-11 General Export parameters

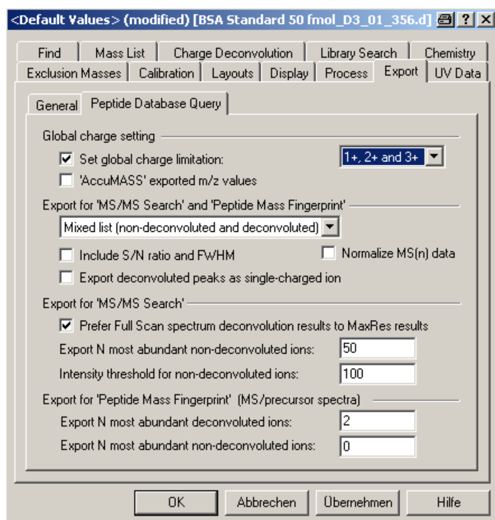


Figure 2-12 Export parameters for peptides

2.5 ProteinAnalysis for Analyzing Large Data Sets

After having set and optimized all parameters described above, **Tools > ProteinAnalysis** can be activated. This action can be viewed as a “short-cut” and corresponds to the consecutive execution of:

- Find > CompoundsAutoMS(n) with MassList >Find
- Deconvolute > MassSpectra
- File > Export > Compounds as MGF and XML files.

The exported peak list files are named:

- ProteinAnalysisResults.xml
- ProteinAnalysisResults.mgf
- ProteinAnalysisResults.ETD.mgf
- ProteinAnalysisResults.CID-ETD.mgf.

They are stored in the `Analysis` folder.
The big advantage of ProteinAnalysis is that the data are analyzed compound by compound, which saves memory and time.
A limitation is that no compound spectra are shown, since they are only created in-situ.
The analyzed retention time can be restricted by selecting a range. In this case, the retention time range will be added to the filename of the exported peak lists. Further details on filtering and limitation of the number of compounds are described in the manual.

2.6 Automatic Processing

Automation scripts are part of the DataAnalysis method. Running the scripts requires:

- Activation of the script in the HyStar Sample table. This is done by enabling Run script in the Compass method in SampleTable > Methods.
- A suitable automation script in the DataAnalysis method.

To check the automation script, open it in **Method > Script** within the DataAnalysis method. It is written in Visual Basic, as shown in Figure 2-13. After editing, the script should be saved using **Method > Save/SaveAs**.

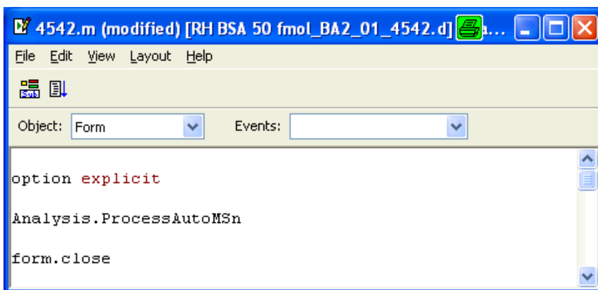


Figure 2-13 Automation script which executes ProteinAnalysis

3 Database Search using ProteinScape

The automatic identification workflow continues with database searches. This can either be done using the database system ProteinScape, which is described below, or using BioTools.
For the automatic transfer of peak lists from DataAnalysis, the Bruker PushDaemon has to be prepared.

- The daemon must be installed on the data processing computer.
- The daemon must be turned ON for the automatic peak list transfer.
- The location of the new XML peak lists is set on the Directories to Watch page.
- The ProteinScape server name is set on the Settings page.

Details are described in the ProteinScape tutorial and manual.

When the peak lists have been transferred to ProteinScape, the pre-defined database search method is applied. The parameters shown in
Figure 3-1 were used for the demo data shown used in this tutorial.

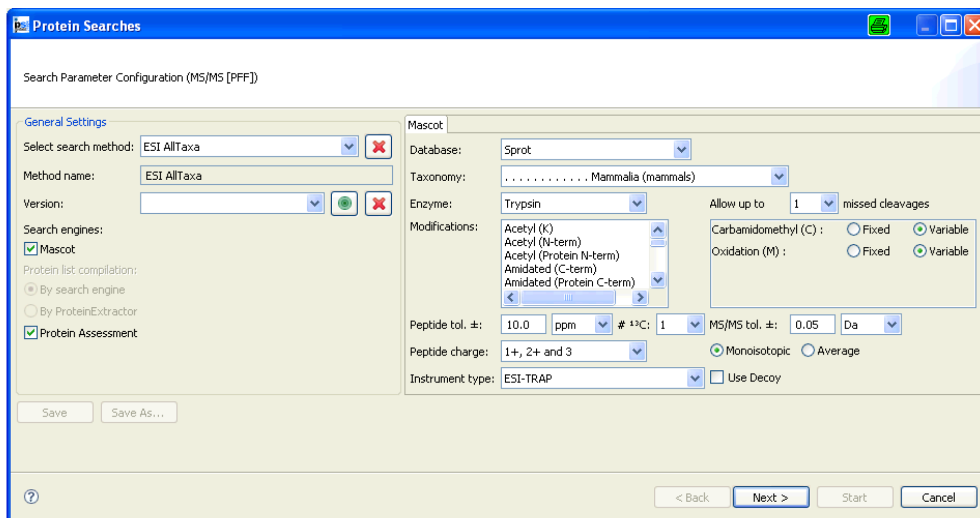


Figure 3-1 Database search parameters

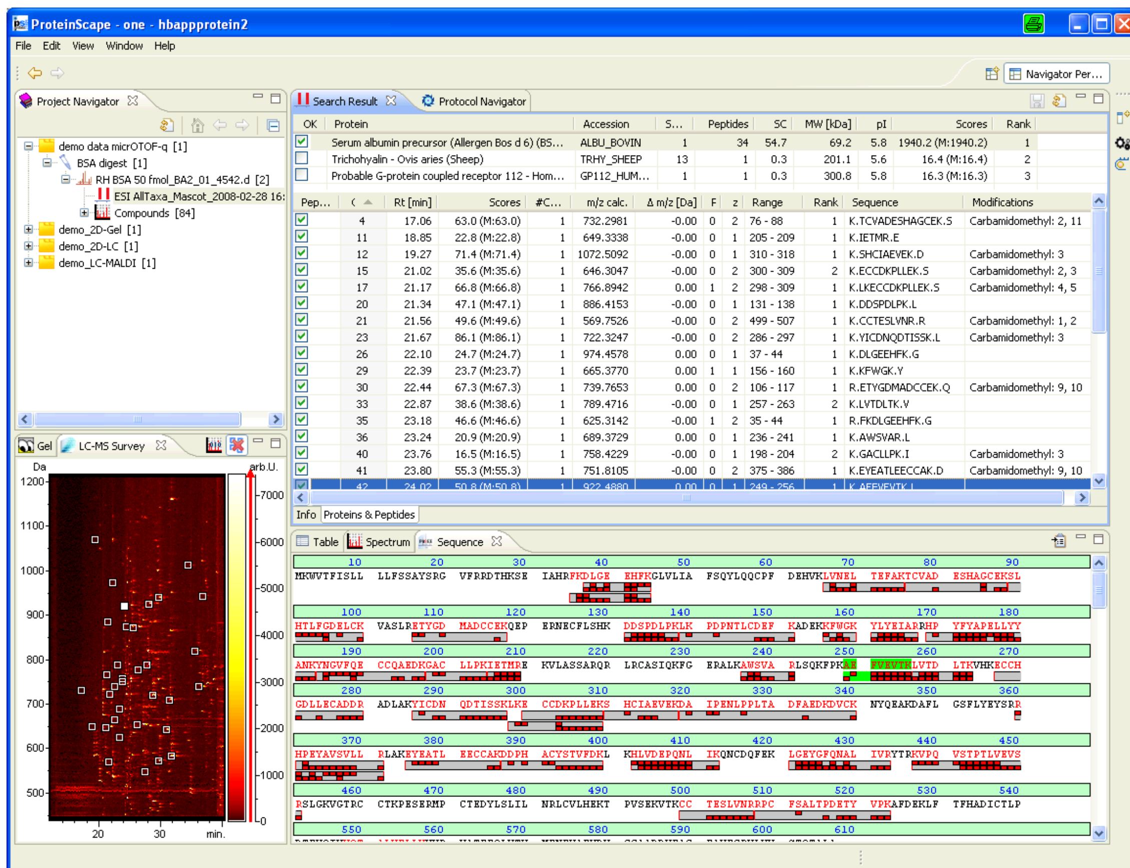


Figure 3-2 ProteinScape perspective after database search. The parameters shown in Figure 3-1 were used