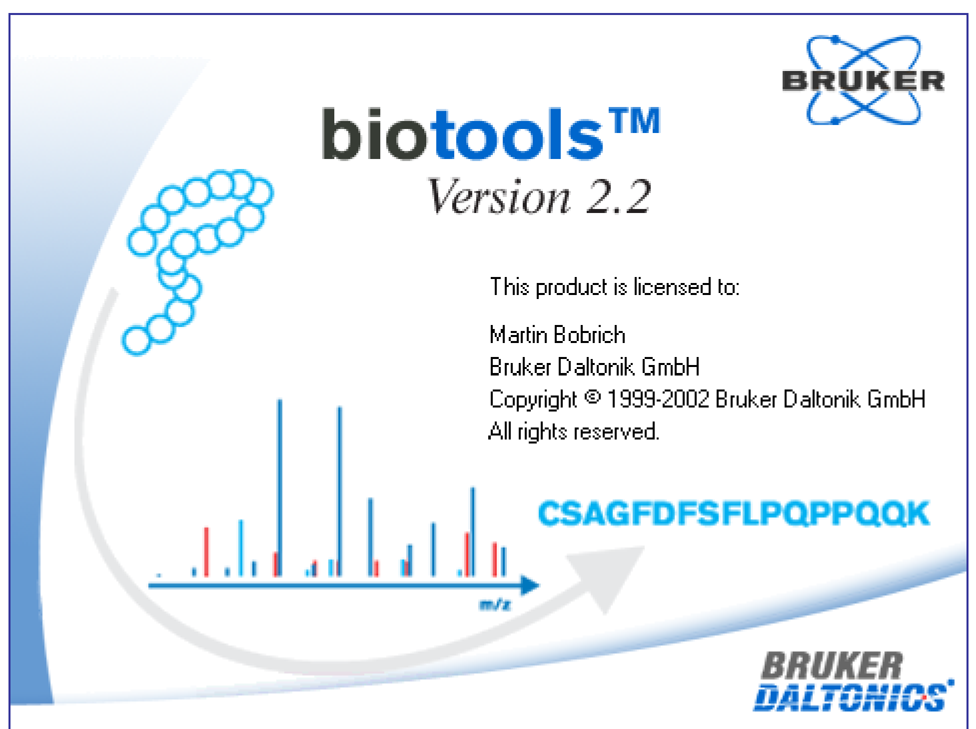
BioTools User Manual



(August 2002)

Copyright

Copyright 2002

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

BioTools User Manual, Version 2.2 (August 2002)
Part #: 216936
First edition: October 2000
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.

Copyright:

Bruker Daltonik GmbH
Fahrenheitstrasse 4
28359 Bremen
Germany

Phone: +49 (4 21) 22 05-200
FAX: +49 (4 21) 22 05-103
Email: mailto:sales@bdal.de
Internet: http://www.bdal.de/

Contents

1 INSTALLATION 1-1
1.1 System Requirements 1-1
1.2 Program Setup 1-1
1.3 Starting the Installation 1-1
1.4 Starting the Program 1-8
1.5 License Manager 1-8

2 QUICKSTART 2-1
2.1 Loading processed data 2-2
2.2 Loading Spectra 2-7
2.2.1. Data from Bruker XMASS / XTOF 2-7
2.2.2. Data from Bruker DataAnalysis 2-8
2.2.3. Display of Picked Peaks 2-9
2.3 Processing of MS Data 2-10
2.3.1. DeNovo Sequencing 2-10
2.3.2. Mascot Database Query (MS/MS) 2-14

A APPENDIX: AMINO ACID RESIDUES AND FRAGMENTATIONS A-1
A.1 Amino Acid Residues A-1
A.1.1. Single letter code A-1
A.1.2. Genetic Code .... A-A
A.1.3. Formulas and Molecular Weights A-4
A.1.4. Chemical Structures A-5
A.2 Peptide Fragmentation A-6
A.3 Menu and Shortcut list A-11
A.4 Toolbar Reference list for BioTools A-14
A.5 Toolbar Reference list for SequenceEditor A-16
A.6 Part Numbers A-16

QUICK GUIDE TO OPERATE BIOTOOLS AND SEQUENCEEDITOR

TUTORIALS FOR BIOTOOLS

TUTORIALS FOR SEQUENCEEDITOR (only available in the pdf-file!)

BIOTOOLS REFERENCE MANUAL (only available in the pdf-file!)

SEQUENCEEDITOR REFERENCE MANUAL (only available in the pdf-file!)

Table of changes

Version	***Date***	***Changes***
2.0	2000-10-14	Software versions: BioTools, Version 2.0 SequenceEditor, Version 1.0
2.1	2002-01-18	Software versions: BioTools, Version 2.1 SequenceEditor, Version 2.1
2.2	2002-08-01	Software versions: BioTools, Version 2.2 SequenceEditor, Version 2.1

1 Installation

This chapter contains information on the hardware and software requirements for running BioTools. It also describes the installation procedure and how to start the program from WINDOWS Program Manager.

1.1 System Requirements

- CPU: Intel Pentium II processor or better.
- Clock: 266 MHz processor for satisfying data handling.
- Main Memory: Minimum 64 Mbytes RAM or better.
- Operating System: Microsoft Windows-NT version 4.0 or higher (Service Pack 3 or better), installed and operating Internet Explorer 4.0 or better.
- Graphic Resolution: 1024 * 768 pixel, 256 colors or better
- CD–ROM drive (4X or better)
- Ethernet connection
- Hard disk: at least 100 Mbytes of free disk space.

1.2 Program Setup

First make sure that Windows NT Version 4.0 with Service Pack 3 and the Microsoft Internet Explorer 4.0 or better is installed on your computer system. We do not recommend proceeding beyond this point when your system does not run this software.

1.3 Starting the Installation

Put the installation CD-ROM into the drive (e. g. D:). If the "Autostart" function is activated, the installation program will start automatically. Otherwise click to the WINDOWS "Run..." element.

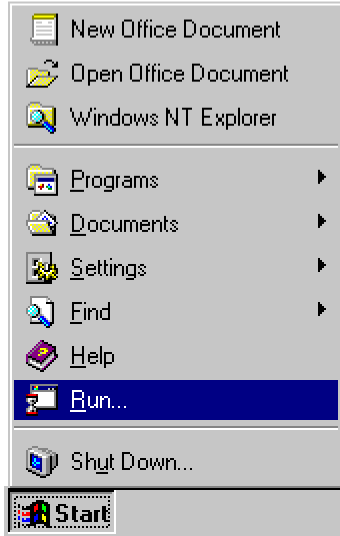


Figure 1-1, The WINDOWS „Run..." element

The dialog box shown in the next figure appears. The installation CD must be inserted into the appropriate drive (e.g. "D") and the command line "D:\SETUP.EXE" typed in. The installation system then starts as soon as the *Ok* button of the dialog box is activated.

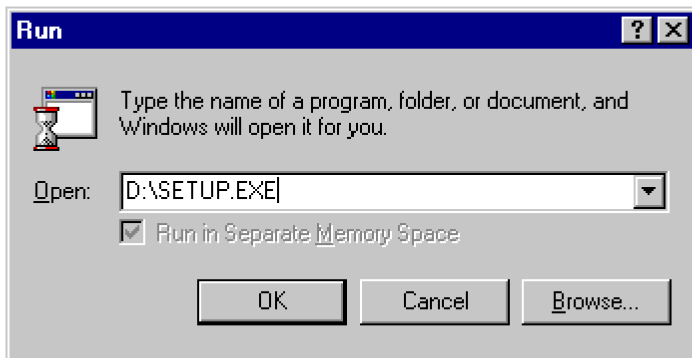


Figure 1-2, Start of the Setup program

The installation starts with a "Welcome" message followed by the software license agreement. Next follow the instructions on the screen: enter your user information, destination directory and select the components to be installed on your computer.

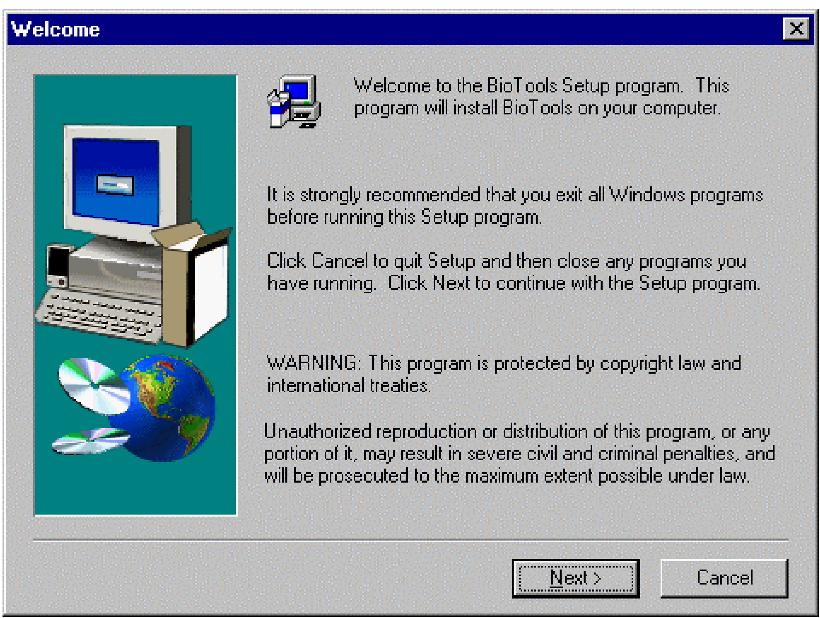


Figure 1-3, Welcome to the BioTools Setup Program

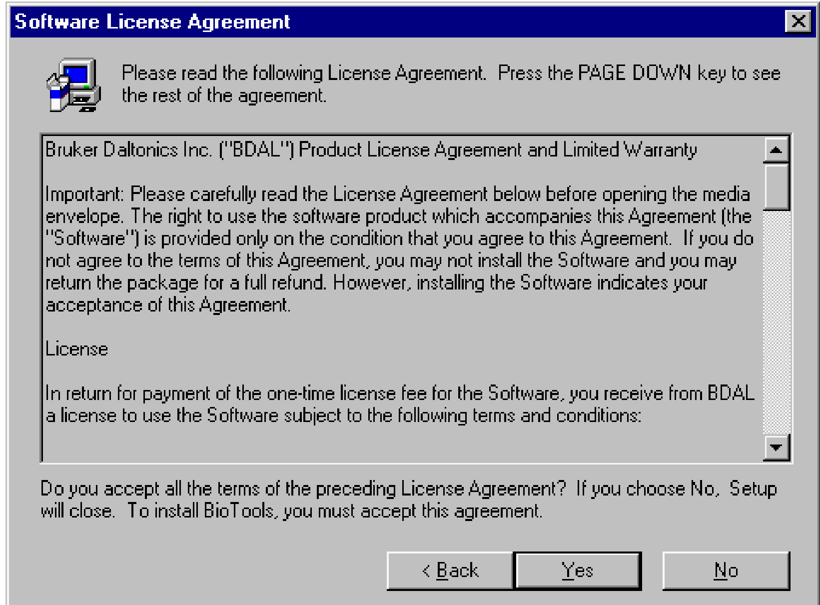


Figure 1-4, Software License Agreement

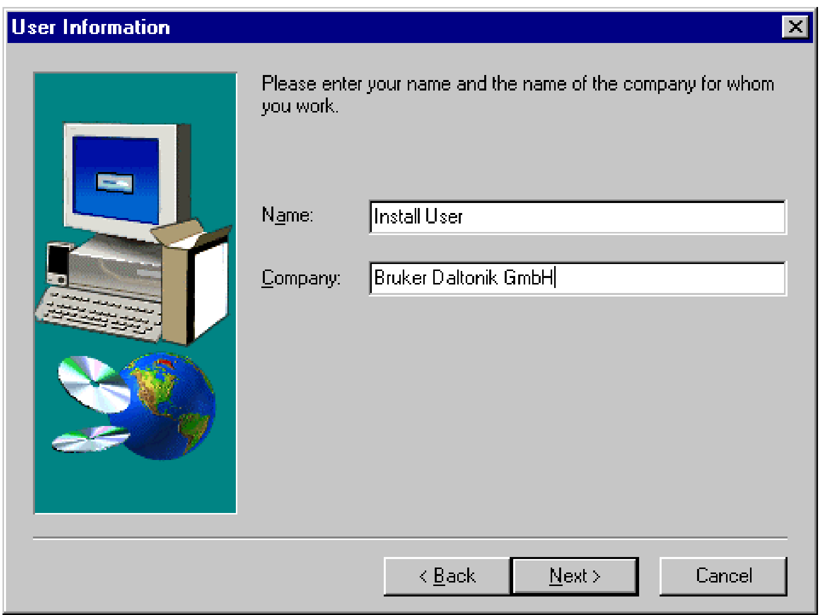


Figure 1-5, User Information

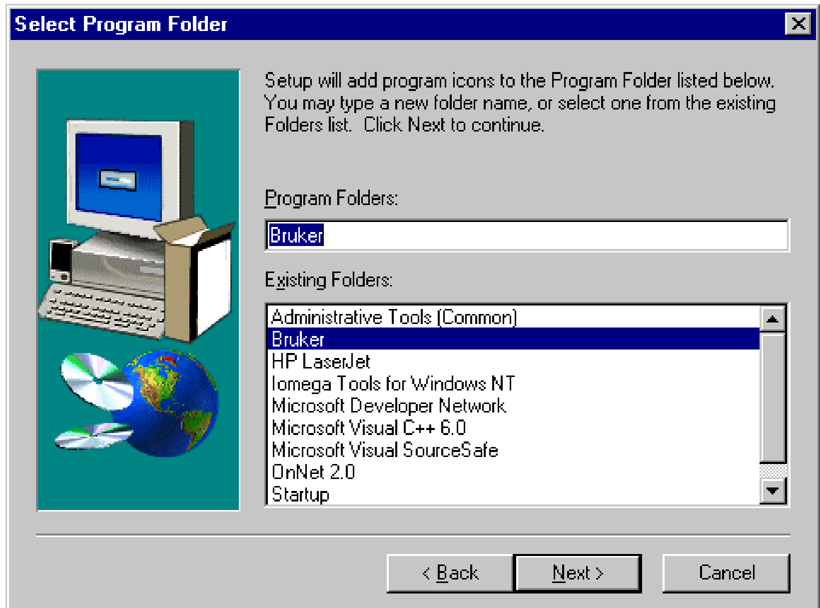


Figure 1-6, Select Program Folder

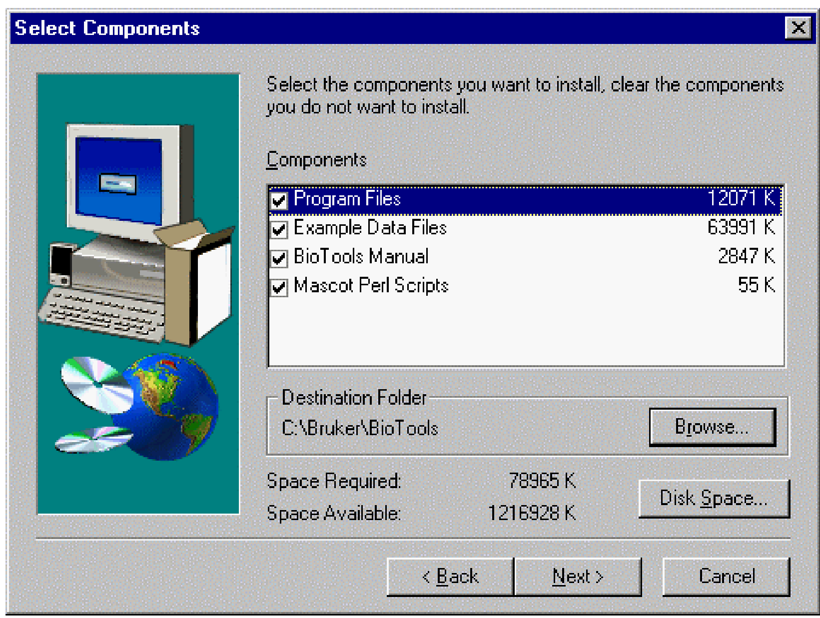


Figure 1-7, Select Components

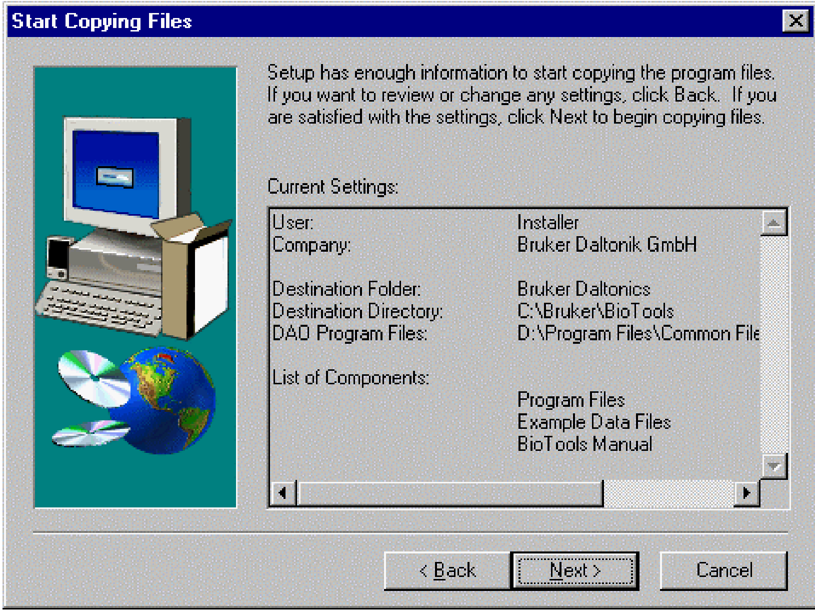


Figure 1-8, Start Copying Files

The entry "Mascot Perl Scripts" is only available if the Mascot Database is installed on your computer. Normally you will use BioTools and Mascot on different machines.

In this case you must perform the BioTools setup on your Mascot server also, but you have to select the component "Mascot Perl Scripts" only.

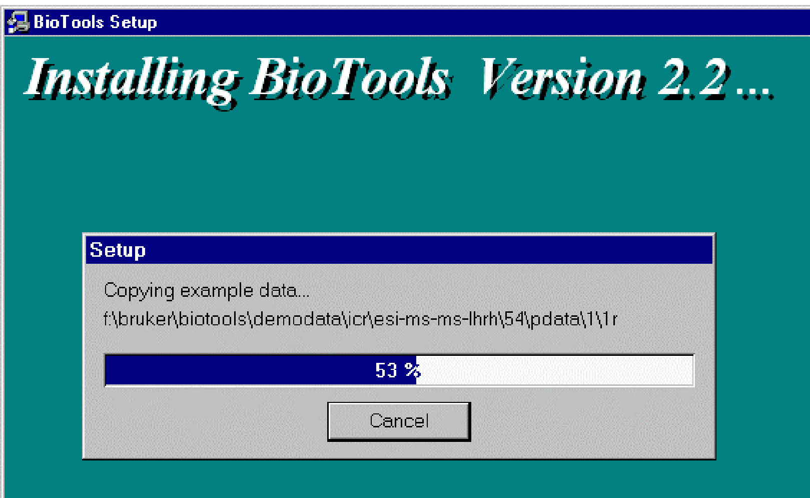


Figure 1-9, Installing BioTools

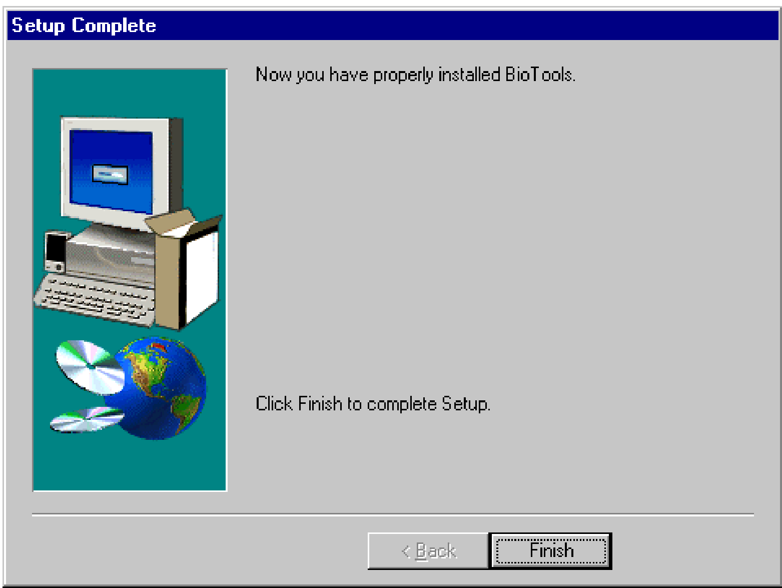


Figure 1-10, Setup Complete

On the other hand you can copy the required Perl scripts by hand from the distribution CD of BioTools (see folder "BioTools Perl Scripts for Mascot") into the Mascot directory (normally "inetpub\Mascot\cgi").

Important Note for use of remote Mascots servers:

Normally you will use BioTools and Mascot on different machines. In this case you must prepare the mascot server after finishing the BioTools installation:

- Either you can perform the BioTools setup on your Mascot server, but you have to select the component "Mascot Perl Scripts" only.
- Or you can copy the Perl scripts manually from the distribution CD of BioTools (see folder "BioTools Perl Scripts for Mascot") into the Mascot directory on the server computer (normally "inetpub\Mascot\cgi").

If you decided to install the BioTools manual and the Acrobat Reader software is not installed yet, you will find an icon "Install Acrobat Reader" in your BioTools program group. Use this icon to install the Acrobat Reader software.

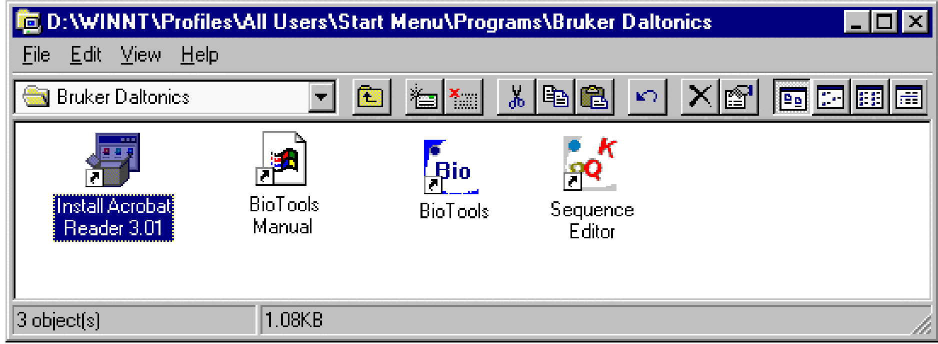


Figure 1-11, Acrobat Reader Program not yet installed

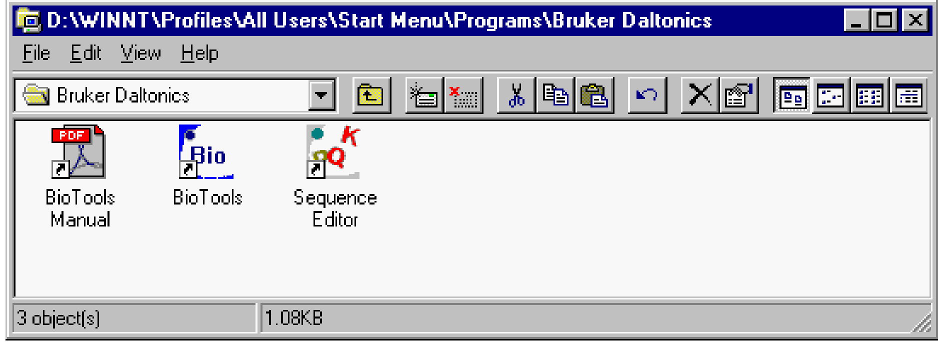


Figure 1-12, After installation of the Acrobat Reader Program

1.4 Starting the Program

Use the BioTools Icon from the program group you specified during program installation. If you receive the error message "A procedure entry point httpsendrequestExA could not be located in the dynamic linked library" during program start, the Microsoft Internet Explorer 4.0 or better is not installed on your system. The installation of the Microsoft Internet Explorer 4.0 or better will replace an existing WININET.DLL by a newer one.

1.5 License Manager

If you start BioTools the first time, you must enter a license key. The key comes together with the BioTools documentation. Use the license manager to add or remove licenses (Menu Help – License Manager).
If an invalid license key is entered, the program can not be used.

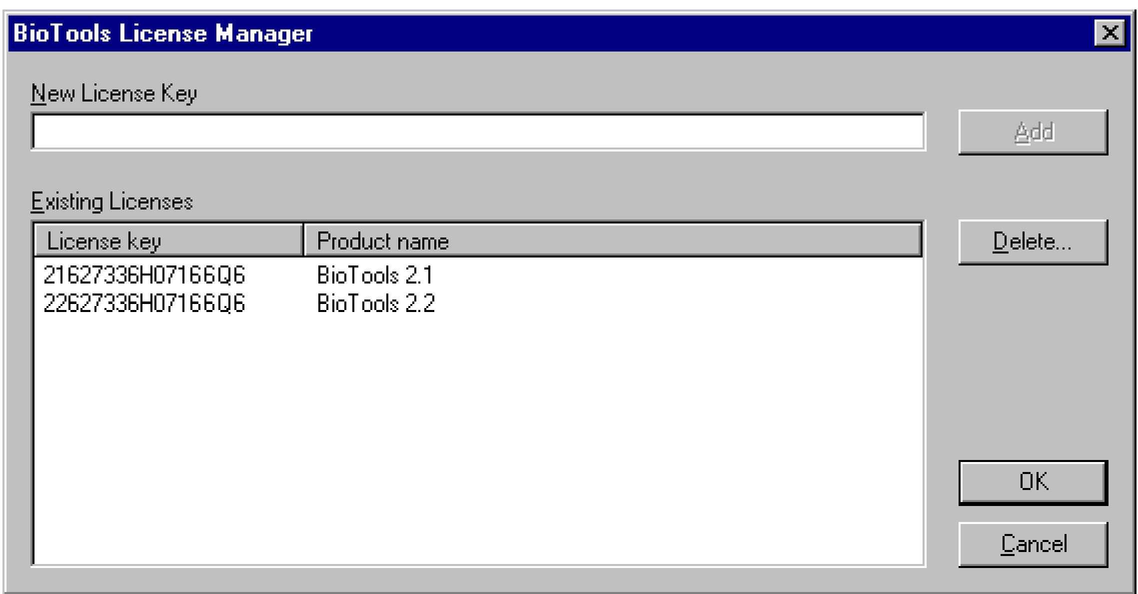


Figure 1-13, License Manager

2 Quickstart

BioTools uses processed data from different Bruker software packages. These data are spectra and peaklists, which must have been created in advance. BioTools will use the data to perform a *DeNovo* sequencing or an Internet search based on the peaklist. As a result you will receive one or more amino acid sequence which will be matched with the experimental spectrum. The best match can be used to annotate the spectrum.

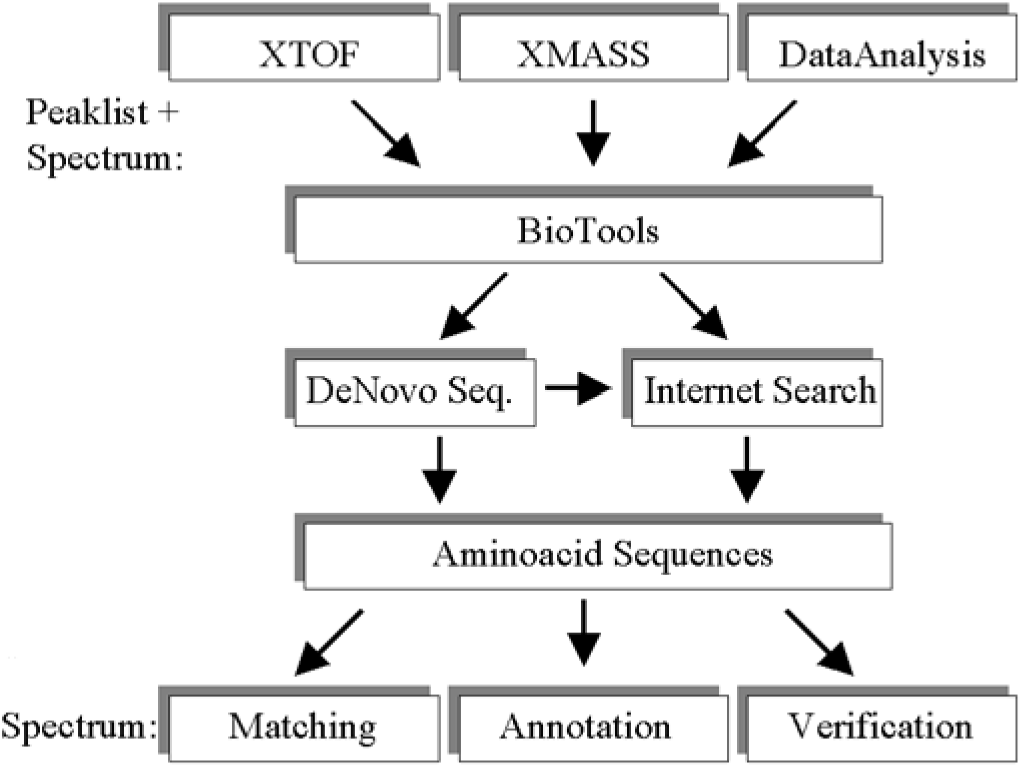


Figure 2-1, BioTools data flow

2.1 Loading processed data

To open an already processed data set, use the option "Find" from the "File" menu . The default file name is "1R" (this is processed data from XMASS/XTOF). The path after "Look in" is the folder where BioTools was saved during program installation. Subfolders will be included for the search also because the button "Include subfolders" is checked. Next click to button "Find Now" to start the search.

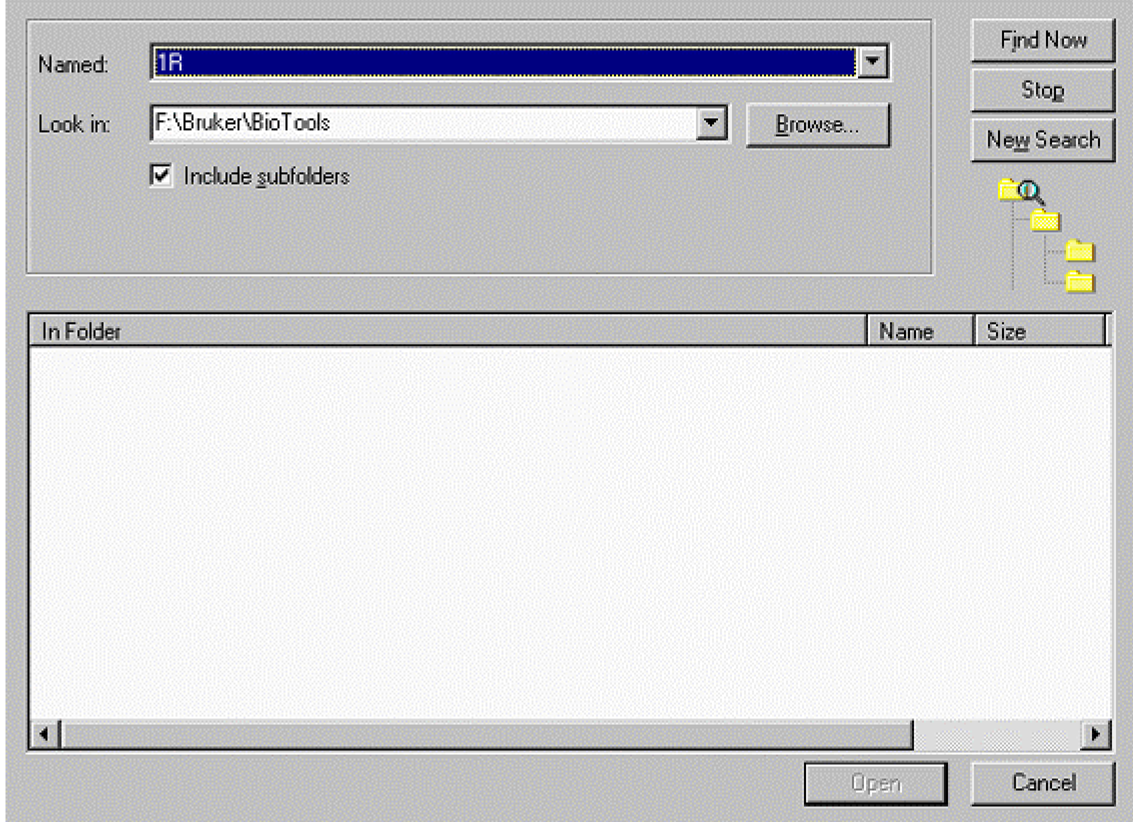


Figure 2-2, Find files

From the list of found files select the entry "ACTH1-17.FAST" and load it using button "Open". This is a MALDI-PSD spectrum and therefore an example for the analysis of MS/MS data.

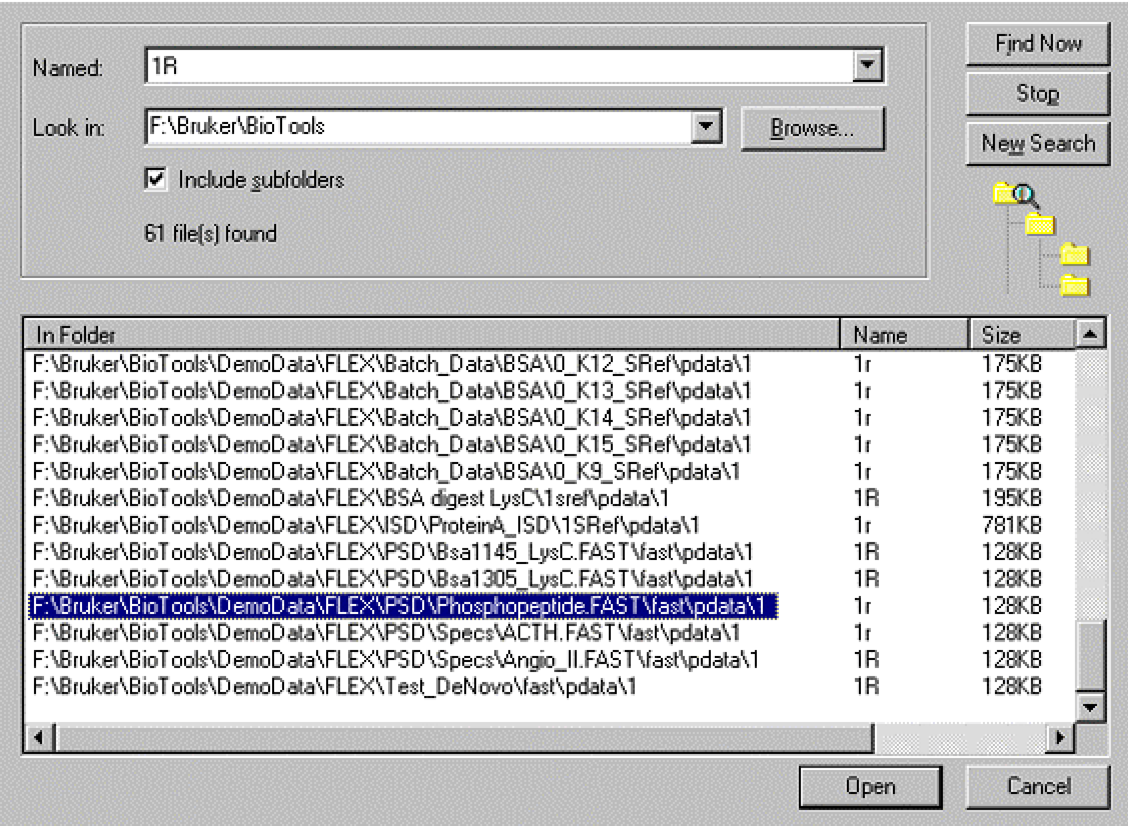


Figure 2-3, Found files

The result is shown in the next figure (in this example the background color was set to white; the default background color after program installation is gray).

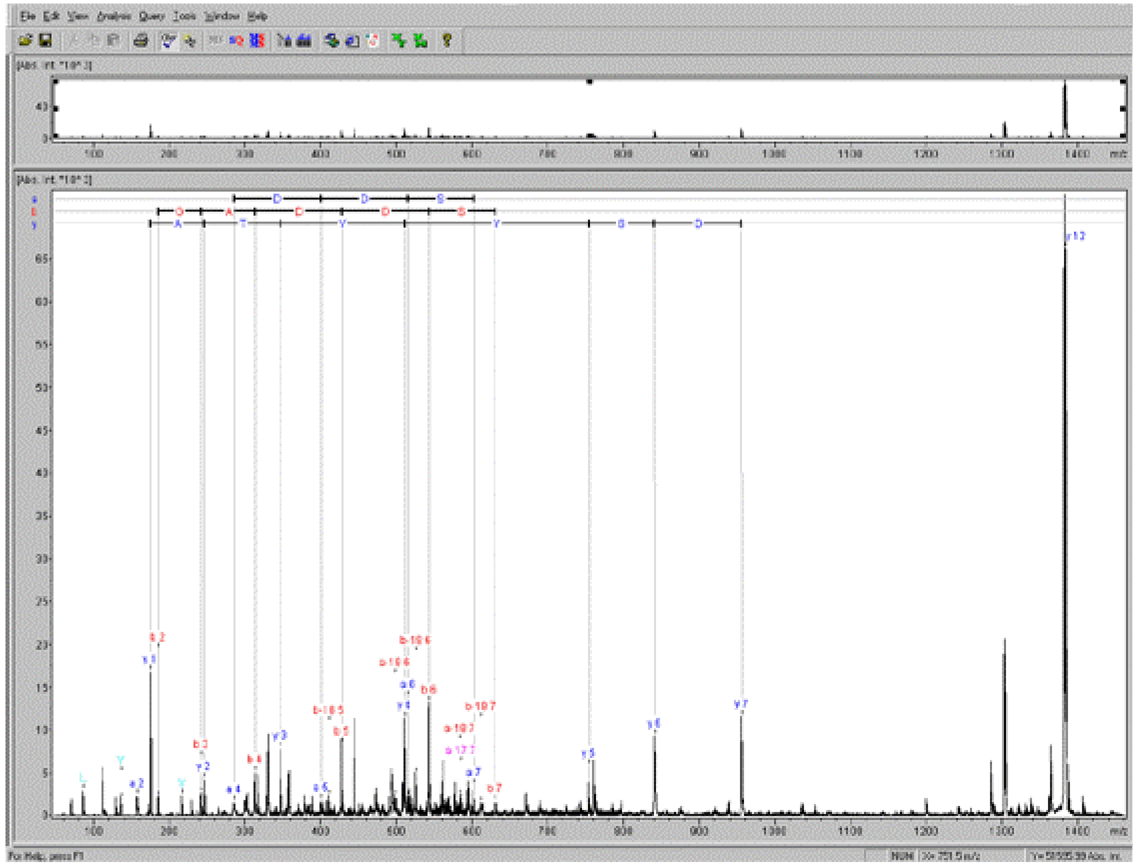


Figure 2-4, Reloaded processed data

The upper slice shows the overview window: this is always the complete spectrum. The lower slice shows the selected range of the spectrum. Use the mouse to zoom in the displayed spectrum: move the mouse cursor into the region of the spectrum window, press the left mouse button, move the mouse and finally release the mouse button. The lower slice now shows a part of the spectrum, the upper slice shows the complete spectrum and a rectangle around the zoomed area.

There are different ways to move within the spectrum:

- Move the mouse pointer into the overview window into the marked rectangle, press the mouse button, move the rectangle with your mouse to an area of interest and release the mouse button. The zoomed area will be updated immediately.

- Move the mouse pointer into the area of the x-or y-axis, press the left mouse button and move the mouse. Spectrum window and overview window will be synchronized.

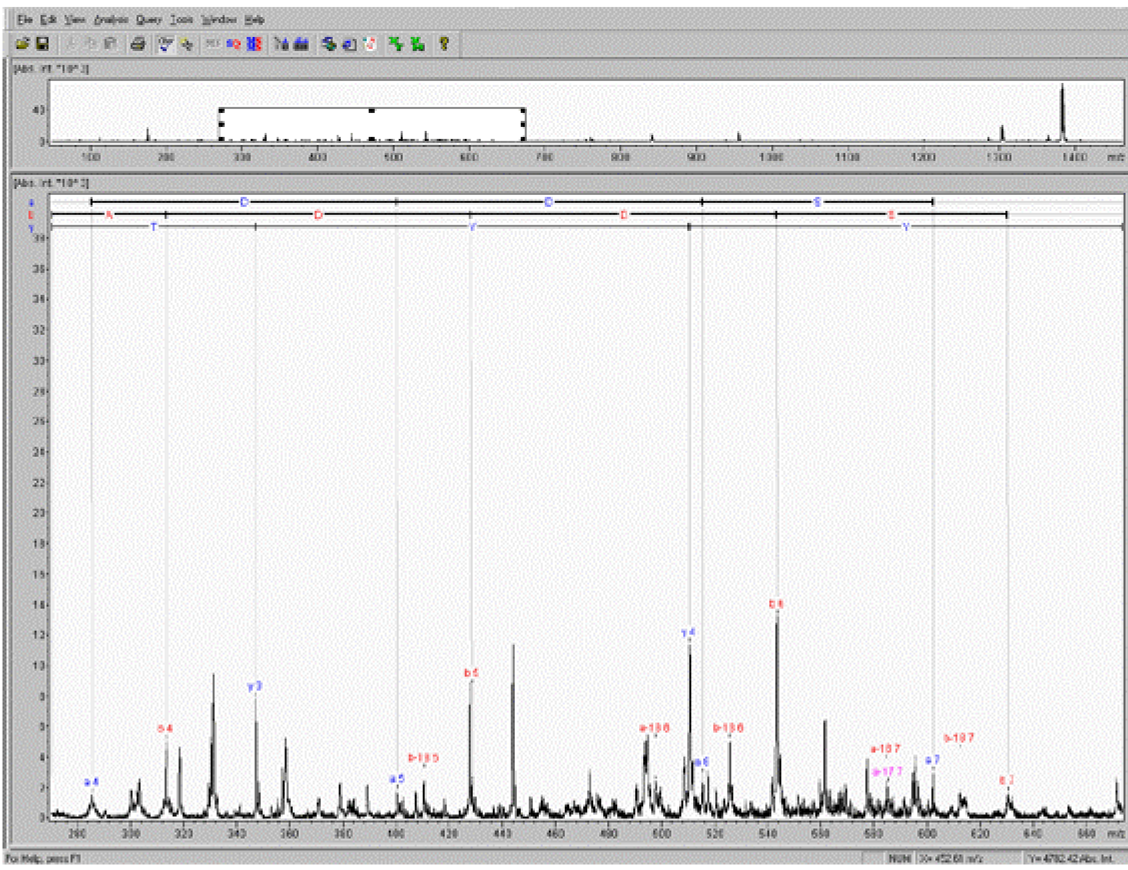


Figure 2-5, Zoomed spectrum

A double click with the left mouse button within the spectrum window will reset the display (a more detailed description of all features of this window will be given later in this manual).

To change the ratio between overview window and spectrum window move the mouse cursor to the small area between both windows. The shape of the cursor will change to an up/down arrow if you have reached the right position. Now press the left mouse button and move the mouse upward to hide the overview window.

Probably you may wonder where the information of annotation comes from. From the toolbar select the "Show/Hide Fragments" button .

The display will be split into two parts: the upper still shows the annotated spectrum, the lower gives information about the sequence and calculated fragments. The text field to the right of the "Sequence" button, displays the amino acid sequence (optionally followed by an additional comment). This sequence is used to calculate the sum-

formula, mono-isotopic, average parent mass and fragment ions (there will be a detailed description how these fragments are calculated later in this manual). The "Sequence" button allows editing of the sequence and of modifications using the ***SequenceEditor***, which is described in the respective manual. Calculated fragment ions are matched to picked peaks in the spectrum . In case of a **match** the single letter symbol of the amino acid and the corresponding mass is **displayed in red color**, **otherwise in gray color**. Matching takes place with a given tolerance of permissible differences between calculated and found peak masses. A threshold may be set to ignore small peaks from background noise.

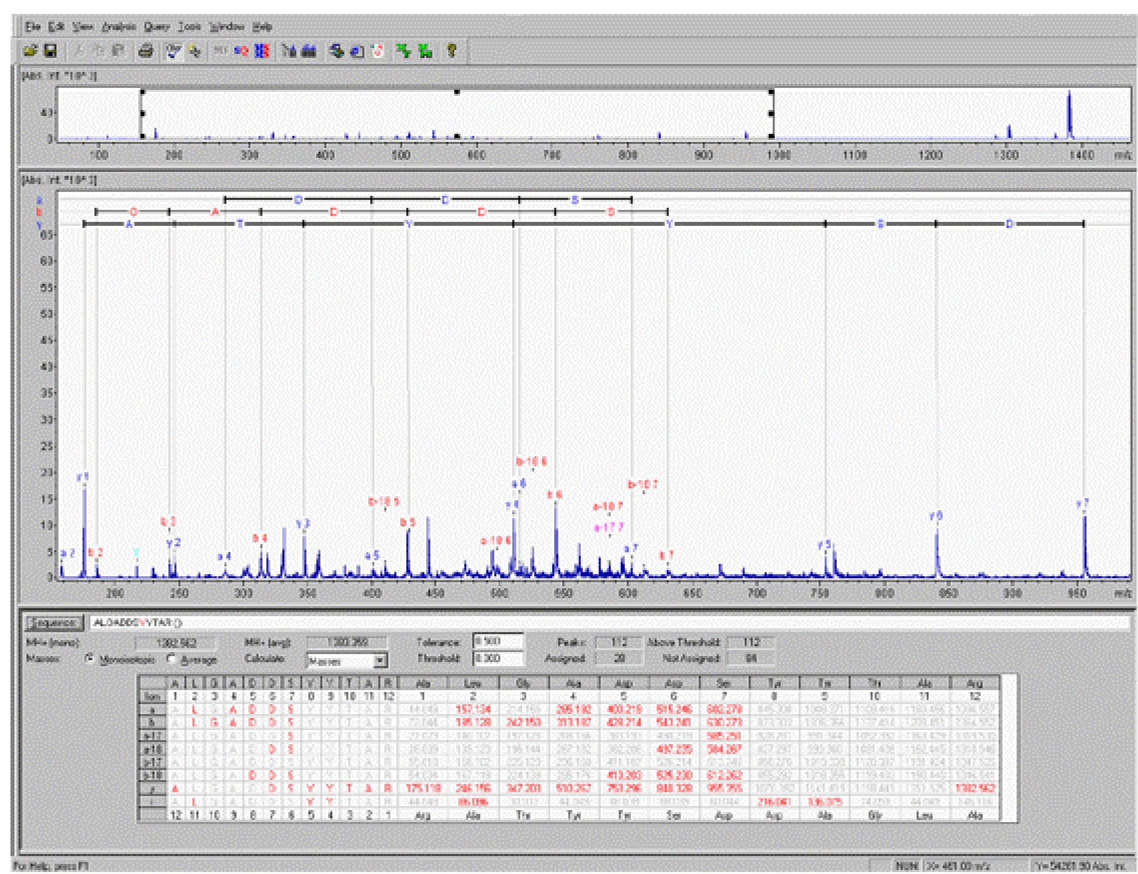


Figure 2-6, Annotated spectrum and calculated fragments

2.2 Loading Spectra

BioTools uses data that has already been processed either by Bruker XMASS/XTOF (Version 5.1 or higher) or Bruker DataAnalysis (Version 2.0 or higher).

2.2.1. Data from Bruker XMASS / XTOF

2.2.1.1. XTOF Windows NT

Using the Windows NT version of XTOF is the easiest way to receive data for processing with BioTools. Start XTOF, open a spectrum, perform peak picking using "label region" and then click the button "Bio" to start the program BioTools. BioTools will automatically load the corresponding spectrum together with the picked peaks.

2.2.1.2. XTOF Unix

Open a spectrum and perform peak picking using "label region". Use the "Bio" button to transfer the data back to BioTools.

For XTOF Unix 5.1 and higher: Use the "Bio" button to store the current peak list as an XML-file.

For XTOF Unix 5.0 and lower: Use the **P2BT** command from the command line level in order to save the peak list in all formats required for the full functionality available in BioTools. If you do not find the command on your system please write an email to DSU@bdal.de and specify your XTOF version. We will email the command to you.

The peak list is then automatically loaded into BioTools together with the spectrum when it is opened manually.

2.2.2. Data from Bruker DataAnalysis

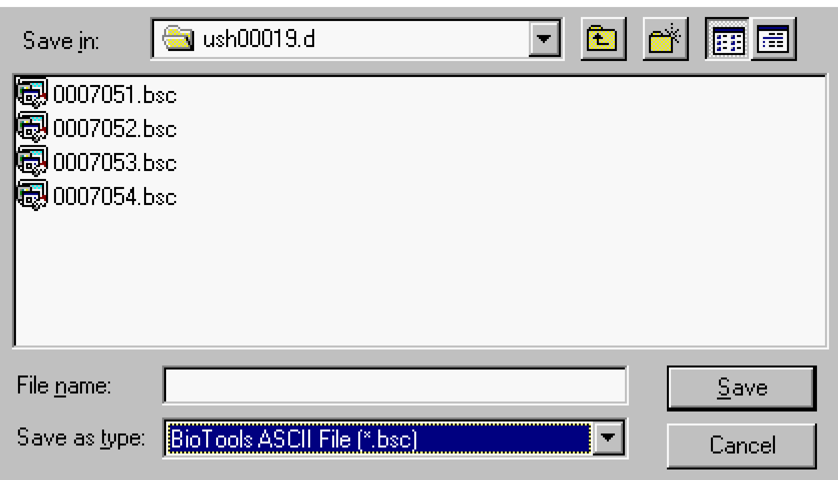


Figure 2-7, Export of data from Bruker DataAnalysis

To export data from DA 2.0 and higher, use "Export – Mass Spectrum" from the "File" menu. The data format must be "BSC". Export the data into the same folder where the other files of the data set are stored.

More detailed information is in the tutorial, Using BioTools for esquireSeries data .

2.2.3. Display of Picked Peaks

To display peaks that have been previously picked by Bruker XTOF or Bruker DataAnalysis 2.0 move the mouse into the spectrum window of BioTools, press the right mouse button and check the entry "Picked Peaks". Picked peaks are displayed as red histogram bars in the spectrum.

To hide the picked peaks uncheck this entry.

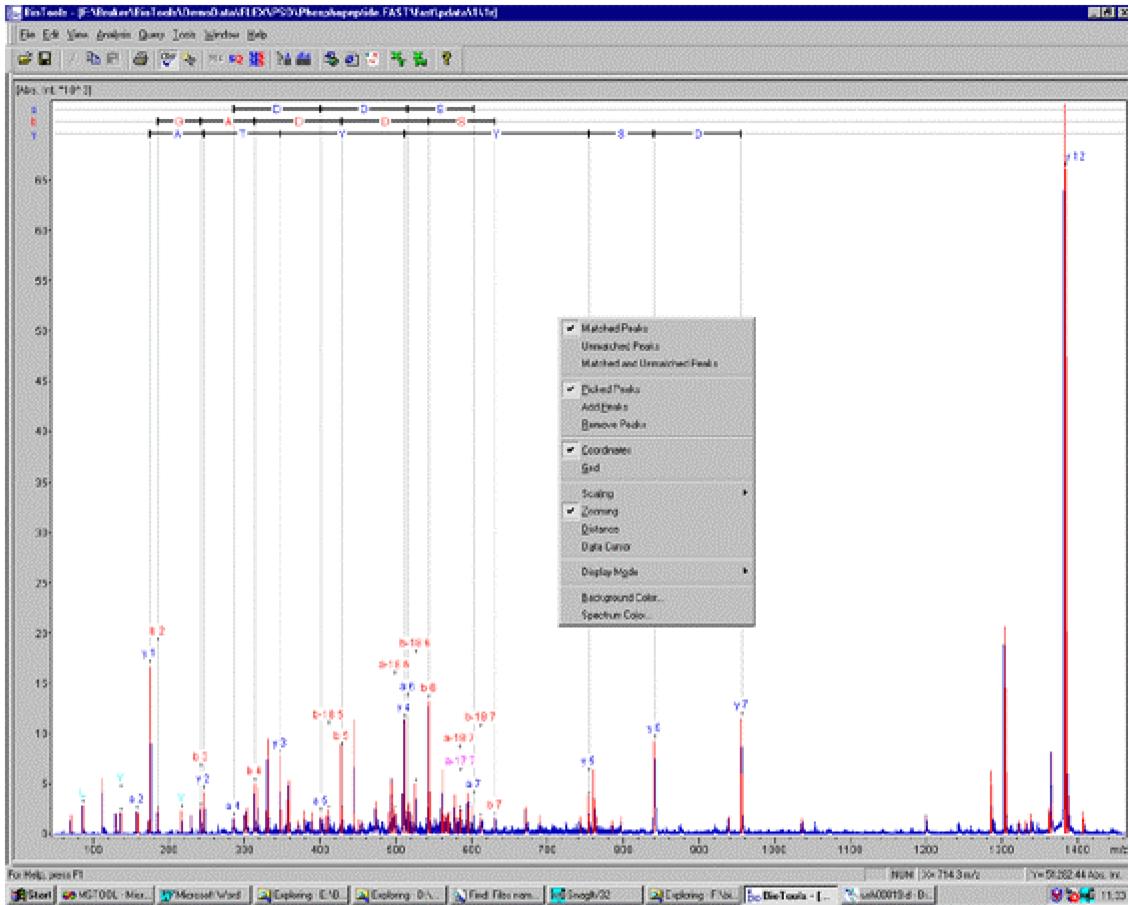


Figure 2-8, Spectrum with picked peaks

2.3 Processing of MS Data

2.3.1. DeNovo Sequencing

To open a data set, use again the option "Find" from the "File" menu. If you did use this option previously the list of files that were found before is displayed again. Otherwise set entry "Named" to "1R". The path after "Look in" should be the folder where BioTools was saved during program installation. Subfolders must be included for the search (check button "Include subfolders"). Next click to button "Find Now" to start the search. From the list of found files select the entry "Test_DeNovo.FAST" and load it using the "Open" button.

Next select button "*DeNovo* Sequencing", button "Show/Hide Treeview" (if the treeview is shown) and button "Show/Hide Fragments" (if the calculated fragments are not shown). The result is shown in the next figure.

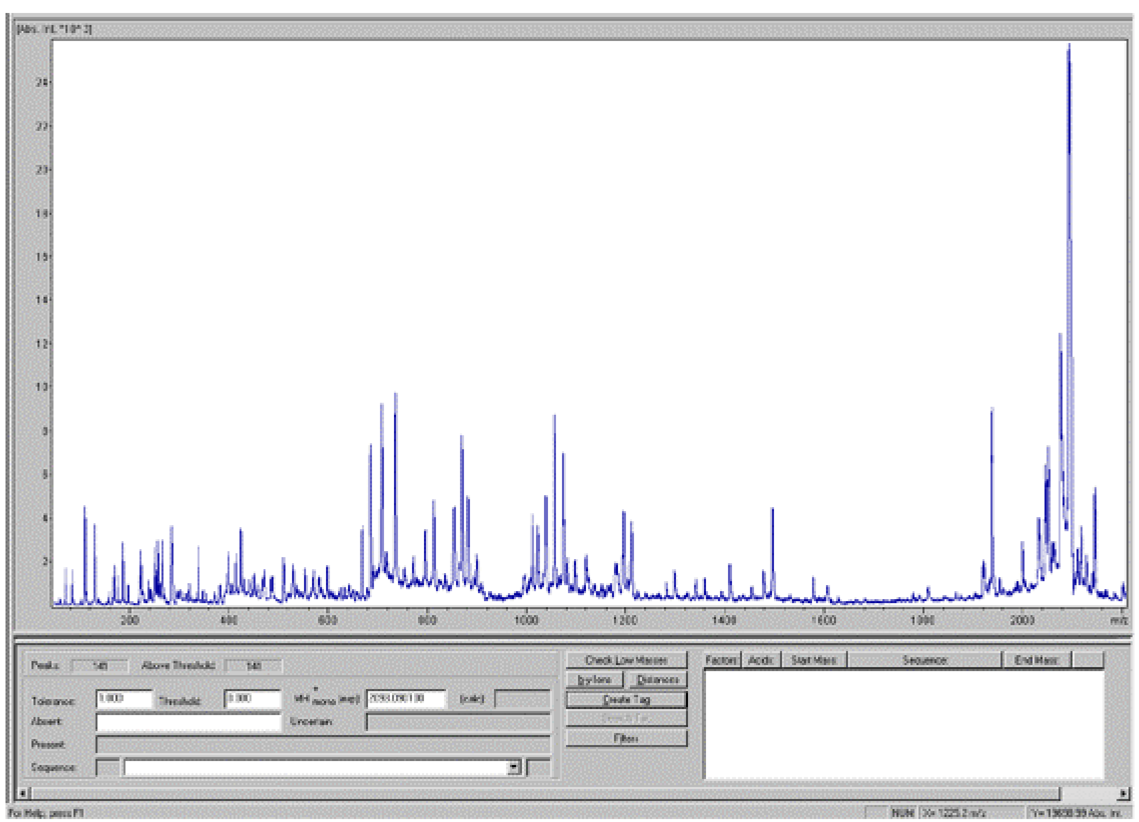


Figure 2-9, Start of *DeNovo* sequencing

In the "Absent" text field type in "**I Q**" if you want to avoid redundant sequence tag suggestions for the isobaric residue pairs I and L, and K and Q. Click the buttons "Create Tag" and then "Search Tag".

Note ***The search with sequence tags in PeptideSearch is only useful when a MASCOT search failed and the assumption is, the reason may a modification in the peptide. Peptide search allows for error-tolerant searches by reduction of the required match regions (standard: 1 and 2 and 3). 1 and 2 allows an error C-terminal of the sequence tag, 2 and 3 N-terminal of the tag and 1 and 3 allows an error within the tag sequence.***

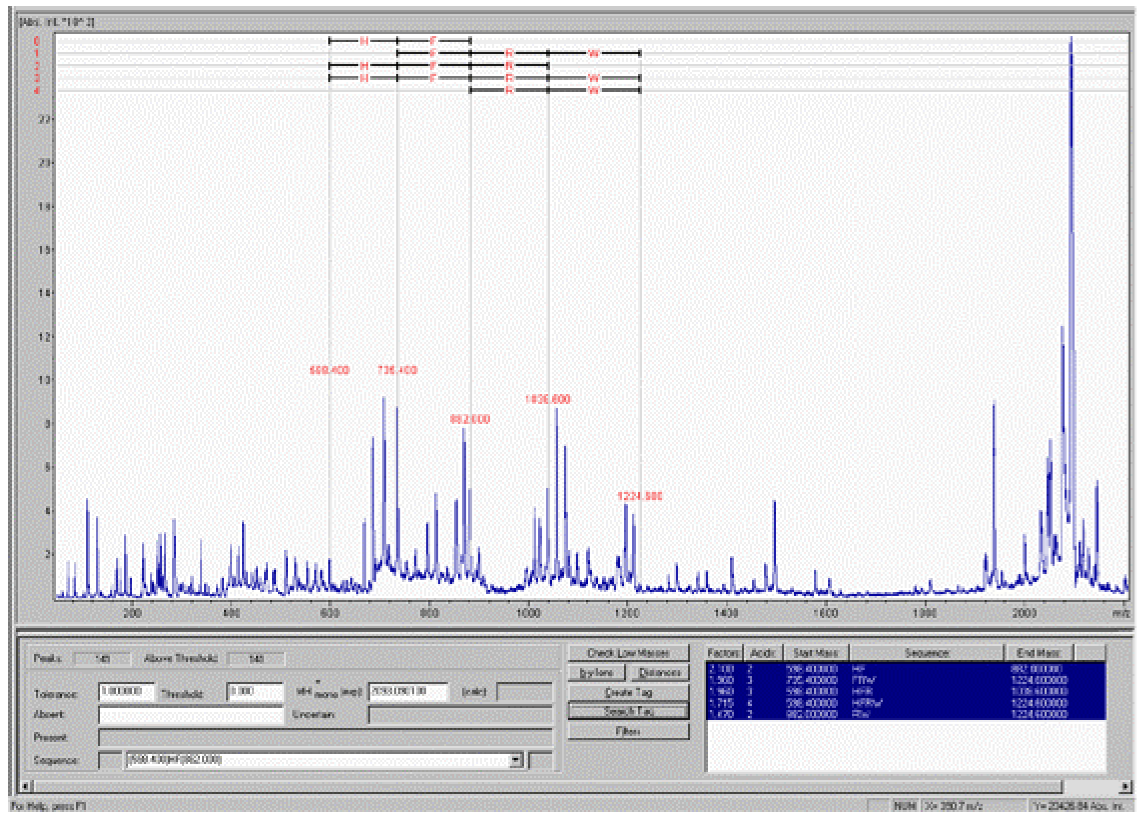


Figure 2-10, Create Tag result

URL: http://193.175.249.95:80/CGI/PPG.PeptSearchSequenceTags.acgi

Protein mass range from (kDa): 0 to (kDa): 300

Cleavage agent: Trypsin

Cysteine is: Cys

Oxidized Methionine

Peptide mass (neutral): 2092.082275 Monoisotopic mass

Mass accuracy: 0.3 Da

Peptide sequence tag: (598.400)HF(882.000)

Match regions: 1 and 2 and 3

Pattern match search by: B-type sequence ions

Edman type search by: N-terminal sequence extension

Allowed number of errors: 0

Cleavage specifities: N-terminal specifity C-terminal specifity

Results per page: 100

Copy Save Start Exit

Figure 2-11, Sequence Tag search parameter

The "Sequence Tag Search" dialog shows search parameters and the "best" found sequence tag resulting from the peak list. A click to the button "Start" will connect you to the Internet (provided your Internet connection is installed properly). After some time the results of the query will be displayed in a window of the Microsoft Internet Explorer. Use the button "Get Results" (on the right side of the result window) to transfer the query results to your PC. The obtained sequences (in this example there are 25 hits, however all found sequences are identical) will be added to the treeview on the left side of the screen. Leave the result window (use the "Exit" button), close the "Sequence Tag Search" dialog and switch back to the "Check Sequence" mode . You will receive the annotated spectrum.

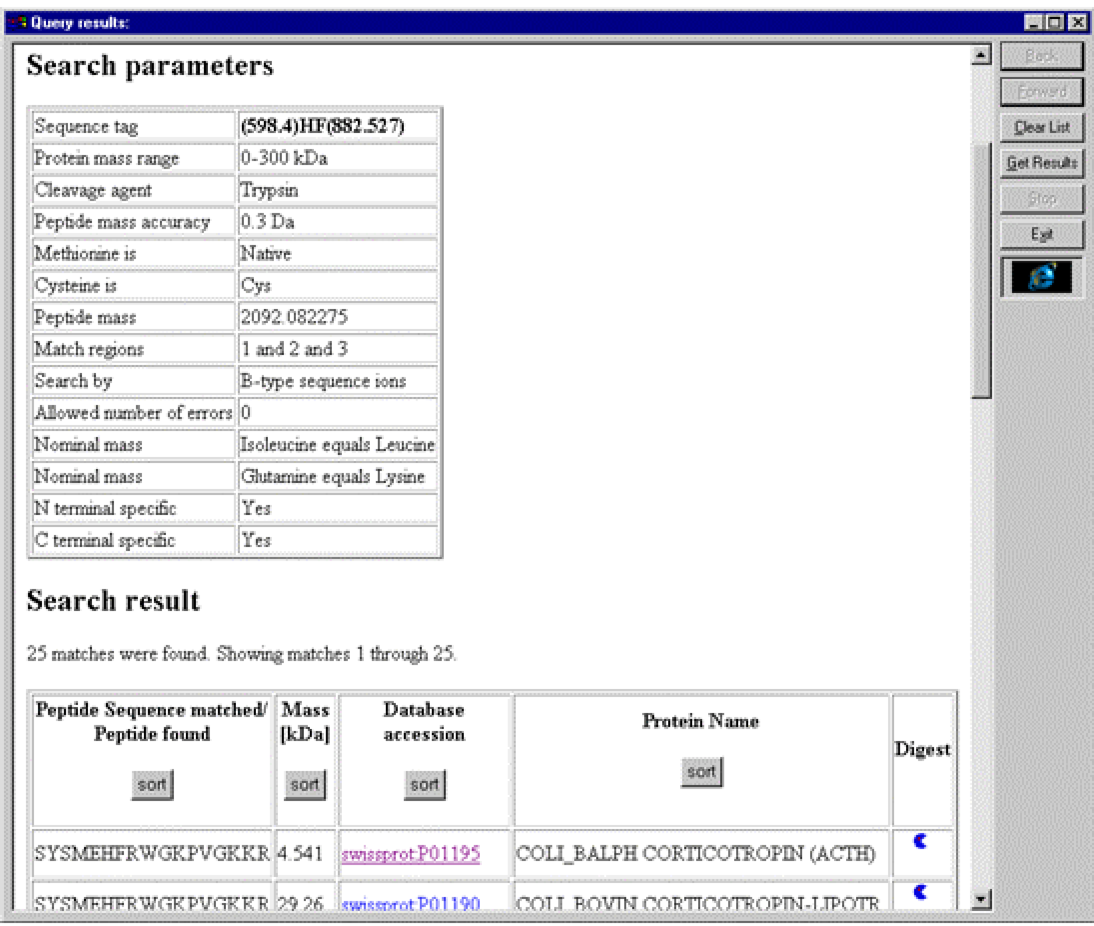


Figure 2-12, Sequence Tag search results

2.3.2. Mascot Database Query (MS/MS)

In all cases, in which the presence of reference sequence information in protein sequence databases can be expected, it is useful to perform a sequence database search. To open a test data set, use the option "Find" from the "File" menu. If you did use this option previously the list of files that were found before is displayed again. Otherwise set entry "Named" to "1R". The path after "Look in" should be the folder where BioTools was saved during program installation. Subfolders must be included for the search (check button "Include subfolders"). Next click to button "Find Now" to start the search. From the list of found files select the entry "Bsa1305_LysC.FAST" and load it using the "Open" button. Next select button "Check Sequence", button "Show/Hide Treeview" (if the treeview is not shown) and button "Show/Hide Fragments" (if the calculated fragments are not shown). The result is shown in the next figure.

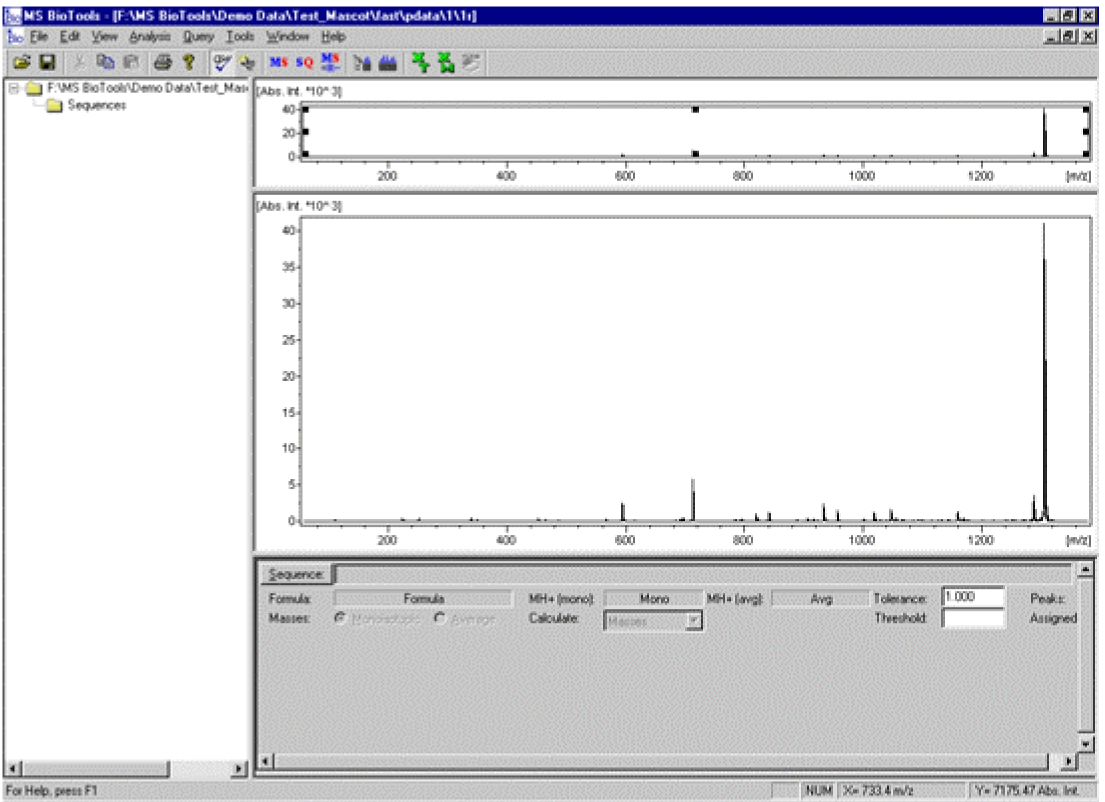


Figure 2-13, Start of Mascot search

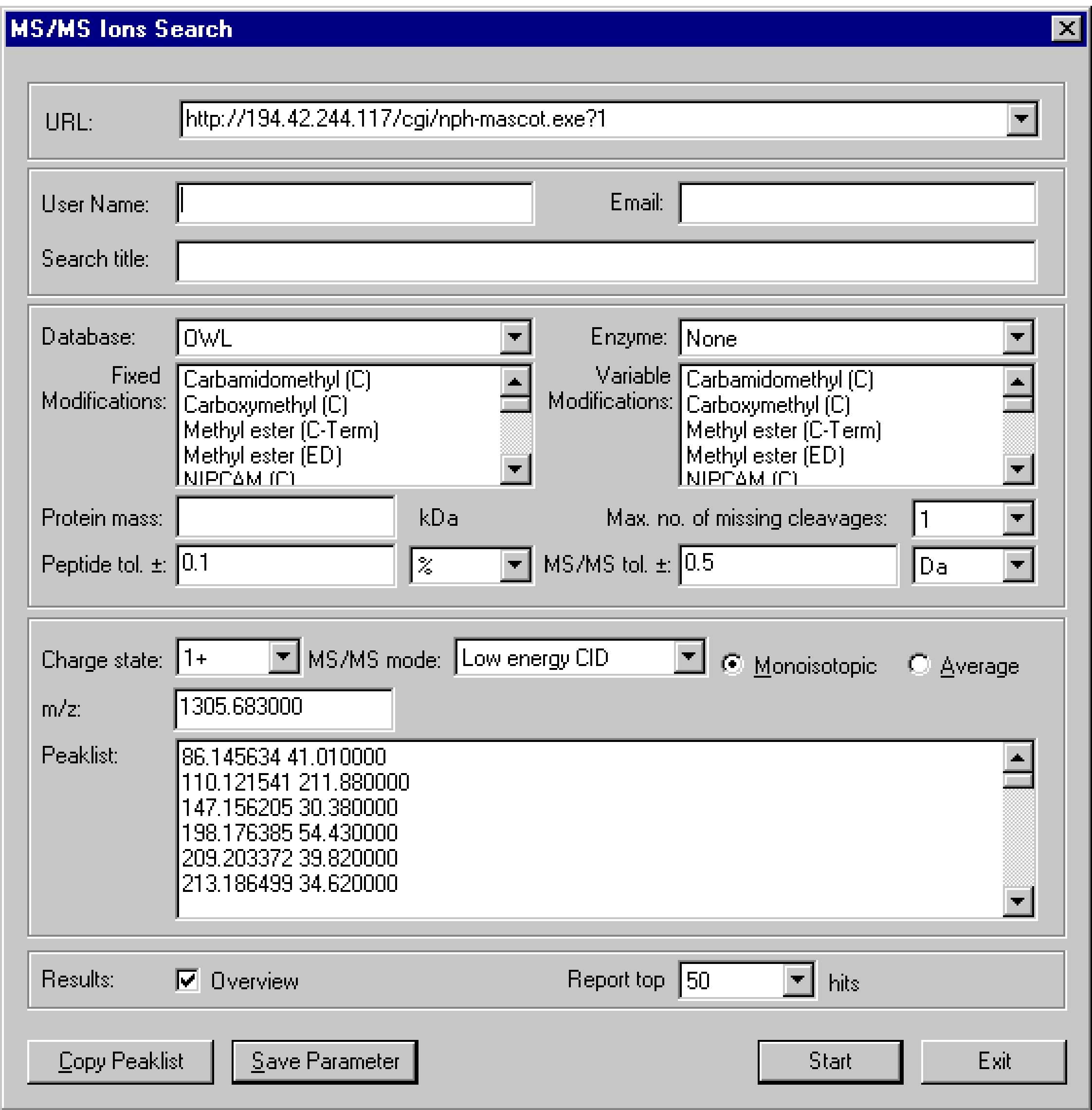


Figure 2-14, MS/MS Ion Search Parameter

Next click to the button "Mascot MS/MS Ion Search" and start the query using the "Start" button with the parameters: Enzyme **Lys C**, Fixed Modifications **Carbamidomethyl (C),** Variable Modifications **none**, Peptide Tol. **50 ppm**, MS/MS Tol. **0. 5 Da**, Charge state **+1**, MS/MS mode **PSD**, **monoisotopic**, Report Top **10** hits. For details of Mascot query parameters see http://www.matrixscience.com/, follow the link

to "Mascot" - "MS/MS Ion Search" – "Search Form" and select any underlined entry for detailed information. After a short period, the results are shown in the browser window. To transfer the results use the "Get Results" button. Leave the results window (use the "Exit" button), close the "MS/MS Ion Search" dialog and you will receive the annotated spectrum.

Note ***The MASCOT top score protein name is color-coded in green in the treeview. The Bruker MS/MS score is applied to each peptide sequence and all sequences are sorted according to the Bruker score (in brackets left of the peptide sequence).***

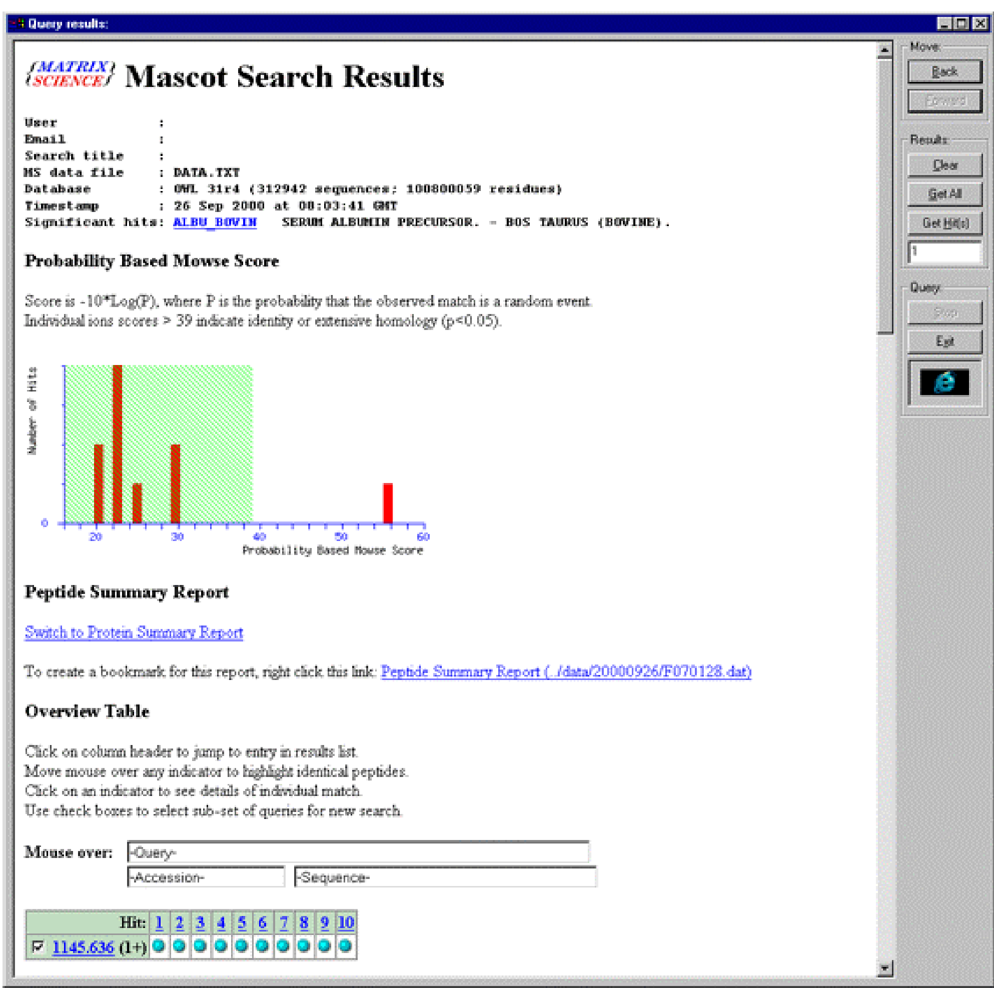


Figure 2-15, MS/MS Ion search results

A Appendix: Amino acid Residues and Fragmentations

A.1 Amino Acid Residues

A.1.1. Single letter code

Most of the analysis programs use the single letter code - learn it by heart from: http://alpha2.bmc.uu.se/~kenth/bioinfo/singleletter.html

Name	Three Letter Code	Single Letter Code	Mnemonic
Alanine	Ala	A	(Alanine)
Cysteine	Cys	C	(Cysteine)
Aspartic Acid	Asp	D	(aciD)
Glutamic Acid	Glu	E	(E comes after D)
Phenylalanine	Phe	F	(Ph=F)
Glycine	Gly	G	(Glycine)
Histidine	His	H	(Histidine)
Isoleucine	Ile	I	(Isoleucine)
Lysine	Lys	K	(L follows K)
Leucine	Leu	L	(Leucine)
Methionine	Met	M	(Methionine)
Asparagine	Asn	N	(AsparagiNe)
Proline	Pro	P	(Proline)
Glutamine	Gln	Q	(Qlutamine)
Arginine	Arg	R	(aRginine)
Serine	Ser	S	(Serine)
Threonine	Thr	T	(Threonine)
Valine	Val	V	(Valine)
Tryptophan	Trp	W	(Double ring - W)
Tyrosine	Tyr	Y	(tYrosine)

and:

http://www-lehre.img.bio.uni-goettingen.de/edv/Bio_Inf/MolBiol/intro1.htm

Single Letter Code	Three Letter Code	Mnemonic
A	Ala	**A**lanine
C	Cys	**C**ysteine
D	Asp	aspar**D**ic acid
E	Glu	glu**E**tamic acid
F	Phe	**F**enylalanine
G	Gly	**G**lycine
H	His	**H**istidine
I	Ile	**I**soleucine
K	Lys	before **L**
L	Leu	**L**eucine
M	Met	**M**ethionine
N	Asn	Asparagi**N**e
P	Pro	**P**roline
Q	Gln	**Q**-tamine
R	Arg	a**R**ginine
S	Ser	**S**erine
T	Thr	**T**hreonine
V	Val	**V**aline
W	Trp	t**W**o rings
Y	Tyr	t**Y**rosine

A.1.2. Genetic Code

1.			2.		3.
	T	C	A	G	
T	TTT Phe [F]	TCT Ser [S]	TAT Tyr [Y]	TGT Cys [C]	T
	TTC Phe [F]	TCC Ser [S]	TAC Tyr [Y]	TGC Cys [C]	C
	TTA Leu [L]	TCA Ser [S]	TAA Ter [end]	TGA Ter [end]	A
	TTG Leu [L]	TCG Ser [S]	TAG Ter [end]	TGG Trp [W]	G
C	CTT Leu [L]	CCT Pro [P]	CAT His [H]	CGT Arg [R]	T
	CTC Leu [L]	CCC Pro [P]	CAC His [H]	CGC Arg [R]	C
	CTA Leu [L]	CCA Pro [P]	CAA Gln [Q]	CGA Arg [R]	A
	CTG Leu [L]	CCG Pro [P]	CAG Gln [Q]	CGG Arg [R]	G
A	ATT Ile [I]	ACT Thr [T]	AAT Asn [N]	AGT Ser [S]	T
	ATC Ile [I]	ACC Thr [T]	AAC Asn [N]	AGC Ser [S]	C
	ATA Ile [I]	ACA Thr [T]	AAA Lys [K]	AGA Arg [R]	A
	ATG Met [M]	ACG Thr [T]	AAG Lys [K]	AGG Arg [R]	G
G	GTT Val [V]	GCT Ala [A]	GAT Asp [D]	GGT Gly [G]	T
	GTC Val [V]	GCC Ala [A]	GAC Asp [D]	GGC Gly [G]	C
	GTA Val [V]	GCA Ala [A]	GAA Glu [E]	GGA Gly [G]	A
	GTG Val [V]	GCG Ala [A]	GAG Glu [E]	GGG Gly [G]	G

A.1.3. Formulas and Molecular Weights

Table A-1, Amino acid residue masses (mono-isotopic and average) together with 3-and 1-letter code and elemental composition

Name	Symbol	S.	C	H	N	O	S	Monoisotopic Mass	Averaged Mass
Alanine	Ala	A	3	5	1	1	0	71,03712	71,079
Cysteine	Cys	C	3	5	1	1	1	103,00919	103,145
Aspartic acid	Asp	D	4	5	1	3	0	115,02695	115,089
Glutamic acid	Glu	E	5	7	1	3	0	129,0426	129,116
Phenylanaline	Phe	F	9	9	1	1	0	147,06842	147,177
Glycine	Gly	G	2	3	1	1	0	57,02146	57,052
Histidine	His	H	6	7	3	1	0	137,05891	137,141
Isoleucine	Ile	I	6	11	1	1	0	113,08407	113,159
Lysine	Lys	K	6	12	2	1	0	128,09497	128,174
Leucine	Leu	L	6	11	1	1	0	113,08407	113,159
Methionine	Met	M	5	9	1	1	1	131,04049	131,199
Asparagine	Asn	N	4	6	2	2	0	114,04293	114,104
Proline	Pro	P	5	7	1	1	0	97,05277	97,117
Glutamine	Gln	Q	5	8	2	2	0	128,05858	128,131
Arginine	Arg	R	6	12	4	1	0	156,10112	156,188
Serine	Ser	S	3	5	1	2	0	87,03203	87,078
Threonine	Thr	T	4	7	1	2	0	101,04768	101,105
Valine	Val	V	5	9	1	1	0	99,06842	99,133
Tryptophan	Trp	W	11	10	2	1	0	186,07932	186,213
Tyrosine	Tyr	Y	9	9	1	2	0	163,06333	163,176

A.1.4. Chemical Structures

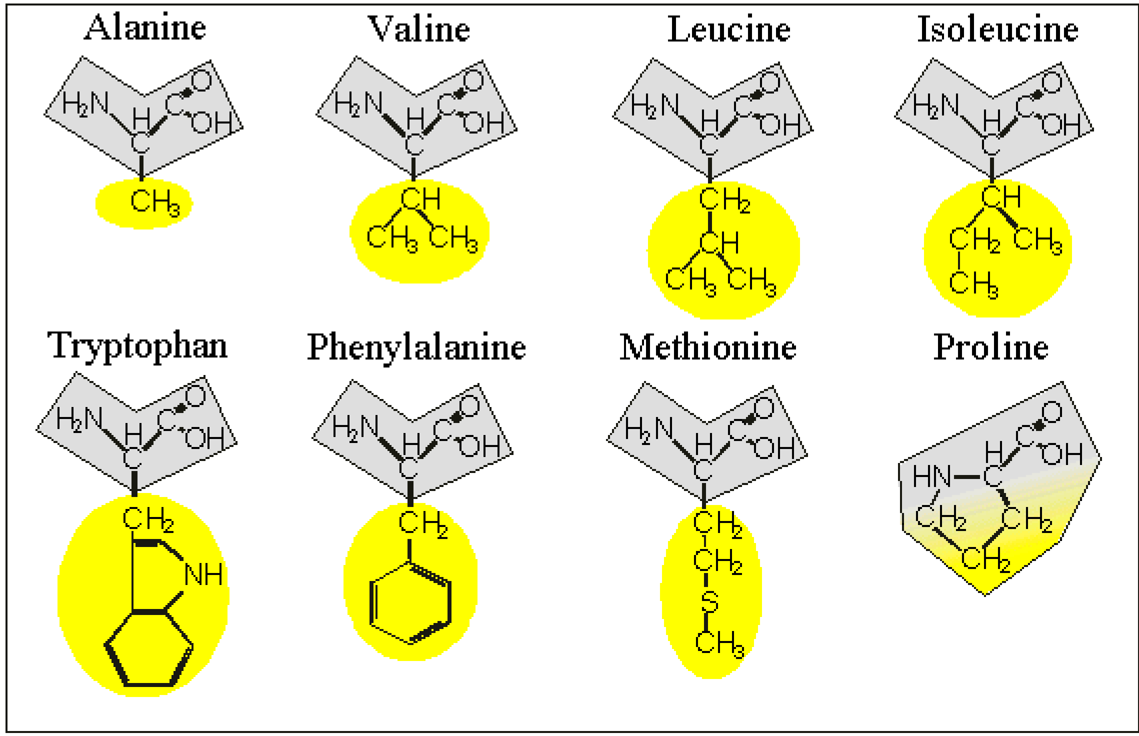


Figure A-1, Neutral hydrophobic Amino acid

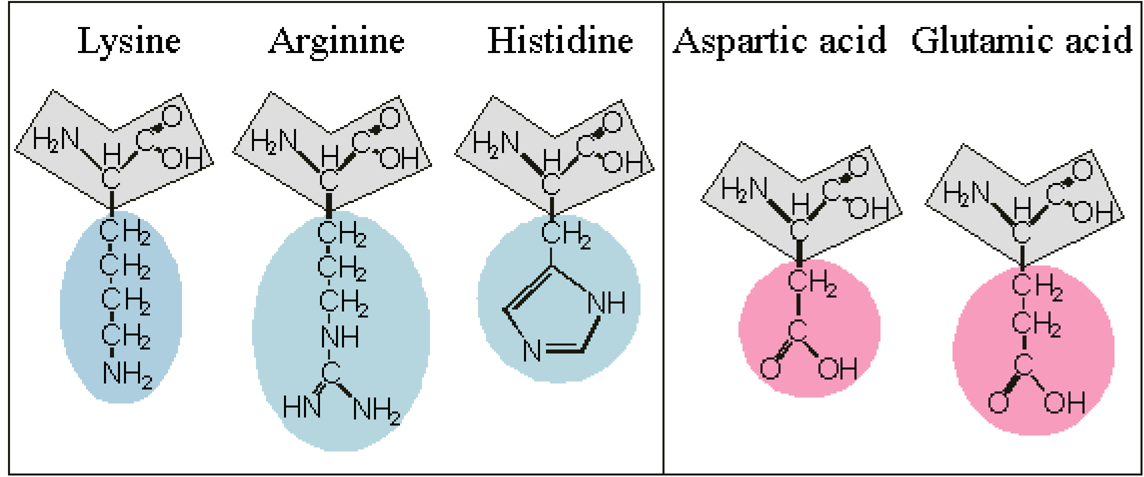


Figure A-2, Basic and acid Amino acids

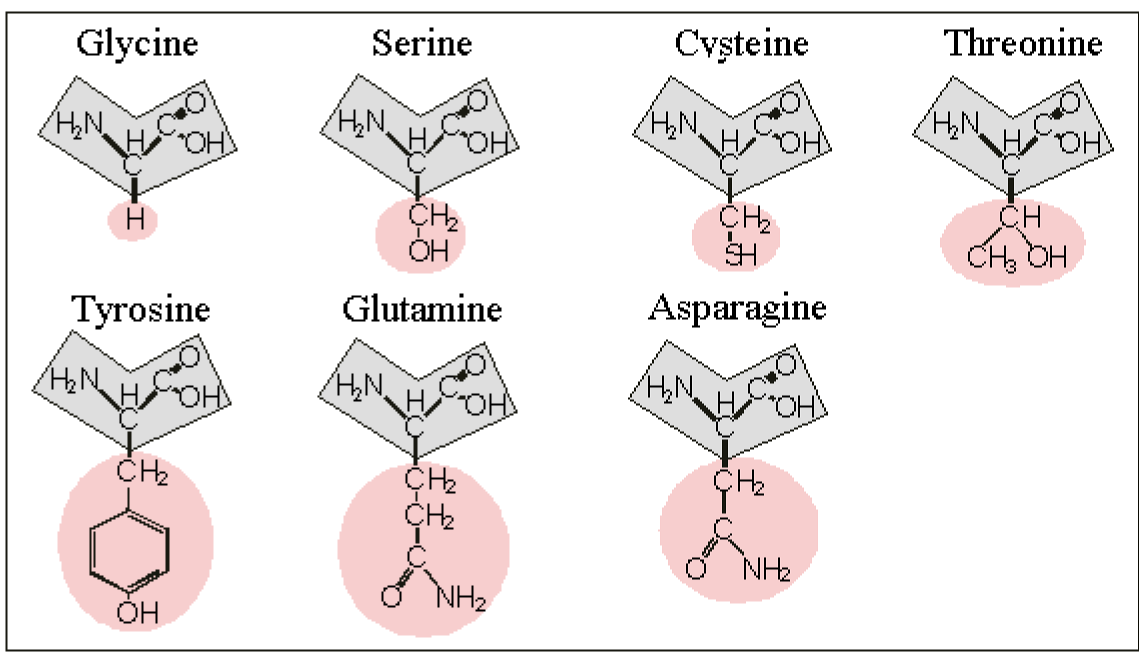


Figure A-3, Neutral polar Amino acids

A.2 Peptide Fragmentation

The types of fragment ions observed in an MS/MS spectrum depend on many factors including primary sequence, the amount of internal energy, how the energy was introduced, charge state, etc. The nomenclature used for fragment ions is the Biemann nomenclature (R. S. Johnson, S. A. Martin & K. Biemann (1988), Int. J. Mass Spec. Ion Procs. 86, 137-154).

Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either *a*, *b* or *c*. If the charge is retained on the C terminal, the ion type is either *x*, *y* or *z*. An index indicates the number of residues in the fragment. In addition to the proton(s) carrying the charge, c ions and y ions abstract an additional proton from the precursor peptide. Note that these structures include a single charge carrying proton. In electrospray ionization, tryptic peptides generally carry two or more charges, so those fragment ions may carry more than one proton.

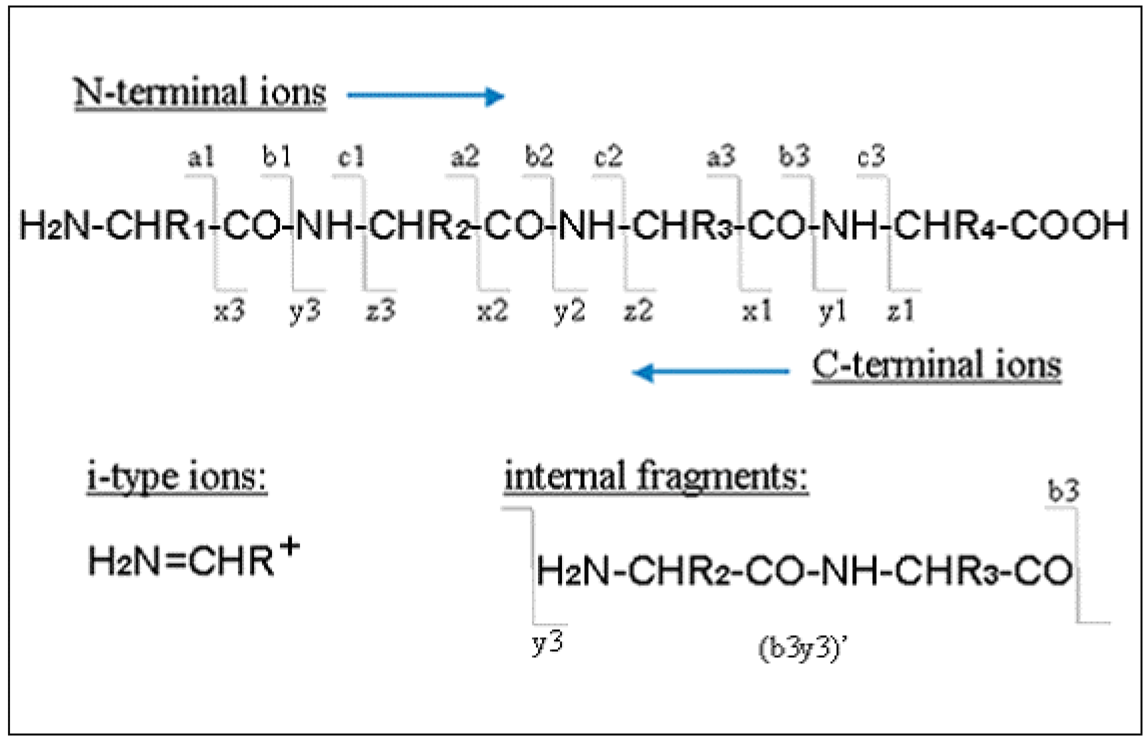


Figure A-4, Peptide Fragmentation

Typical fragment ions observed are:

- Low energy CID: b and y
- PSD: a, b, y and i, including neutral losses of NH_3 from a and b
- ISD: c and y
- ECD-FTICR: c and z

Fragmentation of the backbone at two sites gives rise to internal fragments. Usually, these are formed by a combination of *b*-type and *y*-type cleavage to produce the illustrated structure, amino-acylium ion. Sometimes, internal ions can be formed by a combination of *a*-type and *y*-type digest, an amino-immonium ion.

An internal fragment with just a single side chain formed by a combination of *a* type and *y* type digest is called an immonium ion. The immonium ions can be used for *DeNovo* sequencing. The values from the following table are used to find these ions.

Table 4-2, Immonium and related ion masses

Residue	3-letter	1-letter	Immonium	Related ions
Alanine	Ala	A	44	
Cysteine	Cys	C	76	
Aspartic acid	Asp	D	88	
Glutamic acid	Glu	E	102	
Phenylalanine	Phe	F	120	148
Glycine	Gly	G	30	
Histidine	His	H	110	82 155
Isoleucine	Ile	I	86	44 72
Lysine	Lys	K	101	84
Leucine	Leu	L	86	44 72
Methionine	Met	M	104	60
Asparagine	Asn	N	87	
Proline	Pro	P	70	98
Glutamine	Gln	Q	76	
Arginine	Arg	R	112	100 87 70 60
Serine	Ser	S	60	
Threonine	Thr	T	74	
Valine	Val	V	72	
Tryptophan	Trp	W	159	
Tyrosine	Tyr	Y	136	

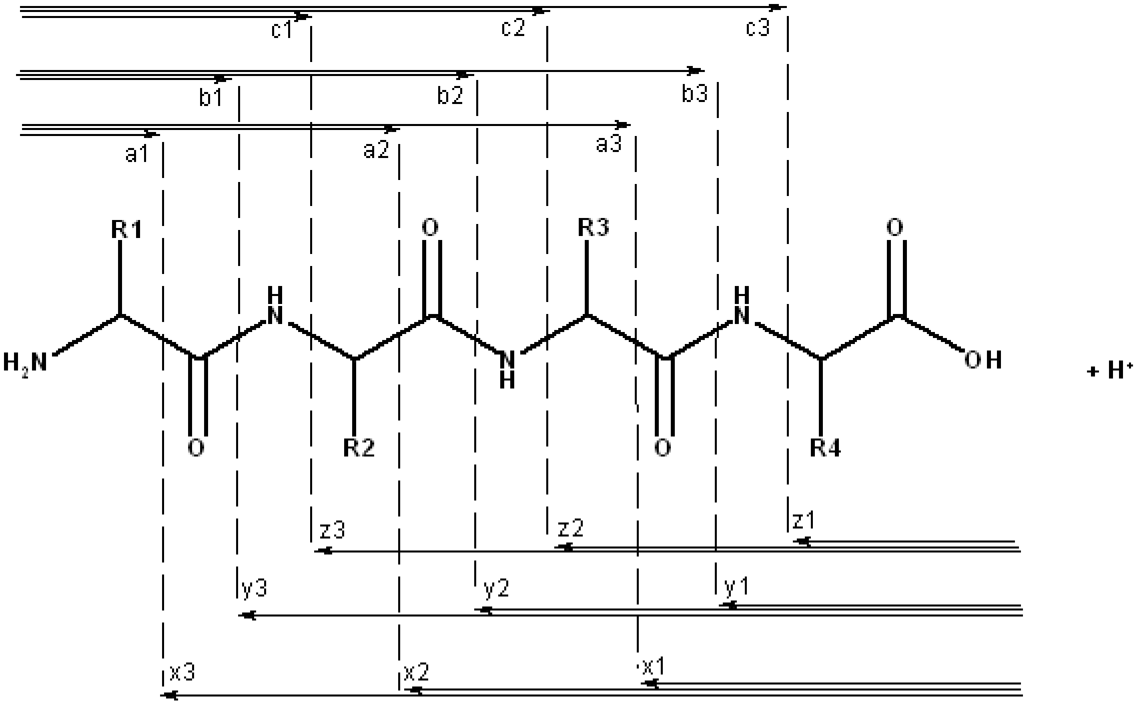


Figure A-5, Fragmentation pattern for peptides

Figure A-6, Structures of the fragments

A.3 Menu and Shortcut list

Toolbar button	*Menu option*	Shortcut F-key	Description
	***File** – Find*		Opens a file finder
	File – Open Spectrum	Ctrl + O	Opens a file manager
	File – Multiple open 1r		**All** existing data files in a folder are opened.
	File – Close		Closes the active window
	File – Combine multiple LIFT spectra		Multiple spectra can be combined for use in one data file.
	File – Save		Saves the active data file
	File – Print		Opens the print dialog
	File – Print Preview		Starts the print preview
	File – Print Setup	Ctrl + P	Opens the printer setup dialog
			Prints the active data file immediately in accordance to the Print Setup (page one landscape, other portrait)
	File – Send		Opens an email program to send the active spectrum to an email address as an attached file.
	File – Last used data files		The last used data files are listed here
	File – Exit	ALT + F4	Terminates the program
	***Edit** – Undo*	Ctrl + Z	The previous action is undone
	Edit – Cut	Ctrl + X	Deletes and copies into clipboard
	Edit – Copy	Ctrl + C	Copies from clipboard to cursor position
	Edit – Paste	Ctrl + V	Pastes from clipboard to cursor position
	Edit – Sequence…		Starts the SequenceEditor
	***View** – Toolbar*		Show or hide the toolbar
	View – Status Bar		Show or hide the status bar
	View – Query Results		Reloads results of a former query
	View – View Fingerprint Results		Changes fragment window to fingerprint view
	View – View MS/MS Results		Changes fragment window to sequence view
	View – Matched Peaks		Displays masses and sequence positions for peaks matching a protein sequence (**black** in general, **blue for peptides containing optional modifications**)

Toolbar button	*Menu option*	Shortcut F-key	Description
	View – Unmatched Peaks		Displays masses in **red** for peaks not matching a protein sequence
	View – Matched and Unmatched Peaks		Displays all peaks
	View – Picked Peaks		Shows the peaks in the spectrum window
	View – Coordinates		Show or hide mass and intensity information within the status bar
	View – Grid		Show or hide a grid within the spectrum window
	View – Scaling		Opens the expand manual dialog or resets to whole spectrum
	View – Zooming		Enables or disables zoom mode
	View – Undo zooming	-	The previous zoom action is undone
	View – Redo zooming	-	The previous zoom action is redone
	View – Distance		Enables or disables the line distance mode
	View – Data Cursor		Enables or disables display of mass and intensity of current peak
	View –Colors		Opens a dialog to change colors
	View – Display Mode		Offers different display modes
	Analysis *– Check Sequence*		Changes the Fragments window to Check Sequence
	Analysis – DeNovo Sequencing		Changes the Fragments window to DeNovo Sequencing
	Analysis – Full DeNovo Sequencing		Opens the DeNovo Sequencing procedure performs a stepwise build-up of the amino acid sequence.
	Analysis – ISD Data		Opens a dialog to select the ions to be calculated
	Analysis – Set Threshold		Defines a threshold (only peaks above the threshold will be taken into account)
-y- -b-	*Analysis – Annotation*		Defines which calculated ions shall be displayed and performs or resets an annotation.
	Analysis – Picked Peaks		Shows the peaks in the spectrum window
	Analysis – Add Peaks		Switches to peak picking mode (max. data point assignment only!).
	Analysis – Remove Peaks		Switches to peak deletion mode.
	Search *– EMBL*		Opens the internet search using EMBL protein identification by peptide masses

Toolbar button	*Menu option*	Shortcut F-key	Description
	Search – Prowl Peptide Mapping		Opens the internet search using Prowl protein identification by peptide mapping
	Search – Mascot Peptide Mass Fingerprint		Opens the Mascot Peptide Mass Fingerprint Query dialog
	Search – Mascot Sequence Query		Opens the Mascot Sequence Query dialog
	Search – Mascot MS/MS Ion Search		Opens the Mascot MS/MS Search dialog
	Search – Mascot Batch	-	Opens the Mascot Batch Mode window
	Search – Search Mass Search SequenceEditor	-	Opens the Search for mass dialog box in SequenceEditor (*Search – Mass Search)*
	Search –Digest SequenceEditor	-	Opens the Perform Digest dialog box in the SequenceEditor (*Search – Perform Digest*)
	Tools *– Authentication*		Opens a dialog to define web access
	Tools – Options		Opens a dialog to define general, display and print options
	Tools – Maldi Spectrum Parameter		Displays basic information about loaded spectrum
	Tools – Modification info		This window shows information about the currently defined modification within the amino acid sequence.
	Tools – Formula parser		This formula parser calculates the mass weight of a chemical formula.
	Tools – Start XTOF NT		Starts the XTOF NT program
	Tools – Start XMASS NT		Starts the XMASS NT program
	Tools – Execute XMASS commands		Executes commands on XMASS running on a remote computer
	Tools – Customize		Customize the toolbar(s) and the menu bar
	Window *– Show/Hide Treeview*		Show or hide the Treeview window
	Window – Show/Hide Fragments		Show or hide the Fragments window
	Window – Show/Hide Browser Window	-	Starts the Internet Browser to get Query results
	Window – Start SequenceEditor		Starts the SequenceEditor
	Window – Reset Window Sizes		The arrangement of the window in BioTools will be set to the default settings

Toolbar button	*Menu option*	Shortcut F-key	Description
	Window – Cascade		Shows all windows cascaded
	Window – Tile Horizontally		Shows all windows tiled horizontally
	Window – Tile Vertically		Shows all windows tiled vertically
	Window – Arrange icons		Arranges all icons on the bottom
	Window – Active data files		All active data files are listed here
	Help *– License Manager*		Opens the License Manager dialog to verify or set the license keys
	Help – What's new in BioTools 2.1		Opens a presentation with new features in BioTools 2.1
	Help – What's new in BioTools 2.2		Opens a presentation with new features in BioTools 2.2
	Help – Open BioTools Manual		Starts the Acrobat Reader (if installed) and opens the BioTools manual
	Help – Help Topics		Opens the online help
	Help – About BioTools		Opens About BioTools window

A.4 Toolbar Reference list for BioTools

To hide or display the Toolbar, choose menu *View - Toolbar*.

Toolbar button	*Menu option*	Shortcut / Function key	Description
	File – Open Spectrum	Ctrl + O	Opens a file manager
	File – Save	Ctrl + S	Saves the state of the active data file
	Edit – Cut	Ctrl + X	Deletes and copies into clipboard
	Edit – Copy	Ctrl + C	Copies from clipboard to cursor position
	Edit – Paste	Ctrl + V	Pastes from clipboard to cursor position
	-	-	Prints the active data file immediately in accordance to the Print Setup

Toolbar button	*Menu option*	Shortcut / Function key	Description
	Analysis – Check Sequence	-	Changes the fragment window to check sequence mode
	Analysis – DeNovo Sequencing	-	Changes the fragment window to *DeNovo* sequencing mode
	Analysis – Annotation Parameter	-	Opens the annotation options dialog box
	Search – Search Mass Search SequenceEditor	-	Opens the Search for mass dialog box in SequenceEditor (*Search – Mass Search)*
	Search –Digest SequenceEditor	-	Opens the Perform Digest dialog box in the SequenceEditor (*Search – Perform Digest*)
	Search – Mascot Peptide Mass Fingerprint	-	Opens the internet search via Peptide Mass Fingerprint
	Search – Mascot Sequence Query	-	Opens the internet search via Sequence Query
	Search – Mascot MS/MS Ion Search	-	Opens the internet search via MS/MS Ion Search
	Window – Show/Hide Treeview	-	Show or hide the treeview window
	Window – Show/Hide Fragments	-	Show or hide the fragment window
	View – Undo zooming	-	The previous zoom action is undone
	View – Redo zooming	-	The previous zoom action is redone
	Search – Mascot Batch	-	Opens the Mascot Batch Mode window
	Window – Show/Hide Browser Window	-	Starts the Internet Browser to get Query results
	Edit – Sequence…	-	Loads a sequence into the Sequence Editor and starts this program
	Tools – Start XTOF NT	-	Starts the XTOF NT program
	Tools – Start XMASS NT	-	Starts the XMASS NT program
	Help – About BioTools	-	Opens About BioTools window

A.5 Toolbar Reference list for SequenceEditor

To hide or display the Toolbar, choose menu *View - Toolbar*.

Toolbar button	*Menu option*	Shortcut/ Function key	Description
	File – New Sequence	Ctrl + N	Creates a new sequence, also from the web (Sequences from the Web).
	File – Save	Ctrl + S	Saves the active sequence with its current name. If you have not named the sequence, your SequenceEditor displays the Save As dialog box.
	Edit – Cut	Ctrl + X	Removes selected data from the sequence and stores it on the clipboard.
	Edit – Copy	Ctrl + C	Copy the selection to the clipboard.
	Edit – Paste	Ctrl + V	Inserts the contents of the clipboard at the insertion point.
Bio	-		The modified sequence or a marked range of it will be Send data to BioTools for further processing.
	-		Prints the active data file immediately
	Help – About SequenceEditor…		Opens About SequenceEditor window
	-		Activates the context sensitive help.
	File – Print	Ctrl + P	Opens the printer dialogue
	File – Exit	ALT + F4	Terminates the program

A.6 Part Numbers

# 216941	Software-Package BioTools 2.2
# 216935	License BioTools 2.2
# 217032	License RapiDeNovo Sequencing
# 216936	BioTools User Manual

Quick Guide to Operate BioTools and SequenceEditor

0 INTRODUCTION Q-2

0.1 What can I BioTools use for? Q-2
0.2 How to use this Manual and the Help Functionality within BioTools Q-3
0.3 Send Mascot search result pages via email to somebody who does not have BioTools Q-3

1 WINDOW OVERVIEW BIOTOOLS Q-4

1.1 Menu Bar Q-4
1.2 Toolbar Q-5
1.3 Status Bar Q-6
1.4 Working Area Q-6
1.4.1 Treeview Window Q-7
1.4.2 Overview Window Q-8
1.4.3 Spectrum window Q-8
1.4.4 Fragments window Q-9
1.5 Context Menus Q-10
1.5.1 Treeview window Q-10
1.5.2 Spectrum window Q-11
1.5.2.1 Display and cursor Q-12
1.5.2.2 Show Internal Fragments Q-14
1.5.2.3 X-scale and Y-scale Q-15
1.5.3 Fragments window Q-16

0 Introduction

0.1 What can I BioTools use for?

BioTools is for all those using mass spectrometric data to characterize protein structures or to do proteomics. BioTools receives processed spectra from all Bruker mass spectrometers and reference sequence information to allow the visualization of their match. Such sequences can be obtained, e.g., by classical fingerprint- or MS/MS-based library searching, or by generating libraries of structures from a target sequence, which are screened for best match to the data. The match between fingerprint or MS/MS spectrum and target sequence is visualized in the annotated spectrum plus a peptide coverage map.

The list below shows some of the basic tasks enabled by the BioTools software. It also contains links to the corresponding sections of the manual, where they are described in detail:

- Library searches using MASCOT results for importing into MS, MS/MS or LC-MS/MS data sets
- Define Proteomics batch searches with color-coded result visualization
- Edit results from automated batch processes interactively
- Email reporting of proteomic runs
- Full structure of single chain proteins can be analyzed including modifications, mutations and crosslinks.
- Screening of a target sequence for modifications and mutations
- Use calculated digest masses to verify modification and processing profiles by LC-MS/MS
- Use calculated digest masses to investigate spectra for low S/N peaks
- Localize post-translational modifications (PTMs) using MS/MS data
- MS/MS support for LIFT, PSD, ISD spectra, esquire*3000*^{plus}, APEX and BioTOF/BioTOF-Q; basic *DeNovo* sequencing supported
- "Click on a sequence" spectrum annotation
- Copy and paste publishing of sequence coverage maps and annotated spectra

0.2 How to use this Manual and the Help Functionality within BioTools

The Operator Manuals are meant to be online help documents, which can be accessed using the F1 key at any point of entry of the program. In addition they can be read like a book, which is not recommended whatsoever.

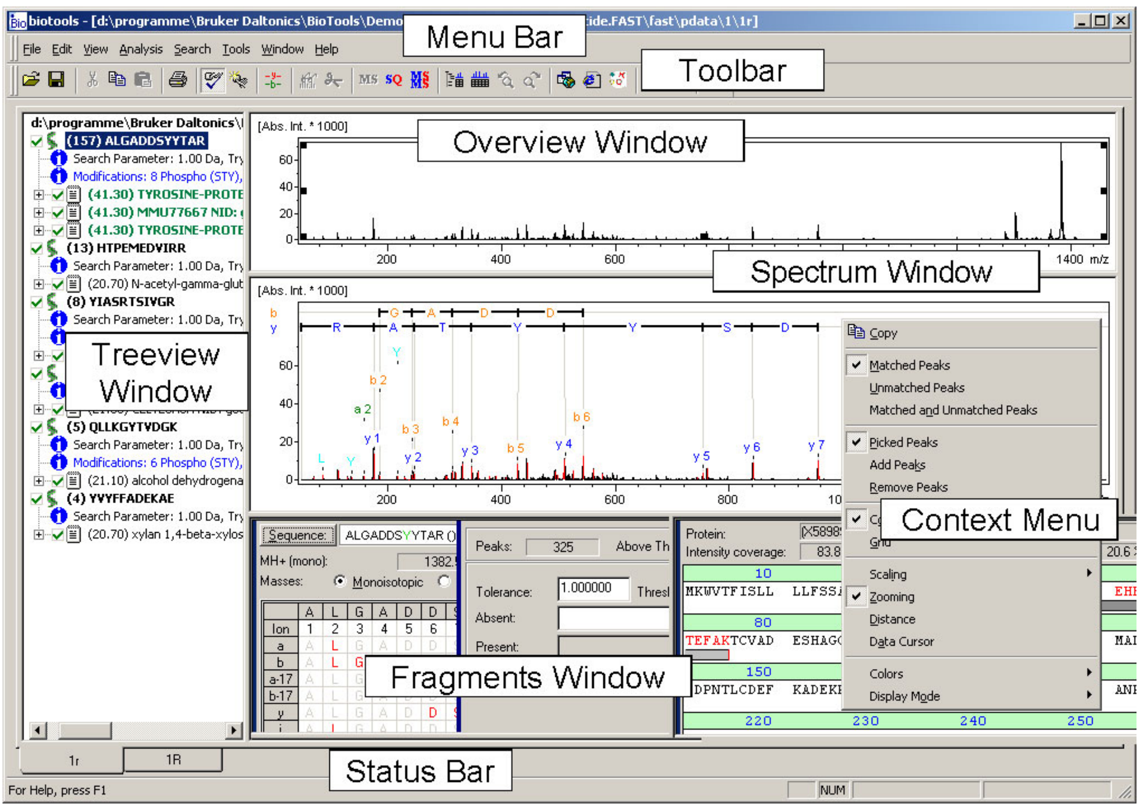
The two tutorial blocks with focus on either the SequenceEditor or BioTools contain procedure descriptions for specific analytical problems. Work your way through those applications, which you want to learn about and use the F1 access to online help only where needed.

0.3 Send Mascot search result pages via email to somebody who does not have BioTools

- Perform the search under BioTools (in the data tree under 1SRef BioTools generates the file search_result.html)
- On the same computer open the file search_result.html in Netscape or the Internet Explorer (the first page opens up and connects to the Mascot server to display the full result page)
- Save it under a different name (now you have the full information in an HTML file)
- This file can be sent to and opened by anybody.

1 Window Overview BioTools

This chapter will give an explanation of all BioTools window areas and how to handle data and sequences.



The active window area is marked on the left with a blue beam.

Further information is given in the BioTools Reference Manual.

1.1 Menu Bar

The menu bar can be moved with the mouse. Clicking on the background of the menu bar with the left mouse button and move the menu bar with held mouse button to the desired position.

1.2 Toolbar

By activating this option the upper toolbar is shown (standard) or hidden. The toolbar can also be moved with the mouse. Clicking on the background of the toolbar with the left mouse button and move the toolbar with held mouse button to the desired position.

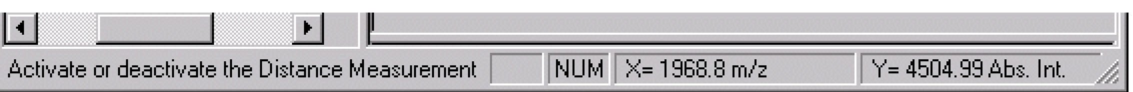
Toolbar

Toolbar button	*Menu option*	Shortcut / Function key	Description
	File – Open Spectrum	Ctrl + O	Opens a file manager
	File – Save	Ctrl + S	Saves the state of the active data file
	Edit – Cut	Ctrl + X	Deletes and copies into clipboard
	Edit – Copy	Ctrl + C	Copies from clipboard to cursor position
	Edit – Paste	Ctrl + V	Pastes from clipboard to cursor position
	-	-	Prints the active data file immediately in accordance to the Print Setup
	Analysis – Check Sequence	-	Changes the fragment window to check sequence mode
	Analysis – DeNovo Sequencing	-	Changes the fragment window to *DeNovo* sequencing mode
	Analysis – Annotation Parameter	-	Opens the annotation options dialog box
	Search – Search Mass Search SequenceEditor	-	Opens the Search for mass dialog box in SequenceEditor (*Search – Mass Search)*
	Search –Digest SequenceEditor	-	Opens the Perform Digest dialog box in the SequenceEditor (*Search – Perform Digest*)
	Search – Mascot Peptide Mass Fingerprint	-	Opens the internet search via Peptide Mass Fingerprint
	Search – Mascot Sequence Query	-	Opens the internet search via Sequence Query
	Search – Mascot MS/MS Ion Search	-	Opens the internet search via MS/MS Ion Search

Toolbar button	*Menu option*	Shortcut / Function key	Description
	Window – Show/Hide Treeview	-	Show or hide the treeview window
	Window – Show/Hide Fragments	-	Show or hide the fragment window
	View – Undo zooming	-	The previous zoom action is undone
	View – Redo zooming	-	The previous zoom action is redone
	Search – Mascot Batch	-	Opens the Mascot Batch Mode window
	Window – Show/Hide Browser Window	-	Starts the Internet Browser to get Query results
	Edit – Sequence…	-	Loads a sequence into the Sequence Editor and starts this program
	Tools – Start XTOF NT	-	Starts the XTOF NT program
	Tools – Start XMASS NT	-	Starts the XMASS NT program
	Help – About BioTools	-	Opens About BioTools window

1.3 Status Bar

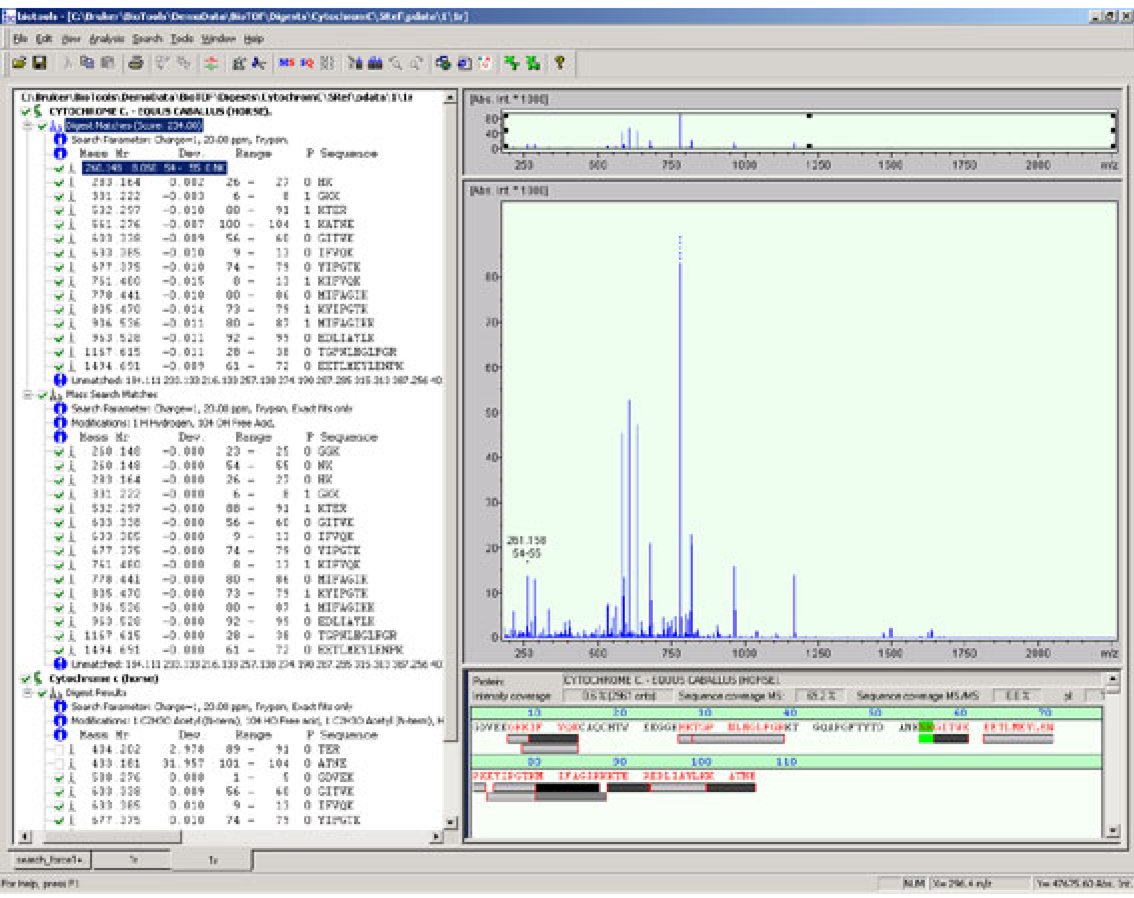
By activating this option in the view menu the lower status bar is shown (standard) or hidden. In the left corner is given a short help text corresponding to the cursor position and actions. The next boxes show the activated caps lock and the numeric function of the numeric block. In the right corner the x-position and y-position (or x-distance) of the cursor (if not deactivated) are given.



Status bar

1.4 Working Area

The working area within BioTools is separated in four windows:



1.4.1 Treeview Window

Handling with sequences

- Send sequences to SequenceEditor by double-click on the sequence
- Jump between the same levels of the tree with tab and shift tab
- Step up and down with arrow keys ⬆ and ⬇
- Show(Hide) tree below with arrow key ➡ (⬅)
- The marked sequence in the treeview is shown in green in the fragments window and in the spectrum window the mass is also shown
- To hide a peak uncheck it in the treeview, it will also disappear in the fragments window and the peak is marked "missing" in the spectrum window.

1.4.2 Overview Window

The overview window always shows the whole spectrum. The part of the spectrum shown in the spectrum window is bordered by a zoom rectangle. This zoom rectangle can be arranged with the mouse by left-clicking on the borders. It can be moved by left-clicking in the rectangle and drawing.

The arrangement of the rectangle can also be set with the arrow keys and/or Ctrl and/or shift key.

Arrow	-	Shift	Ctrl	Ctrl + Shift
	moving with a little overlap	moving in short steps	scaling the rectangle roughly	scaling the rectangle in short steps
right ➡	right	right	increase width, center fix	increase width, center fix
left ⬅	left	left	decrease width, center fix	decrease width, center fix
up ⬆	up	up	increase height, baseline fix	increase height, baseline fix
down ⬇	down	down	decrease height, baseline fix	decrease height, baseline fix

1.4.3 Spectrum window

Shown here is the spectrum – depending on its functionality – with its annotation.

- additional peaks can be added or deleted by activating the cursor via the Context menu and drawing the respective area. The peak will be added to/deleted from the spectrum.
- the zoom area can be drawn with the held left mouse button in the window or by drawing the scale bar with the held right or pushing with the held left mouse button, a double-click in the spectrum window brings back the whole spectrum, with the undo and redo zooming buttons can be changed between the previous zoom views
- the cursor can be set to data or distance mode
- change to the default zoom cursor by clicking in the spectrum window with the right mouse button
- several colors and graphical settings can be set

1.4.4 Fragments window

In the fragments window either the sequence info, or the fingerprint info, or the DeNovo Sequencing info is shown.

Sequence info

With the option *Check Sequence* a sequence can be chosen, their amino acids are shown in the fragments window and the data will be annotated in the spectrum window corresponding to this sequence, see Check Sequence.

DeNovo Sequencing

Under the option *DeNovo* sequencing, tools are available, which allow to extract sequence information from MS/MS data. Depending on the data quality and type, these can be either rather complete sequence suggestions or only "seed" sequences or sequence tags, which require further interactive work, see DeNovo Sequencing.

Fingerprint view

In the fingerprint mode the found sequences are shown, by left-clicking on a sequence, its background change to green

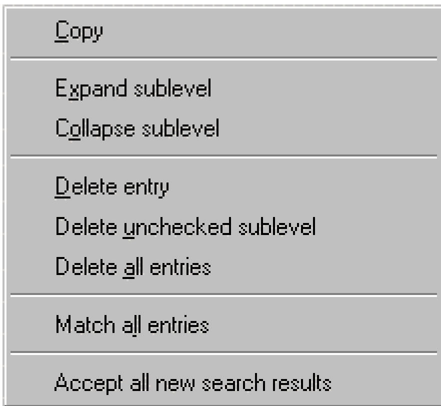
- the actual sequence is marked green and is also shown in the treeview window and the peak is shown in the spectrum window
- N-Glycanes are highlighted with yellow background
  (first protein: N, second p.: ACDEFGHIKLMNPQRSTVWY, third p.: S or T)
- The colors and fonts can be set via the Context menu Fragments window

1.5 Context Menus

There are several context menus available, one from the Treeview window and several from the spectrum window. Clicking in the respective area with the right mouse button opens the context menu.

1.5.1 Treeview window

With this context menu the appearance of the Treeview window and its sequences can be changed, sublevels can be deleted. Choose first in the treeview window the desired sequence by clicking on it with the left mouse button. Click then with the right mouse button on the active sequence. The context menu will be opened.



Context menu to work with sequences

Options	**Description**
Copy	Copy the whole sequence data file in the clipboard.
Expand sublevel	All Sublevel structures of the treeview **of all nodes of the selected level are expanded**. For single-entry sublevel expansion click on the "+" symbol.
Collapse sublevel	All Sublevel structures of the treeview **of all nodes of the selected level are collapsed**. For single-entry sublevel collapse click on the "-" symbol.
Delete entry	Deletes the selected entry and all sublevel entries. If the very top line in the treeview is selected **all treeview entries are completely deleted**. (The same function is available from the *Clear* button in the Mascot result browser window.
Delete unchecked sublevels	Deletes all sublevel entries, which do not contain the red checkmark. E.g., uncheck all peptides from a list of proteolytic peptides, which you think were mis-assigned and use this operation from any higher level.
Delete all entries	Deletes all entries.

Options	**Description**
Match all entries	Determine the MS/MS score for all sequences in the treeview and sort the list of sequences by decreasing score values
Accept all new search results	Determine the MS/MS score for all sequences in the treeview and sort the list of sequences by decreasing score values

In general, the treeview information is used for display purposes in the spectrum just by a left mouse button click. The treeview entry is opened in the SequenceEditor in the case of MS/MS data by a double click on the sequence.

1.5.2 Spectrum window

Easy zooming of the spectrum display is available if the context is on the overview window:

1. The spectrum window can be changed with the arrow buttons:

Description	**Shortcut / Function key**
*y-scale * 2 or /2*	**CTRL Arrow** Down or Up
*x-scale * 2 or /2*	**CTRL Arrow** Left or Right
Move zoom box	All **Arrow** keys

If the **SHIFT** key is used in conjunction with above short keys, the step size of movement or expansion is only 10 %.

2. The spectrum window can also be changed with the mouse buttons:

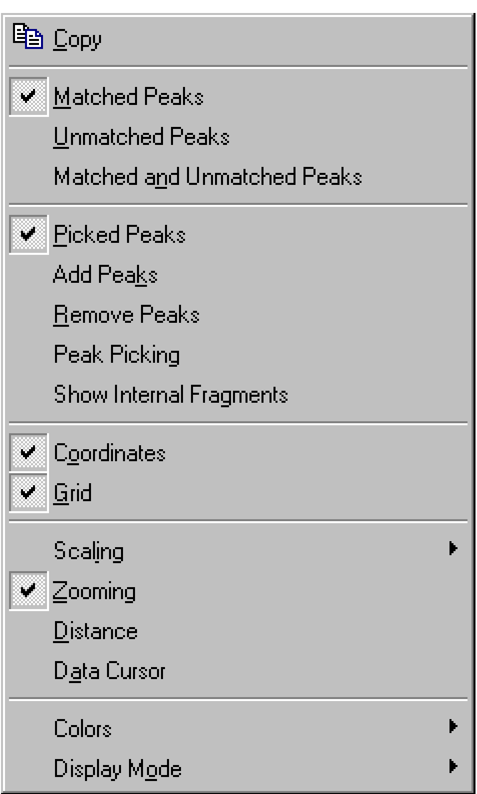
Click in the overview window with the left mouse button and scale or move the area of interest or draw a new area.

Further display control of the Spectrum window is described in the following sections, also it is possible to jump with Undo and Redo zooming between the last actions.

1.5.2.1 Display and cursor

Clicking in the spectrum window with the right mouse button opens the context menu.

The appearing context menu is similar to the *View menu*.

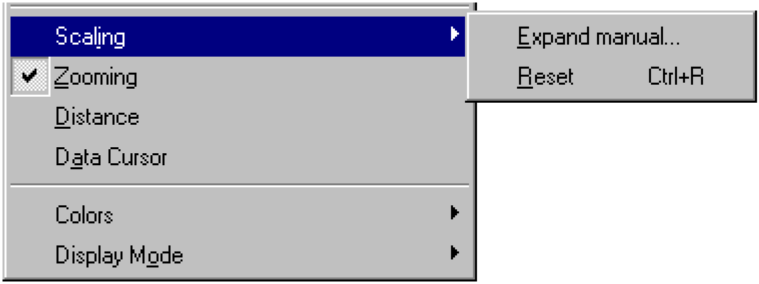


Context menu of spectrum window

Options	**Description**
Copy	Copies the actual contents of the spectrum window into the clipboard
Matched Peaks	Displays masses and sequence positions for peaks matching a protein sequence (**black** in general, **blue for peptides containing optional modifications**)
Unmatched Peaks	Displays masses in **red** for peaks not matching a protein sequence
Matched and Unmatched Peaks	Displays all peaks
Picked Peaks	Shows the peaks in the spectrum window
Add Peaks	Switches to peak picking mode (max. data point assignment only!).

Options	**Description**
	Click right mouse button to revert too zoom mode. **Peaklist not stored for general purposes!**
Remove Peaks	Switches to peak deletion mode. Click right mouse button to revert too zoom mode. **Peaklist not stored for general purposes!**
Peak Picking	For TOF data BioTools can perform peak picking with the SNAP or Sum Peak Finder algorithms
Show Internal Fragments	The calculated internal fragments matched to yet unassigned peaks in the spectrum can be displayed on a dialog.
Coordinates	Displays the cursor coordinates in the status bar
Grid	Displays a grid
Scaling	Opens the Scaling menu (see down below)
Zooming	Reverts to zoom mode (default)
Distance	Switches to distance mode to determine the mass difference between two selected vertical lines. Click right mouse button to revert to zoom mode
Data Cursor	Activates the data cursor. Click right mouse button to revert to zoom mode
Colors	Opens the Color setting for Background, Spectrum, Peak Picking and the Annotation of the matched and unmatched Peaks
Display Mode	Opens the Display Mode menu (see down below)

Scaling options recommended for function key addicts:



Options	**Shortcut / Function key**	**Description**
Expand manual	-	Opens the Manual Scaling window to enter x-range and y-range manually
Reset	Ctrl + R	Moves to total overview

Alternative easier accessible zoom box controls are available if the context is on the overview window:

1. Arrow keys

Description	Shortcut / Function key	Description
*y-scale * 2 or /2*	**CTRL Arrow** Down or Up	
*x-scale * 2 or /2*	**CTRL Arrow** Left or Right	
Move zoom box	All **Arrow** keys	

If the **SHIFT** key is used in conjunction with above short keys, the step size of movement or expansion is only 10 %.

2. The spectrum window can also be changed with the mouse buttons:

Click in the overview window with the left mouse button and scale or move the area of interest or draw a new area.

If the context is on the spectrum window:

Move the cursor on the x-axis (y-axis) of the spectrum window; the cursor changes.

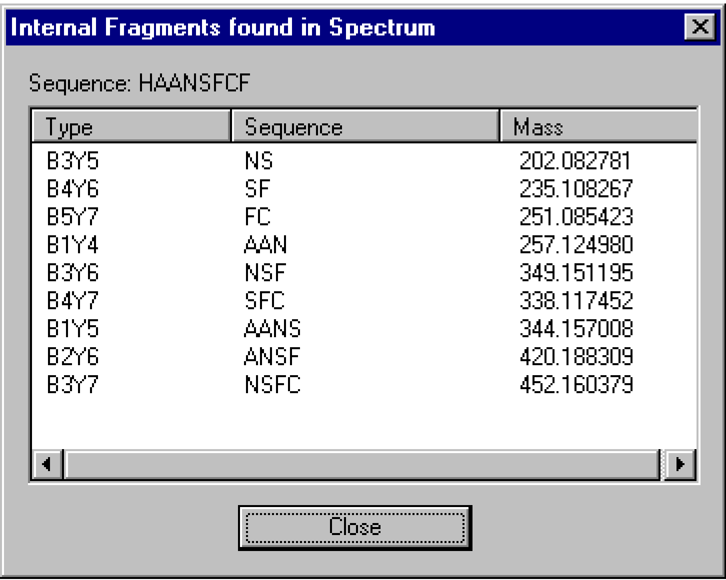
Drag the spectrum along an axis scale: Click with the left mouse button on the axis and move the sector in the desired direction.

Expand the spectrum along an axis scale: Click with the right mouse button on the axis and move the cursor away from the zero-point to stretch the axis scaling, move the cursor towards the zero-point to reduce the axis scaling.

X-axis scale cursor **Y-axis scale cursor**

1.5.2.2 Show Internal Fragments

The calculated internal fragments matched to yet unassigned peaks in the spectrum can be displayed on a dialog accessible via the spectrum context menu with '**Show internal fragments**'. This is only available for PSD and ICR data sets. The name, sequence and mass for the respective a-x-b-y-ions are displayed in a list control that can be sorted by clicking on the respective header.



Show Internal Fragments dialog

1.5.2.3 X-scale and Y-scale

The context menu is opened by clicking in the spectrum window on the y-axis with the right mouse button.

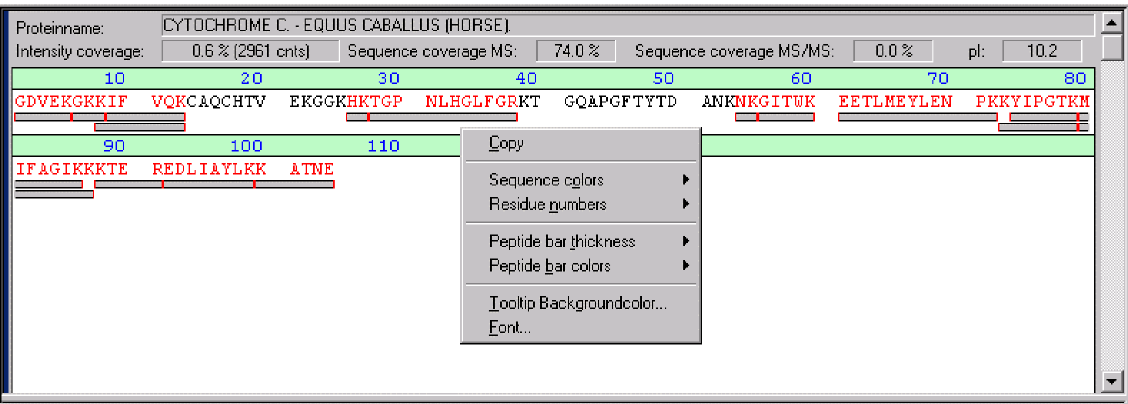
Context menu of y-scale

Options	**Display**
Hide(Show) Y-Axis	Y-axis can be shown (standard) or hidden, disabled for X-axis
Axis Font	The font of the x and y scales and the spectrum annotation objects can be set altogether
Background Color	The background color around the spectrum window and overview window can be set

1.5.3 Fragments window

In the fingerprint mode the found sequences are shown, by left-clicking on a sequence, it will be green back lighted

- the actual sequence is marked green and is also shown in the treeview window and the peak is shown in the spectrum window
- N-Glycanes are highlighted with yellow background (first protein: N, second p.: ACDEFGHIKLMNPQRSTVWY, third p.: S or T)
- Clicking in the fragments window of a protein with the right mouse button opens the context menu.
- If there is a fault in the sequence, a message will be pop up after opening. In addition a symbol is placed in the title of the fragments window.



Context menu of fragments window by fingerprint

Options	**Description**
Copy	Copy the protein data as text in the clipboard and the sequence as graphic. The graphic can be inserted in other programs with "insert special".
Sequence colors	The colors of several residues can be set.
Residue numbers	The numbers can be hidden and the color can be set.
Peptide bar thickness	The thickness of found sequences can be set.
Peptide bar colors	It can be chosen, which value of **Peak intensity**, **MS/MS score** or **Peak goodness** should be shown in the peptide bar.
Tooltip Backgroundcolor	The tooltip background color can be set.
Font	The font and style of the sequence can be set.

T Tutorials for BioTools

L Sequence Database Searches from MALDI Peptide Mass Fingerprints

M Mascot Batch Mode

I In Source Decay (ISD) MALDI-TOF Spectra Analysis

E Using BioTools for esquire3000 data

D DeNovo Sequencing with BioTools

L Sequence Database Searches from MALDI Peptide Mass Fingerprints

L.1 Introduction L-1
L.2 Sequence Database Search using a MALDI Mass Fingerprint L-1
L.3 Define the Search L-2
L.3.1. MASCOT Search Results L-5
L.3.2. Introduction to the MASCOT Query results interface L-6
L.3.3. Work with Search Results in BioTools L-11

L.1 Introduction

The most frequent task in proteomics projects is the identification of a protein sample based on an endoprotease digest, such as trypsin, and a sequence database search using the m/z values of the digested peptides. Such data are called "peptide map" or "mass fingerprint".

BioTools allows performing such searches on all available search engines, Internet access provided, and particularly operates in a seamless way with the MASCOT search program (Matrix Science Ltd., London). A prerequisite is a spectrum with annotated monoisotopic masses either from XMASS, XTOF or DA2 (see esquireSeries tutorial).

In XTOF the macro *Annotate.aura* does an automatic monoisotopic peak labeling in the range of 800-3500 m/z and performs the Autoextract command to eliminate a standard set of trypsin autodigestion masses and the 22 Da adducts from salt impurities.

L.2 Sequence Database Search using a MALDI Mass Fingerprint

Sequence Database Search using a MALDI Mass Fingerprint

An example dataset of this kind of data is in the DemoData directory. Please follow the described steps:

Open the demo data set: *..\Bruker\BioTools\DemoData\Flex\BSA digest LysC*. The red histogram-like peaks indicate the mass-intensity values from the mass labeled spectrum, if "*View/Picked Peaks*" is selected.

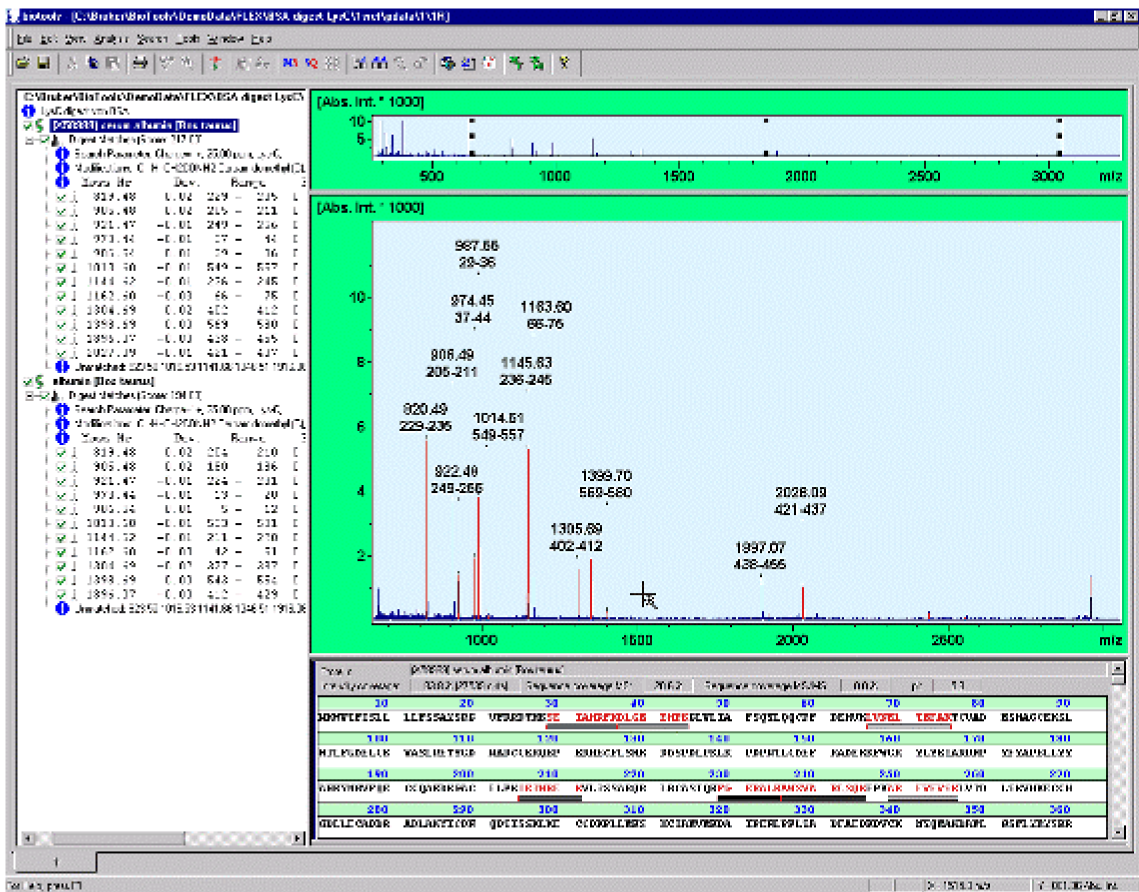


Figure L-1, Starting MS search

The fully annotated Spectrum appears, which is actually the result of a MASCOT search. The following description allows you to generate this information yourself.

L.3 Define the Search

Hit the **MS button** to open the MS data search dialog.

Note: *If no connection to a MASCOT Inter- or Intranet-server can be established, after a significant time you will receive an error message. The local address is of the type* ***http://<servername or IP-address>/mascot/cgi/nph-mascot.exe?1****. It must be*

added to the standard Internet address if you want to do local searches. If the Internet address could not be reached, you need to setup the Internet connection.

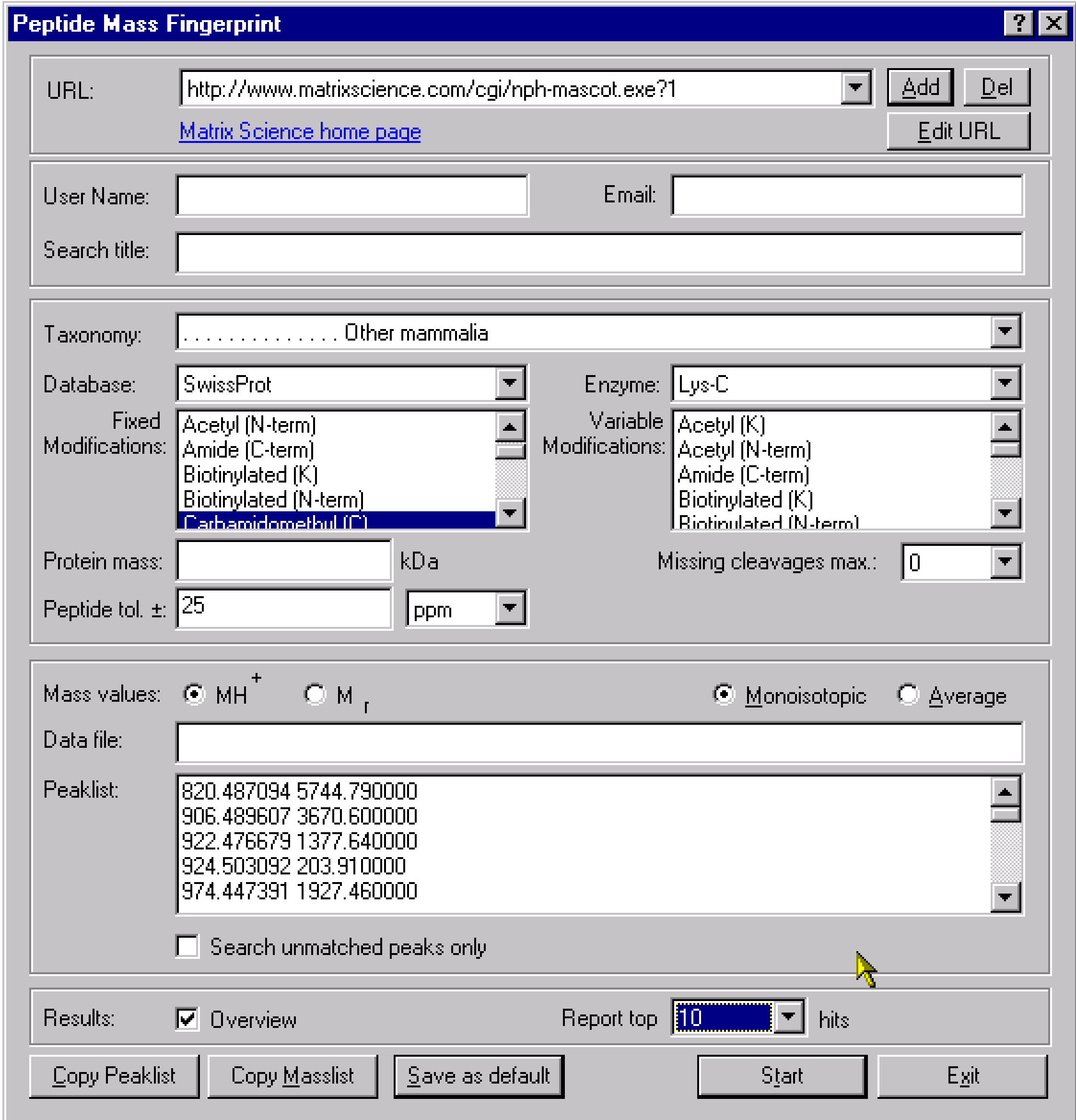


Figure L-2, MS search dialog for local search

Specify the information as in Figure L--2.

Hit the **Start** button for a MASCOT search.

Tutorial library searching and the BioTools search interface

Typically, **fixed modifications** include the known chemistry, such as reductive disulfide cleavage and carbamidomethylation, i.e., reaction with iodoacetamide.
Unknown chemistry, such as the artifact Methionine oxidation or phosphorylation can be specified as **variable modifications**. Allowing for optional modifications may reduce search specificity, but some, like Protein N-term Acetylation/ Formylation/ Pyroglutamylation may help identification of small proteins
Important is the **peptide tolerance** (or mass error), that can be a major source of frustration due to failed identifications, if the error estimation was a bit too optimistic! So: be sure about your data quality. Little rule: at MW 1000 expect as average 1000.5 as exact mass, at MW 2000 exact mass is 2000.0. Using this rule it is easy to estimate the correctness of your calibration. The exact rule is:

Δm = 1.00048 * INT(m), (Matthias Mann, 43rd ASMS Conference, 639)

Giving you the expected first decimal for any peptide ion mass
Typically, the precursor **protein mass** does not need to be specified.
The number of **missed cleavages** (or partials) accounts for tolerated internal cleavage sites in matching peptides. This number should be set to 0 or 1, since higher values reduce the specificity of the search as extensive use of **variable modification** does. If higher values seem to be required on a routine basis, you need to optimize your digest for more complete proteolysis. (In silver gels, destaining might help!).
After you set up the search parameters to a type of application you need to process frequently, hit the **Save as default** button to store conditions. Every spectrum is searched with this set of default conditions initially. If you then modify the conditions for a particular spectrum, the next time you open the spectrum, these last selected parameters will appear, allowing you to reproduce the last accepted result.
Hit **Copy mass list** to paste the mass list into the clipboard.

Note*: From the clipboard you can paste them into any browser-based search engine on the web, such as PeptideSearch, PepSea, Profound or MS-Fit. The search results from these programs, however, cannot be imported back into BioTools, in contrast to MASCOT.*

Hit **Copy Peak list** to paste the list of masses and intensities into the clipboard.
If a search result is already imported into BioTools, and a significant number of unaccounted peaks suggests the presence of another protein in the digest mixture, check the **Search unmatched peaks only** option. Now, all peaks of the list in the treeview of BioTools are not used for this 2nd round of searching. **Note**: *this approach simplifies the setup of a secondary search but may cause problems: elimination of some masses, which are shared by isobaric peptides from the different proteins, may prevent the search engine from identifying the 2nd protein.*

L.3.1. MASCOT Search Results

The basic information in the result header is the **top score**, it's **access number** in the database and the **entry name**, followed by a **histogram representation** of the 50 top scores. Within the green rectangle, the likelihood of a false positive match is 5% or more. Usually, only scores significantly outside this region (Scores > 70) are significant. Good values are > 100.

Attention: *The absolute score for the 5 % false positive likelihood is a function of the database and the search conditions. It may vary.*

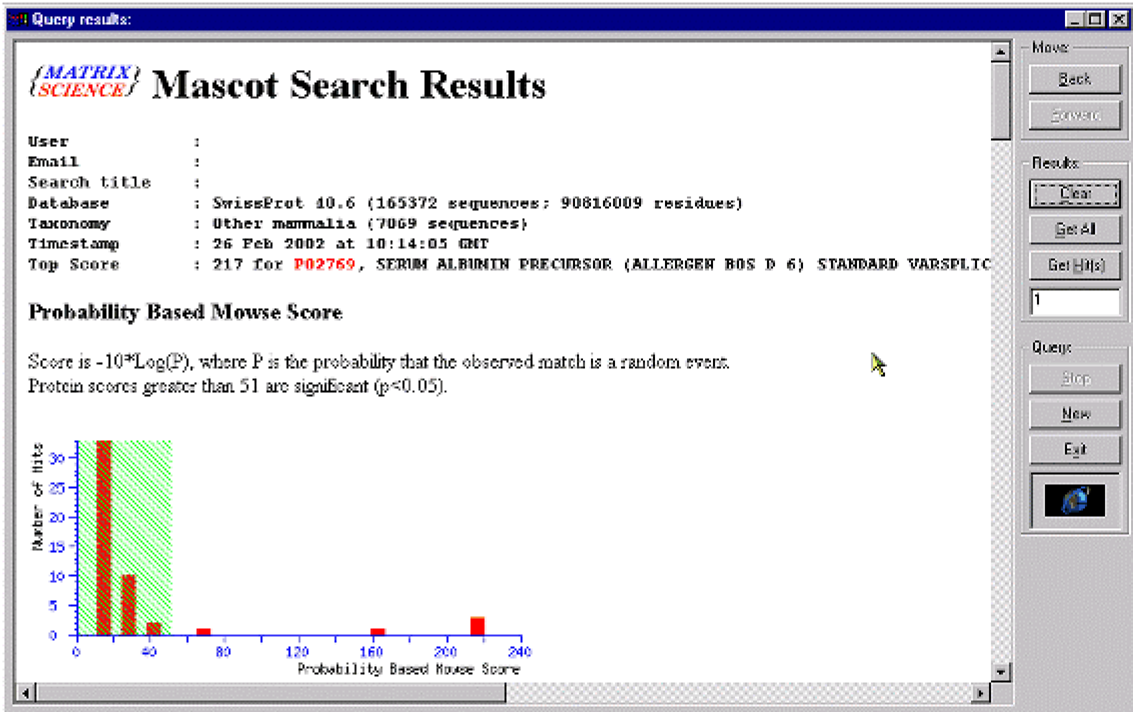


Figure L-3, MASCOT search results overview

To continue with importing the top hit into BioTools, press the **Get Hit(s)** button and continue reading chapter L.3.3 Work with Search Results in BioTools.

To get a short introduction of the MASCOT Query results interface continue with chapter L.3.2 Introduction to the MASCOT Query results interface.

L.3.2. Introduction to the MASCOT Query results interface

Since there frequently are more than one homologue or near identical sequences, splice variants, etc., even in nonredundant databases, the search result may be obscured. MASCOT offers for this case the **Concise Protein Summary Report**. Here, the sequences and scores of the highest scoring sequence for each cluster of homologue sequences is shown.

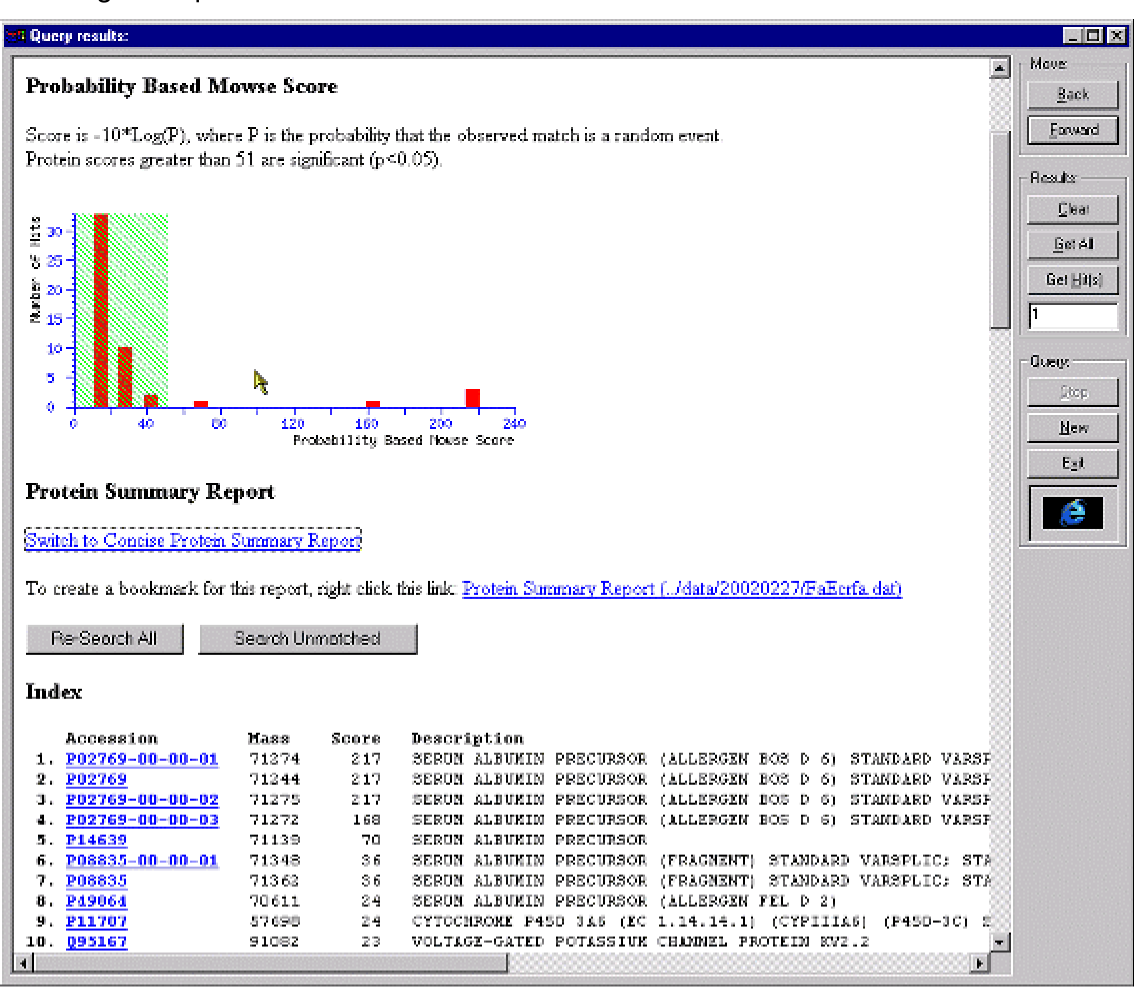


Figure L-4, MASCOT search results Mowse score

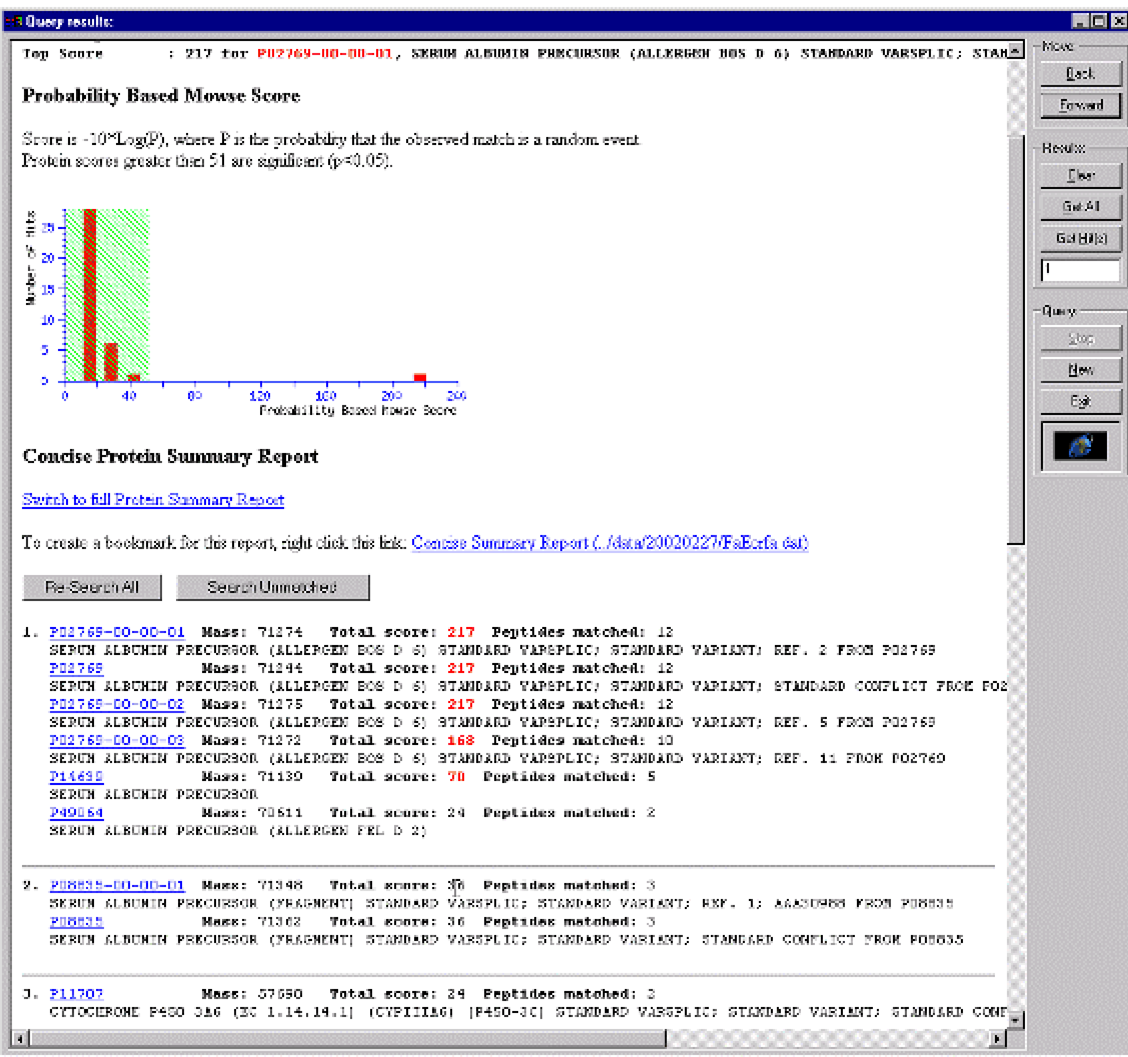


Figure L-5, MASCOT search results summary report

Further down the page there is a peptide match **overview** (Figure L-6), which allows checking the identity of matching peptides across the candidate sequences. Red circles indicate identical sequences, if one of them is under the mouse cursor. This is very useful to get a feeling for the relationship among the retrieved sequences.

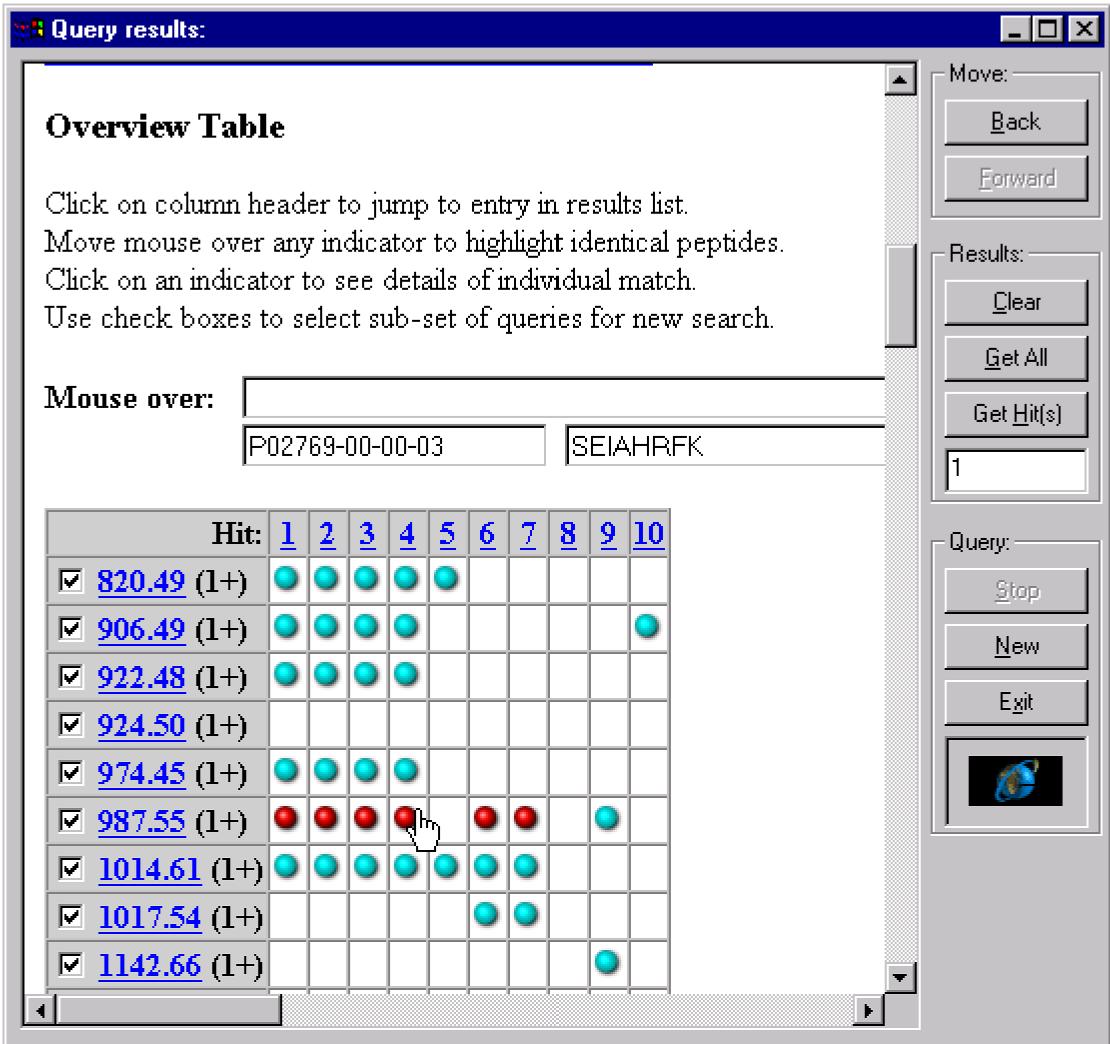


Figure L-6, Search results: overview table of matching peptides

Note ***The peptide overview appears only, if the*** **Overview option** ***was selected in the search dialog prior to searching***

Further down the page under **Index** there is a summary of the result with scores and sequences. The molecular weight of the proteins is often useful to tell false positives due to either excessively high (> 300 kDa) or very low (< 5 kDa) molecular weights. On this level you may decide, which are the entries you would like to visualize within BioTools.

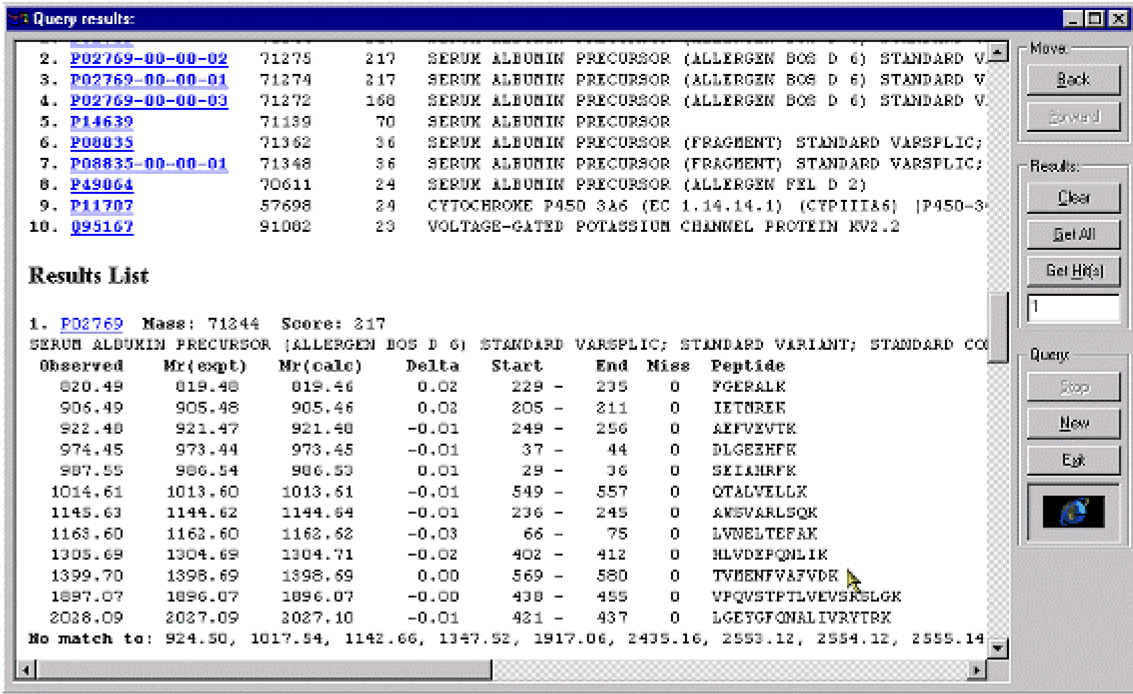


Figure L-7, Search results: index and matching peptides

Under **Results List** for each entry there is a detailed list of all matching peptides and the actual mass error in Da (Delta, irrespective of error dimension in the search). Through the hyperlinked access number you get to the Protein View, which contains a peptide coverage map of the full protein and all available information about the database entry (Figure L-8). Also an error plot is provided, which allows a simple mass error evaluation.

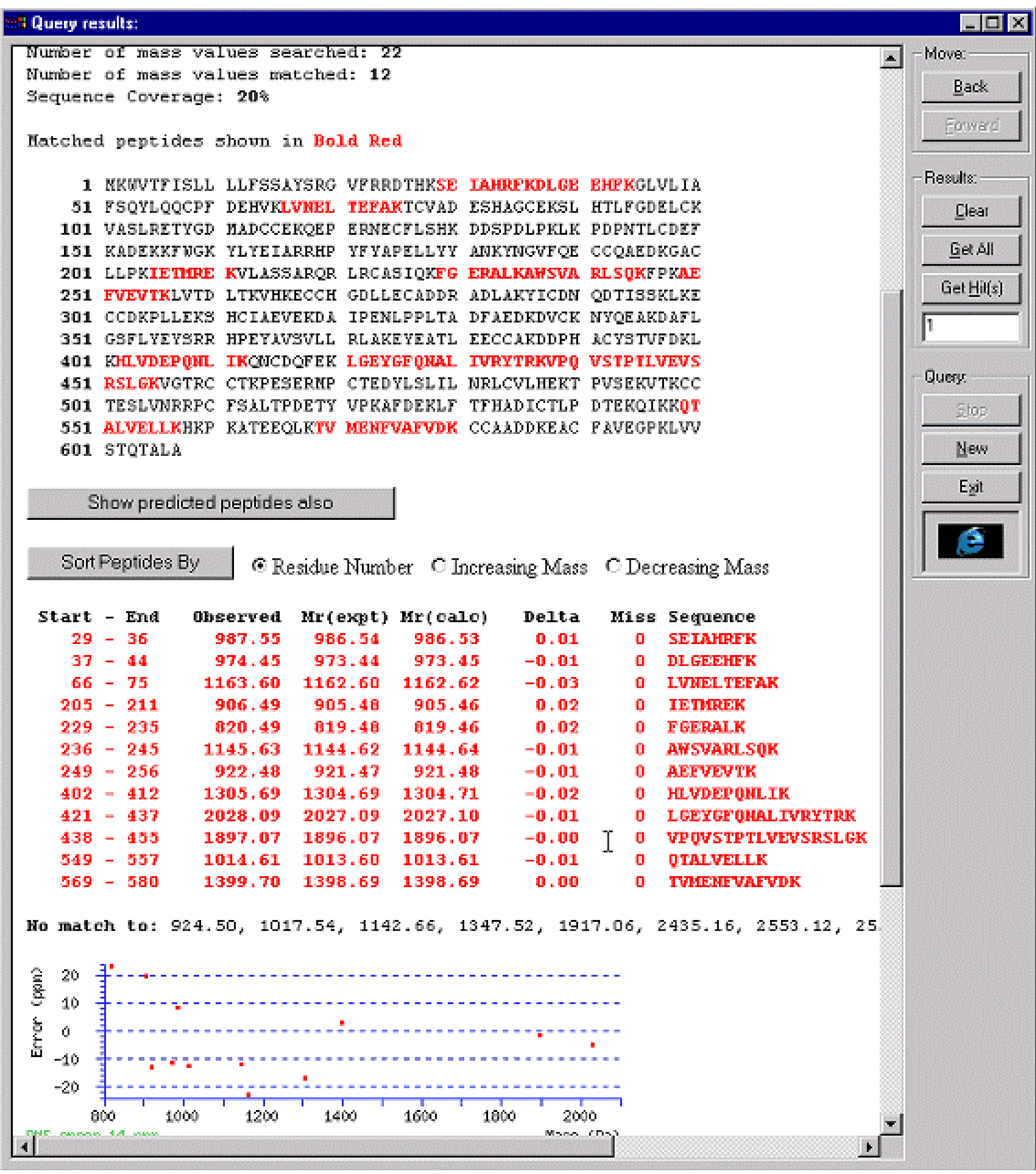


Figure L-8, Protein view to visualize sequence coverage of matching peptides

L.3.3. Work with Search Results in BioTools

If you decided to import candidate sequence # 1 from the MASCOT Query results window for further work in BioTools, hit the **Get Hit(s)** button in the Query Result window with **1** in the entry field below. If you would like to import, e.g., entries 1-4 and 8, specify "1-4,8". If you like to import all entries, hit the **Get All** button, but it is recommended not to do it, since the download of all sequences may cause long waiting times. In particular, if a database is accessed via the web and the total number of hits was selected to be >10.

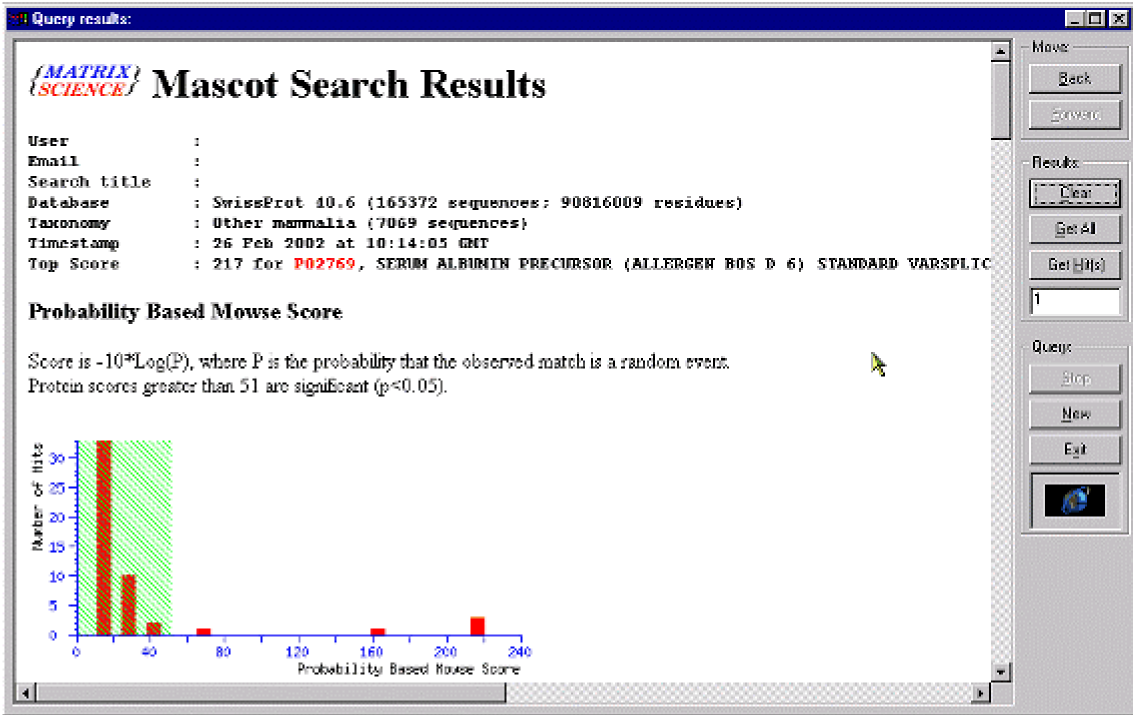


Figure L-9, MASCOT search results Mowse score

To clear the BioTools tree view from previously imported search results, press the **Clear** button before you import the new data.

If you like to do another search, e.g., a secondary search, which does not include all masses, which matched the first protein sequence, press **New** to return to the Search dialogue window.

If you like to exit the Query result page to continue working with the spectrum in BioTools, press **Exit**.

The Tree View

The tree view on the left side contains the data file information with up to two info lines – these contain the comments 1 and 2 provided by the operator during spectra acquisition. Information from the imported MASCOT search: the **list of sequence names**, which were retrieved, with a sublevel called **Digest matches**, which contains the MASCOT score. This is followed by the basic information about the **search parameters** and the **chemical modifications** specified.

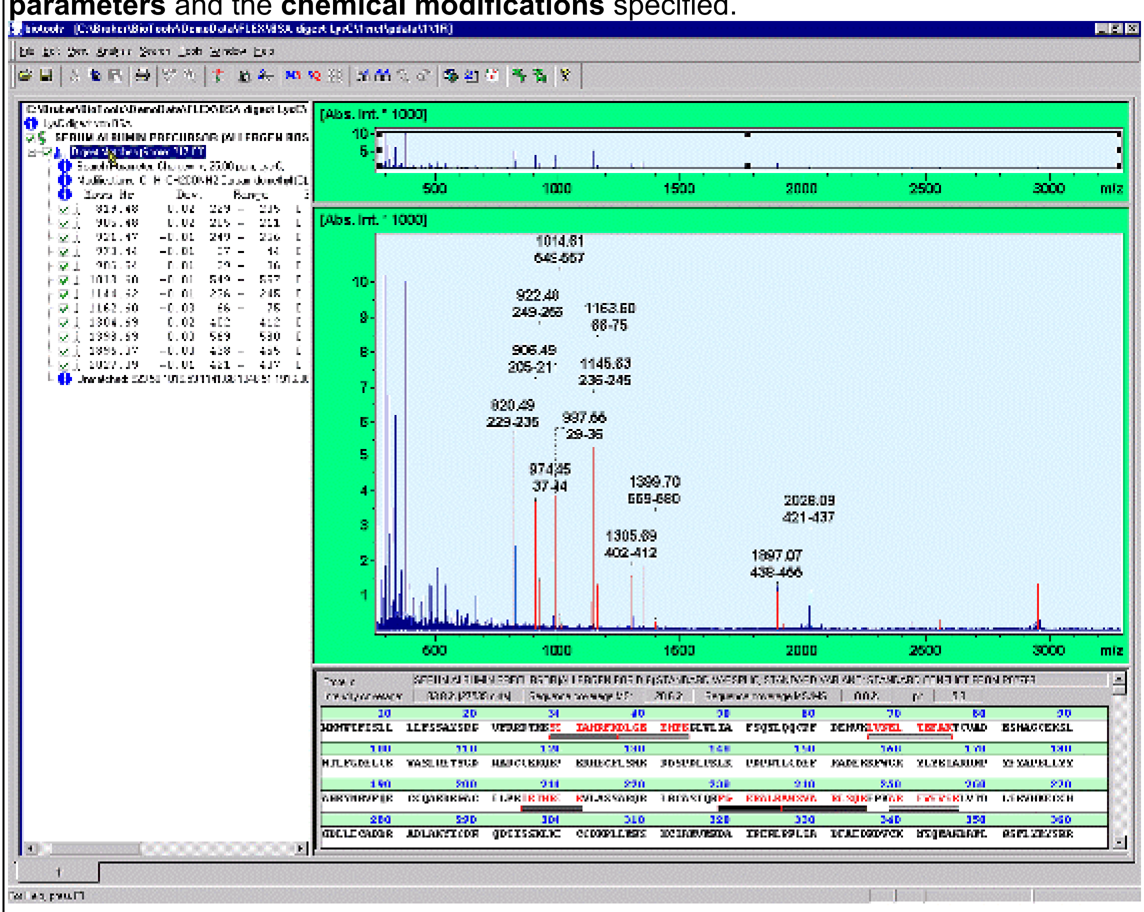


Figure L-10, Imported Search Result in BioTools

The most important **information in the tree view** is the list of peptides in the particular sequence, which match the experimental masses at the specified conditions of the search. Each peptide entry consists of **neutral experimental mass (Mr)**, **mass error (Dev.)**, **sequence range** of the peptide, **partials** (**P**, number of missed internal cleavage sites) and the **sequence**. Selection of the tree view entries at any level can be visualized in the spectrum as well as in the sequence view underneath the spectrum. Be it a single peak or a whole set of peptides if, e.g., Digest **matches** is

selected. The black numbers with additional sequence position information indicate matched peptides. Peptides, which contain an **optional modification** are color-coded in blue . **Red** mass labels in the spectrum indicate those peaks, which were **not matched** at the selected search conditions, if the display is set to display unmatched peaks.

Note: ***All masses in the tree view are always neutral molecular weights [Da].*** *They are* ***experimental neutral masses****, wherever possible. The mass values in the spectrum are data dependent and are typically MH+ for MALDI (m/z) or neutral M [Da] for charge deconvoluted ESI spectra.*

The Sequence Viewer

Simultaneous to the treeview information, also the sequence is loaded into the BioTools Sequence Viewer following the Get Result(s) operation in the MASCOT results window. The viewer is directly linked to the tree view as well as the spectrum, which means they all together display information about the same set of peptides within the same downloaded sequence. The matching peptides are represented here as bars underneath the covered sequence range, which allows you to visualize the information extracted from the spectrum on the sequence level. The view can be configured using the pull down menu opened by right mouse button click in context with the sequence viewer.

Important information about the global match between spectrum and sequence is displayed in the **header of the sequence viewer**:

Protein database entry text, and the Values for isoelectric point **pI** and molecular weight **MW** [kDa] contain plain protein sequence information and no details like modifications are used for these calculations.

The **Intensity coverage** gives an idea about the fraction of the currently selected peaks vs. the total picked peak intensities, which are related to the selected protein. A coverage of larger 80 % means that you achieved a fairly complete extraction of information from the spectrum, while coverage of 20 % means, you probably missed the point in analyzing the spectrum so far.

Sequence coverage MS is the fraction of annotated sequence in a mass fingerprint vs. the total sequence length. In MALDI fingerprints, this value typically varies between 10 and 90% depending on protein size and data quality - good quality spectra of small proteins may yield 90 % while larger proteins like BSA will yield only 15-30 %.

Sequence coverage MS/MS is the fraction of fragmented peptide bonds (b and y ions) vs. the total number of peptide bonds in the sequence and gives you an idea about the MS/MS confirmation level in particular of a LC-ESI MS/MS analysis.

The interactive work using this interface enables you to really judge the data from a mass spec point of view (mass errors, signal shape/intensity/ isotopic distribution) and from a protein chemical view. It is basically a result editing board from which you can initiate various further investigations and to which the respective results of are reported to for your further judgment.

The basic observations, assumptions to explaining them and the possible procedures to check them are:

Problem: Many peaks remain unaccounted for after import of a search result (Intensity coverage poor)

Assumption: There are more proteins in the mixture and I didn't find them all, yet: repeat the MASCOT search and select the **Search unmatched peaks only** option in the search dialog.

Assumption: Several peaks may match actually to the protein, but not in the simple way assumed for database searching. I want to check for higher mass deviations, tolerate more incomplete digestion or even unspecific cleavages (typically trypsin gives raise to further peptides resulting from cleavage after H, Y, W, F, L and I. I may even want to check for the presence of various suspected modifications or sequence errors or point mutations.

For further work on the identification of the unmatched peaks at this stage of analysis, please refer to the Sequence Editor Tutorial about Protein Digests, chapter P.5.1 Search for Unexplained Masses after MASCOT search.

Problem: Sequence coverage is too poor after import of a search result

Assumption: I (or a script) may have missed picking the weak peaks in the spectrum so far and need to find out: do a theoretical digest of the identified protein and send the predicted masses to the spectrum. Then add the missed peaks to the peaklist; please refer to the Sequence Editor Tutorial about Protein Digests, chapter P.2 Perform Enzymatic Digest, chapter P.3 Format the Digest Results and chapter P.4 Export Digest Results to Spectrum.

Assumption: I need to do an LC-ESI-MS/MS run for better coverage and want to set up a preferred or exclusion mass list. Do a theoretical digest of the identified protein and export the predicted m/z values to EsquireControl; please refer to the Sequence Editor Tutorial about Protein Digests, chapter P.2 Perform Enzymatic Digest, chapter P.3 Format the Digest Results and chapter P.4 Export Digest Results to Spectrum.

Problem: A particular peak remains unaccounted for in the mass fingerprint after all my efforts and I really want to know what it is!

Run an MS/MS spectrum (LIFT, PSD, ECD, whatever) first and try a library search in any case with that spectrum, even without enzyme specification. If it fails:

Assumption: The peak is related to an interesting, since unknown structural detail of my identified protein. Search for those masses in protein sequence and allow all thinkable modifications to occur and even allow tolerating single position sequence variations. Use the MS/MS spectrum to judge the calculated suggestions; please refer to the Sequence Editor Tutorial about Protein Digests, chapter P.2 Perform Enzymatic Digest, chapter P.3 Format the Digest Results and chapter P.4 Export Digest Results to Spectrum.

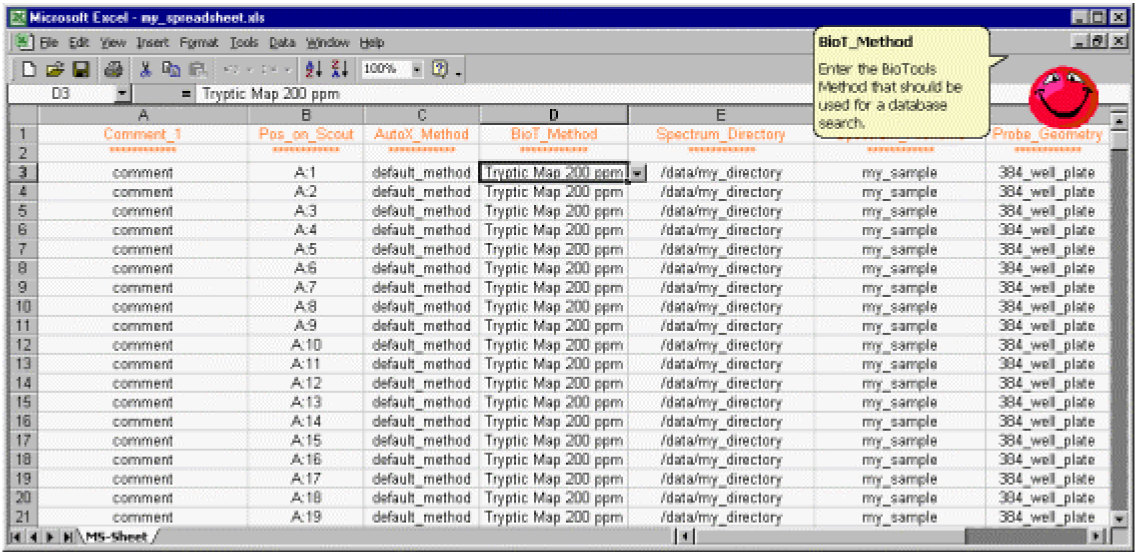
Assumption: The peak is related to another protein, which hasn't been identified in the mass fingerprint and it is not in the protein database. Try searching the ESTdb at the matrix science homepage first and *DeNovo* sequencing second (if the *DeNovo* features in BioTools are not sufficient: please contact ush@bdal.de for esquire support or dsu@bdal.de for flex support).

M Mascot Batch Mode

M.1 How to Automatically Acquire Spectra from a Batch of Samples M-2
M.2 Automatic Spectra Acquisition - Step by Step M-5
M.3 How to Perform a Protein Database Search with a Batch of Spectra......... M-7
M.4 Batch Searches - Step by Step.. M-8
M.5 Batch Processing of AutoLIFT data in BioTools 2.2 .... M-28
M.5.1. Import Spreadsheets M-28
M.5.2. Define Query Parameters... M-31
M.5.3. Batch Search M-32
M.5.4. Importing Search Results from Combined Datasets into BioTools.... M-34
M.6 CAF chemistry in BioTools M-37
M.6.1. MASCOT PMF Searches M-37
M.6.2. Search for Masses M-38
M.6.3. MASCOT MS/MS Searches M-38
M.6.4. Properties of LIFT and PSD Spectra M-38
M.6.5. RapiDeNovo M-38
M.6.6. References M-39

M.1 How to Automatically Acquire Spectra from a Batch of Samples

All sample-specific parameters are set by means of a MS Excel based sample spreadsheet that contains as many rows as spectra are to be acquired plus two header rows. The information for every sample is located in a respective row. Figure M-1 shows an example.



	A	B	C	D	E	F	G
1	Comment_1	Pos_on_Scout	AutoX_Method	BioT_Method	Spectrum_Directory		Probe_Geometry
2	************	************	************	************	************	************	************
3	comment	A:1	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
4	comment	A:2	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
5	comment	A:3	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
6	comment	A:4	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
7	comment	A:5	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
8	comment	A:6	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
9	comment	A:7	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
10	comment	A:8	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
11	comment	A:9	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
12	comment	A:10	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
13	comment	A:11	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
14	comment	A:12	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
15	comment	A:13	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
16	comment	A:14	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
17	comment	A:15	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
18	comment	A:16	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
19	comment	A:17	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
20	comment	A:18	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate
21	comment	A:19	default_method	Tryptic Map 200 ppm	/data/my_directory	my_sample	384_well_plate

Figure M-1, A Sample Spreadsheet for batch acquisition and database search

Since it is certainly not the most convenient way to type in a spreadsheet of 384 rows it is recommended to use MAP Control or – if you do not have a MAP robot – the Bruker Spreadsheet Generator (Figure M-2). This is an Excel environment that allows setting up of sample spreadsheets from a series of templates.

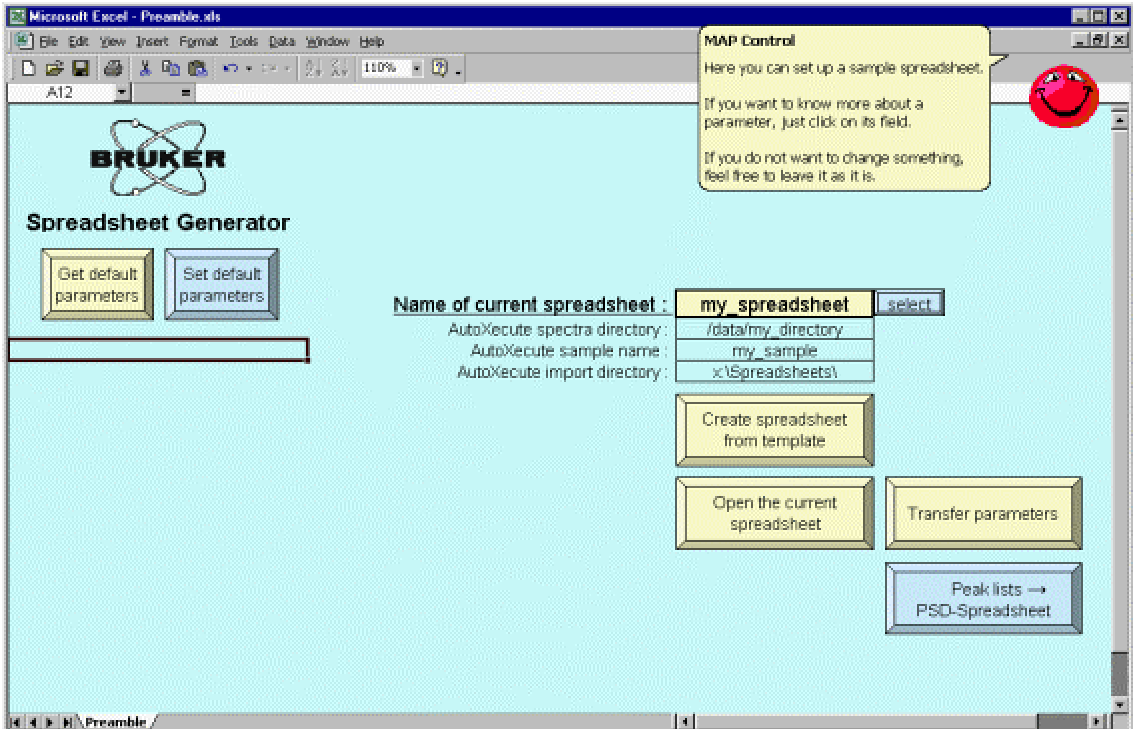


Figure M-2, The Bruker Spreadsheet Generator

The Excel environment contains several features for setting up a sample spreadsheet. The operator can specify the name of a new or existing "current spreadsheet". Further, the directory and the filename that AutoXecute uses for saving any spectra can be specified here. (Alternatively, this can be done in the sample spreadsheet individually for every sample.) If the AutoXecute import directory of the MS workstation is properly mounted on the PC (e. g., the directory /data/Spreadsheets appears as network directory x:\Spreadsheets, see Note on page M-3) this directory should be entered in the field "AutoXecute import directory".

Three buttons are available to set up a new sample spreadsheet (using a template that can be modified in the desired way), to open an existing spreadsheet, and to export the sample spreadsheet to the automation import directory of AutoXecute on the MS workstation.

Since the spreadsheet template is self-explanatory (by means of online help when clicking on a column header) it is described here only very briefly. The operator can enter a comment for every sample in the first column (header: Comment_1). The target positions are specified in the column with the header Pos_on_Scout. For every target position one must specify an AutoXecute Method (header: AutoX_Method) that

sets all parameters necessary for an automatic acquisition run. The directory and filename for the acquired spectra is specified in the columns Spectrum_Directory and Spectrum_Filename. For more detailed information see the online help in the spreadsheet template.

Note ***BioTools, MAP Control and the Spreadsheet Generator can communicate with MS workstations both under Solaris ("Sun") and under Windows NT. However, the path information for the spectra directory and for the AutoXecute import directory depend on the operation system on the MS workstation:***

	Spreadsheet Generator running locally on the MS workstation under Windows NT	**Spreadsheet Generator contacts MS workstation over the network – MS workstation under Windows NT**	**MS workstation under Solaris ("Sun")**
AutoXecute spectra directory (appears as Spectrum_Directory in the spreadsheets)	d:\data\my_directory	d:\data\my_directory	/data/my_directory
AutoXecute import directory	d:\Spreadsheets	x:\Spreadsheets (The directory d:\ on the MS workstation has been mounted on the PC as network drive x:\)	x:\Spreadsheets (The directory /data on the "Sun" has been mounted on the PC as network drive x:\)

M.2 Automatic Spectra Acquisition - Step by Step

Procedure:

- Prepare your samples on a MALDI target. The following steps are described for a target holding 384 samples, but just a few samples will do just as well.
- Open the Excel environment "Spreadsheet Generator".
- Make up a name for a new sample spreadsheet (e. g., my_spreadsheet) and enter this name in the field "current spreadsheet".
- Enter a desired directory (of the MS workstation) where AutoXecute is supposed to save the spectra, in the fields "AutoXecute spectra directory" and "AutoXecute sample name".
  If you do not change the default entries the spectra will appear under /data/my_directory/my_sample (or d:\data\my_directory\my_sample – see Note on page M-3).
- Click on "Create Spreadsheet from Template"
- Select a suitable template, e.g. "AutoX_BioT.xls". A spreadsheet named "my_spreadsheet" containing 384 rows will open up.
  The spreadsheet might contain too many rows. Just delete the rows that you do not need. The spreadsheet should contain only those target positions (under Pos_on_Scout) that you have prepared samples on.
- In the column "Comment_1" you might want to enter the samples' ID numbers, names etc.
  In the column "AutoX_Method" you can select acquisition methods. Select "calibrate" for external calibration on a peptide mixture (Bruker Kit 206195) and "measure" for acquisition, recalibration and annotation of the sample spectra. We recommend a near-neighbor calibration-set like in template "Autox_batch_squares".
- Save the spreadsheet (just press save, do not save it under a different name).
- Close the spreadsheet to get back to the Excel environment.
- Press "Transfer parameters" to transfer the spreadsheet via network to the AutoXecute import directory of the MS workstation.
- Press "Transfer parameters" to transfer the spreadsheet via network to the AutoXecute import directory of the MS workstation. (This transfer step is always necessary – even if the Spreadsheet Generator runs locally on the MS workstation.)

Now you are ready for automatic spectra acquisition and processing by AutoXecute. Insert the SCOUT target into the mass spectrometer. If the "AutoXecute import directory" has been properly set (and mounted!) in the Spreadsheet Generator, the sample spreadsheet will be – upon "transfer parameters" – automatically transferred to the workstation of the mass spectrometer.

Open XTOF. This program has to be active (at least in the background) for automatic processing.

Open XACQ / flexControl and load an appropriate parameter file. (You might wish to check manually weather the current instrument parameters produce a good spectrum quality.

Under Solaris (XACQ):

Now open the AutoXecute sample list (under the menu "AutoXecute" – edit sample list) and import the sample spreadsheet of interest (click the button "import" and select your spreadsheet). Click "start" in the sample list window and the instrument will begin to acquire and analyze data according to the "AutoX_Method" settings in the sample spreadsheet.

Under Windows NT (flexControl):

In the left-top corner of the user interface there is a field AutoX. Here you can select your transferred spreadsheet (As text file in the directory c:\Spreadsheets). The spreadsheet will not appear in flexControl. Just press "Start automatic run" to start the automatic run. (In the AutoXecute tab you might see the note "seq. file not flex control native". Do not worry – just start the run.)

After an AutoXecute run you might want to do some PSD experiments on the samples that are still in the mass spectrometer. To generate an Auto-PSD spreadsheet re-open the "Parameters of MAP Control" and click the button "Peaklists → PSD-Spreadsheet". Now the peaklists of your automatically acquired mass spectra are imported and a PSD-spreadsheet is generated. In the column "PSD_ParentMass" you can select the ion you wish to fragment from small popup menus. The PSD pulser settings will be calculated according to the selected parent masses. After selecting all parent masses of interest save and close the spreadsheet and press "Transfer Parameters and Start Preparation" once more. Now you can "import" the spreadsheet into AutoXecute and "start" the PSD-run.

M.3 How to Perform a Protein Database Search with a Batch of Spectra

BioTools supports a **Batch Mode** that can be used to set up a task, i.e., a batch of spectra combined with search methods, to perform a batch-wise database search and to access the results of all searches. In addition, the results of such automated processes can be manually further refined and these results then be accepted for inclusion into the batch result report.

The **MASCOT Batch Mode** is opened with the button and contains six tabs. The **Status** tab shows the current status of a task with all search results. From here database searches can be started for all spectra of a task or for selected spectra. The **Scout MTP** tab shows a scheme of the SCOUT target with color-coded search results. Here, a click on a position leads to the display of the respective spectrum and search result (For a detailed description assessing an individual search result see tutorial Sequence Database Search using a MALDI Mass Fingerprint, chapter L.2.2 MASCOT Search Results). The **Task Editor** tab is used for setting up a new task by adding spectra to the list or by importing a spreadsheet "from AutoX". This is also the place to link spectra to a BioTools method. The **Query Parameter** tab contains all search parameters , which are stored in a BioTools method. Existing methods can be edited; new methods can be set up. The export functions of BioTools can be set in the **Summary Report** tab. From here it is possible to print out the result of a batch search as a table or to export the results for further use by Excel. The report can be directly communicated by email from here; parameters are set up in the **Mail parameter** tab.

M.4 Batch Searches - Step by Step

The batch search is described here with the BioTools demo data, which can be found under c:\Bruker\BioTools\DemoData\Flex\Batch_Data. This is the "Spectrum Directory" that holds the sample name (BSA) as well as the respective spreadsheet (BatchSearch.txt).

Procedure:

- Open BioTools. Close all spectra and open the MASCOT Batch Mode window with the button.
- Activate the Task Editor tab. Click New to set up a new task.

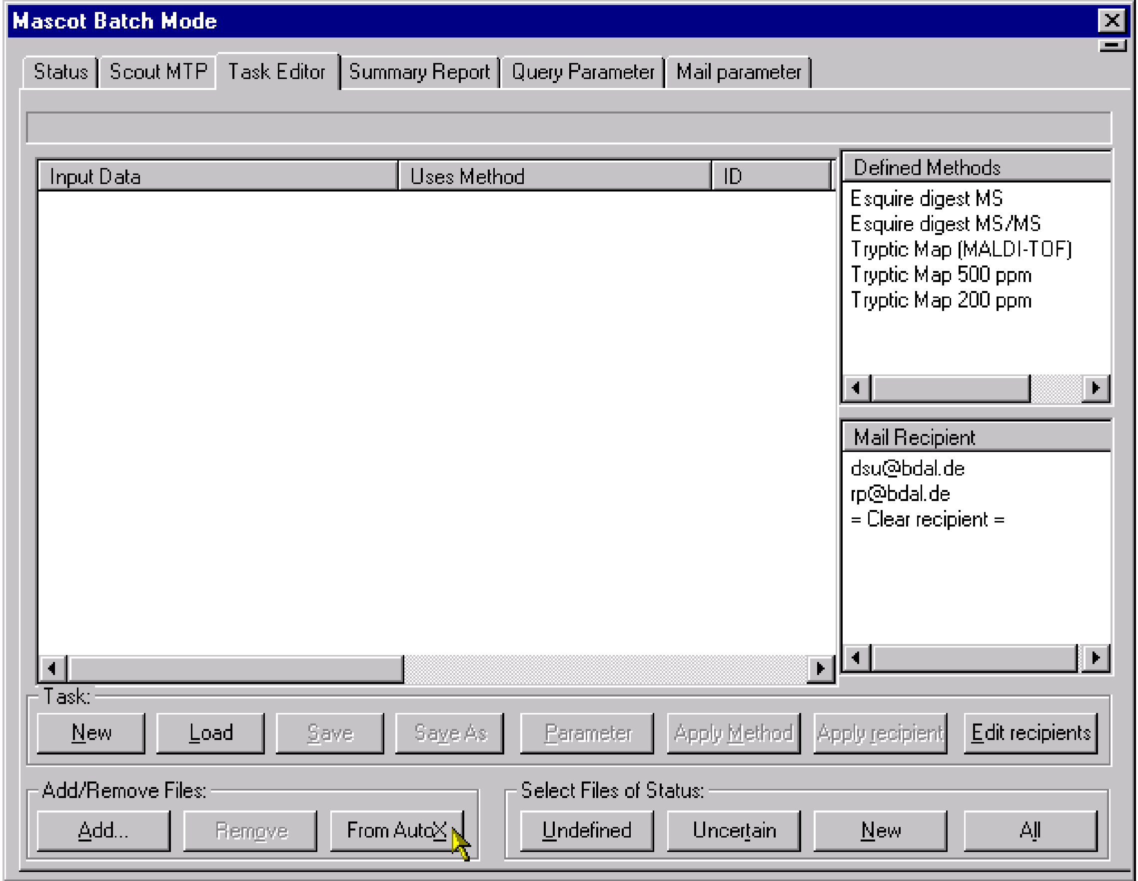


Figure M-3, The Task Editor tab

- Click on From AutoX to import a spreadsheet and the according spectra.
- Next to the field "data spreadsheet" click "browse" and look for the spreadsheet describing the batch process: *BatchSearch.txt* in the DemoData data directory.

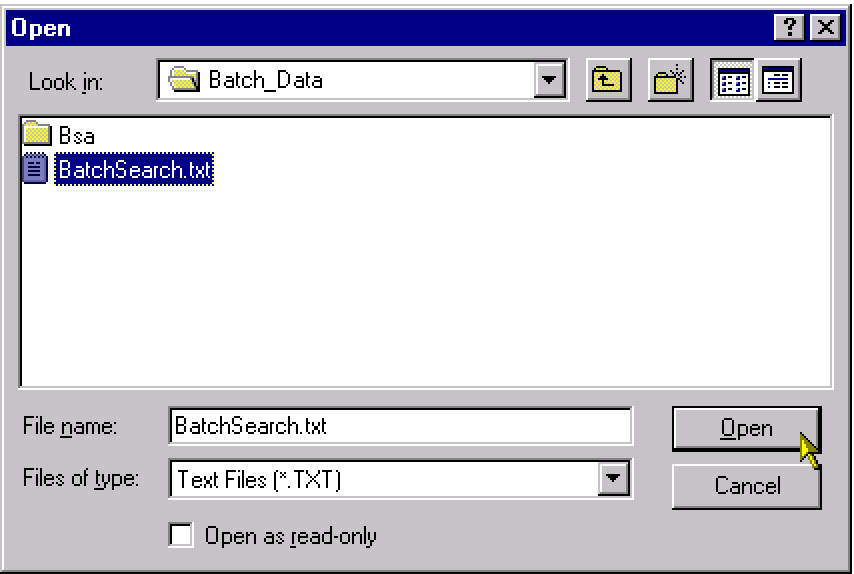


Figure M-4, Open dialog

- After pressing open the spreadsheet's name should be visible in the Data Spreadsheet field. Now press Process Spreadsheet. If AutoXecute has properly acquired the spectra and generated all peaklists there should be a message like "383 peaklists found. Press OK to get back to the batch window.

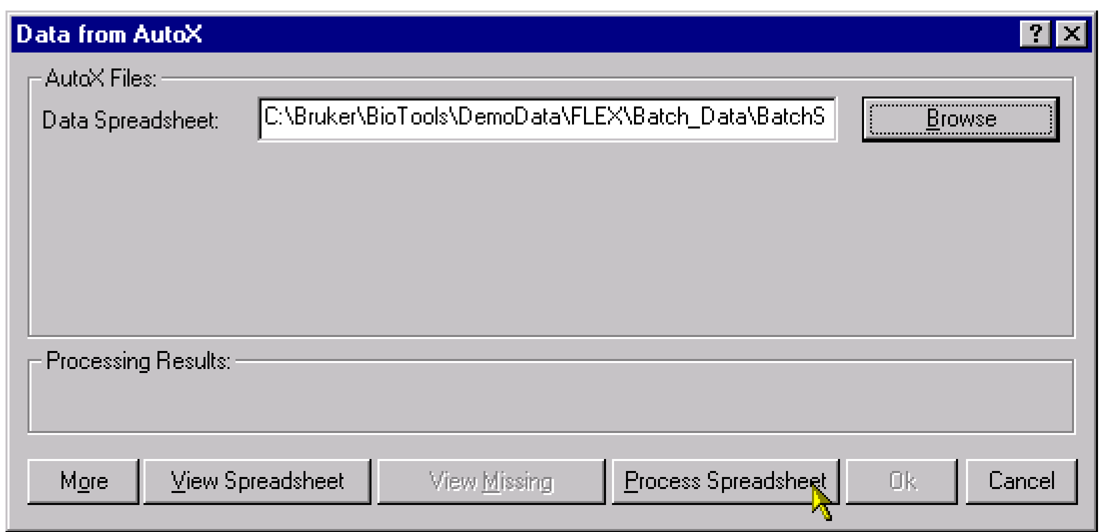


Figure M-5, Data from AutoX

Note	***If you get the message here: "0 peaklists found, xxx missing, there has been a problem finding the peaklists in your data. If you click "view missing" you will see a list of all missing peaklists with the paths that BioTools expects. Make sure that these paths do really contain the files* peptidepeaklistor peaklist.xml *generated by XACQ or flexControl, respectively. By default BioTools looks for* peaklist.xml *under the experiment* 1SRef*. If AutoXecute has saved the data as "*SRef*" you must change the BioTools search string. In the window From AutoXecute press more and you can select either* SRef *or* 1SRef*. Just click on* SRef *and process the spreadsheet again. If you cannot find a file named* peaklist.xml *or* peptidepeaklist *in the subdirectories of your data, you should check if you have used an AutoXecute method or an XTOF macro that generates proper peaklists. An example for a valid AutoXecute method is the standard method* measure_new*, which runs the XTOF macro "*Annotate.aura*.***

- Now you should see a list of your data sets. If the spreadsheet contained a column "BioT_Method" the respective BioTools methods have already been added. If not, you must apply a BioTools method for every spectrum: select data sets or ranges, select an entry under Defined Methods and press the Apply Method button.

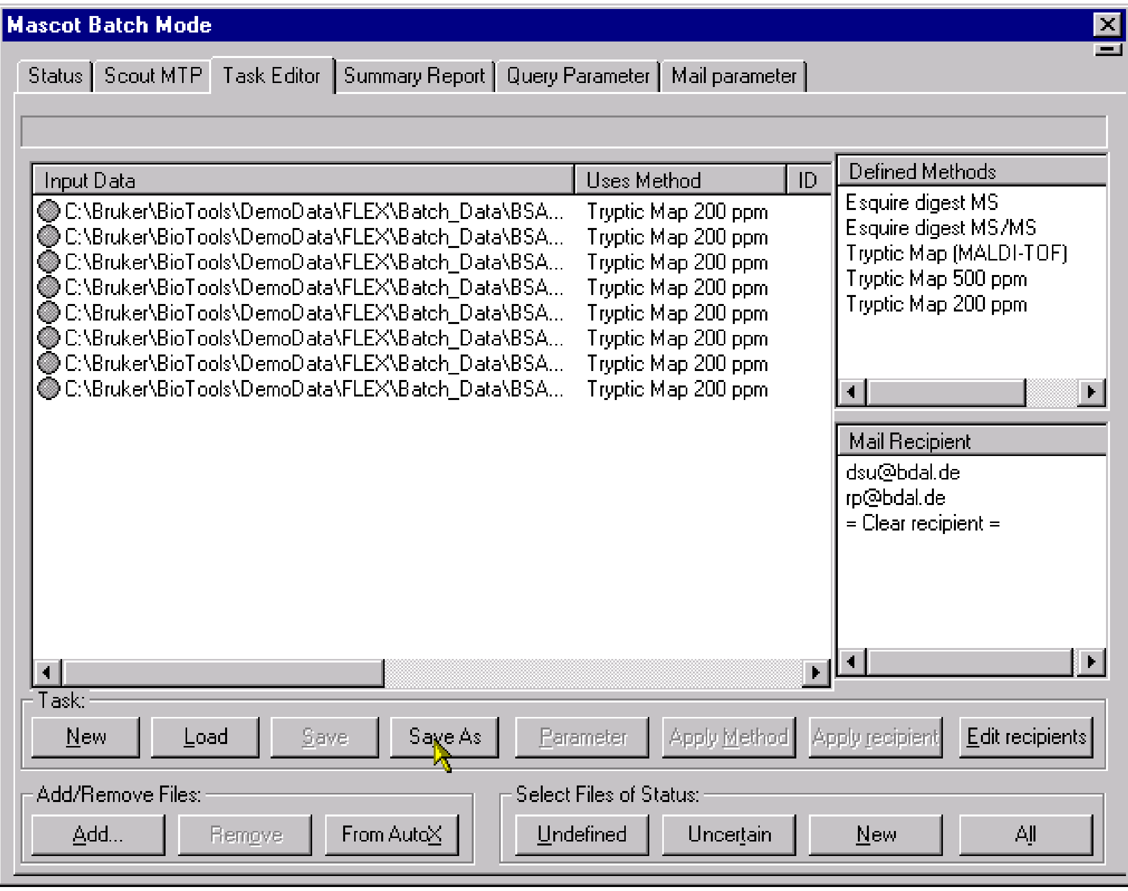


Figure M-6, The Task Editor tab

- Click on "Save as" to save the task under a name you specify, e.g., *DemoDataTest*. Typically, it is a good idea to select the name of the spreadsheet.

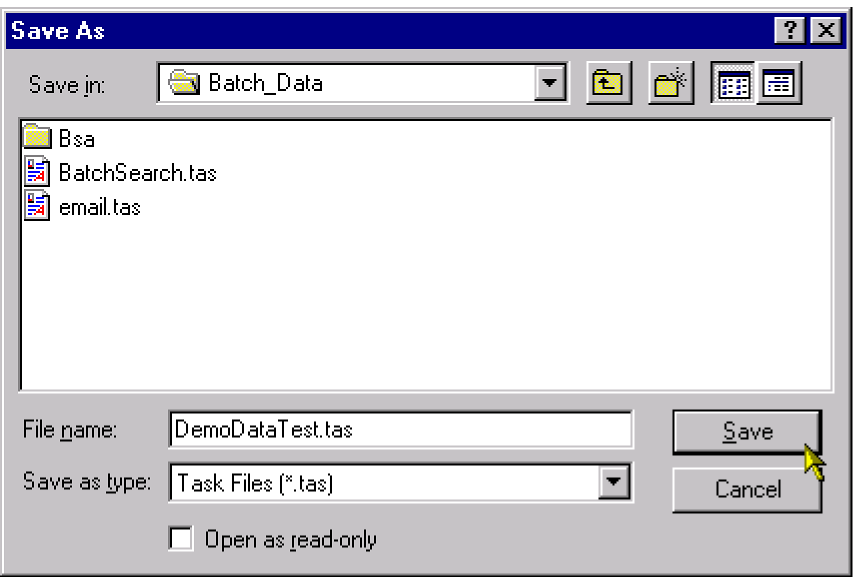


Figure M-7, Save As dialog

- Select the Status tab and Start the search .
- For the example run with the DemoData, we will first define a new BioTools method and configure the batch reporting:
- Switch to Query Parameter tab, press the Method Edit button and add a new method name.

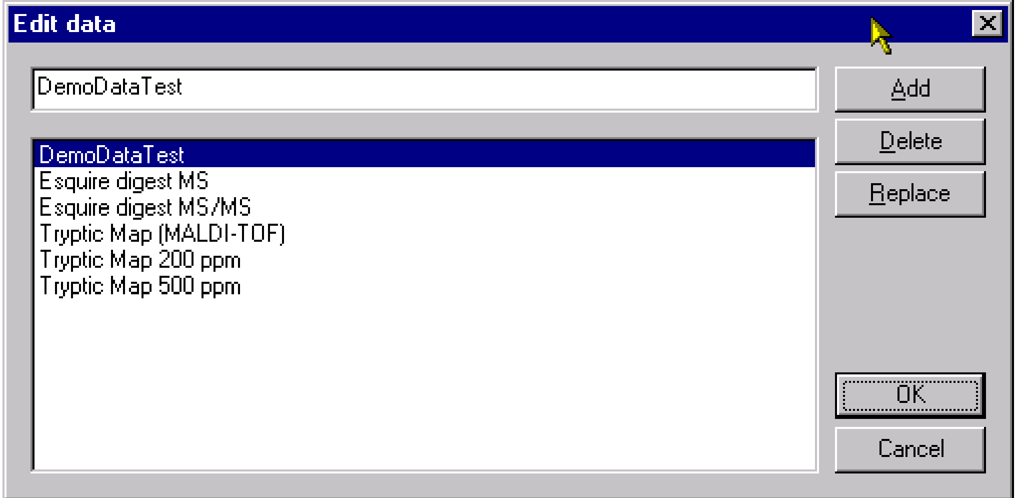


Figure M-8, Edit data dialog

- Define the search parameters according to Figure M-9 and press the **Apply** button. A new method *DemoDataTest* is now available, which we shall later select.

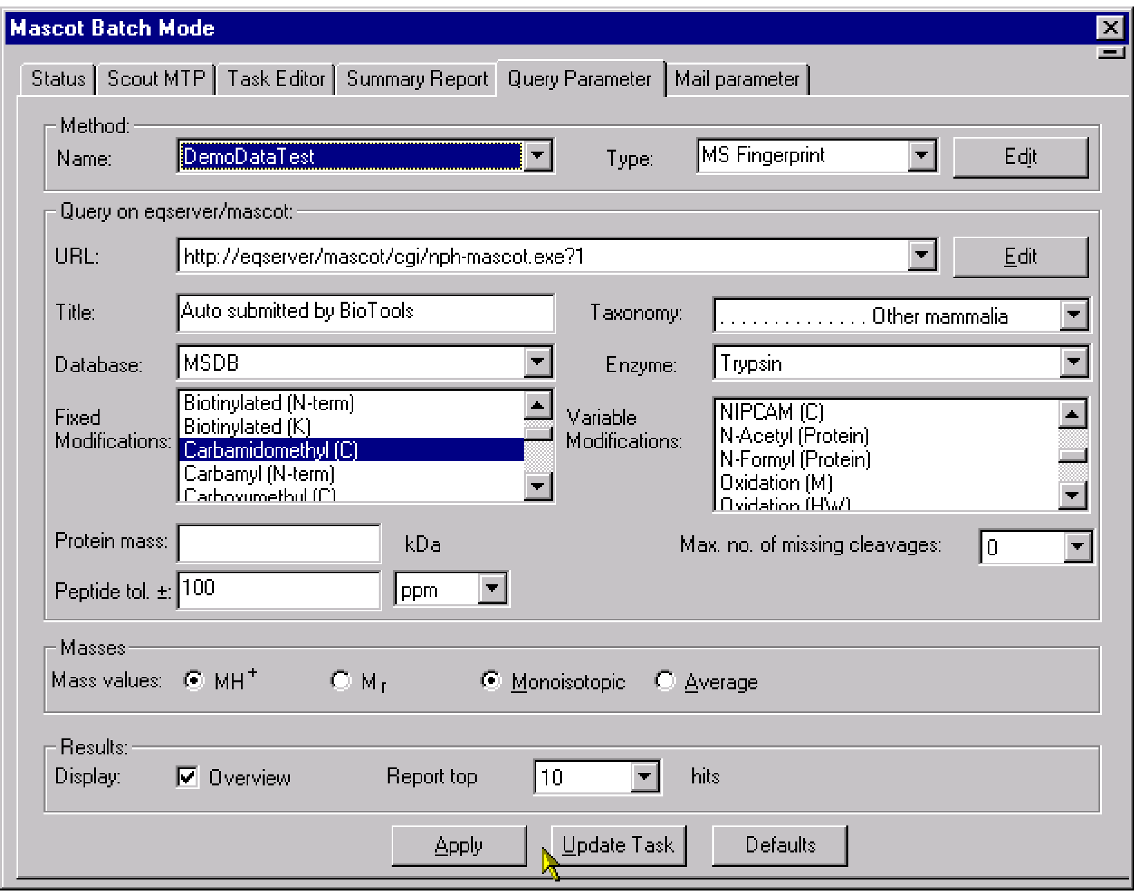


Figure M-9, The Query Parameter tab

- In the Summary Report tab we may add or remove individual parameters from the Report list by selection and by pressing the appropriate button Remove or Add. The order of appearance can be changed by selection and pressing of the Up and Down buttons.

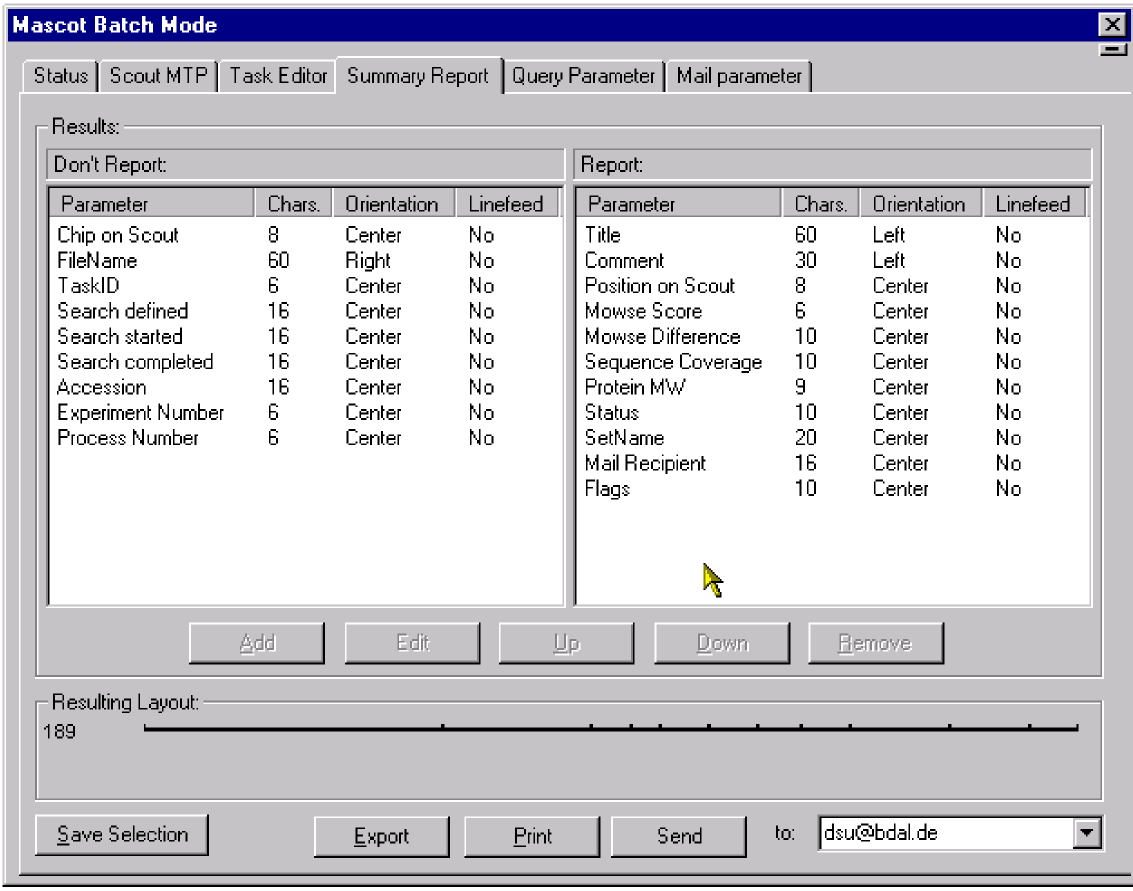


Figure M-10, The Summary Report tab

- In the Mail parameter tab mail reporting can be defined. In Figure M-11 reporting is defined such to automatically send the summary report to the email recipient but not the individual MASCOT reports for each spectrum (please don't use the mail addresses shown there, add your own address for test purposes to the list of recipients).

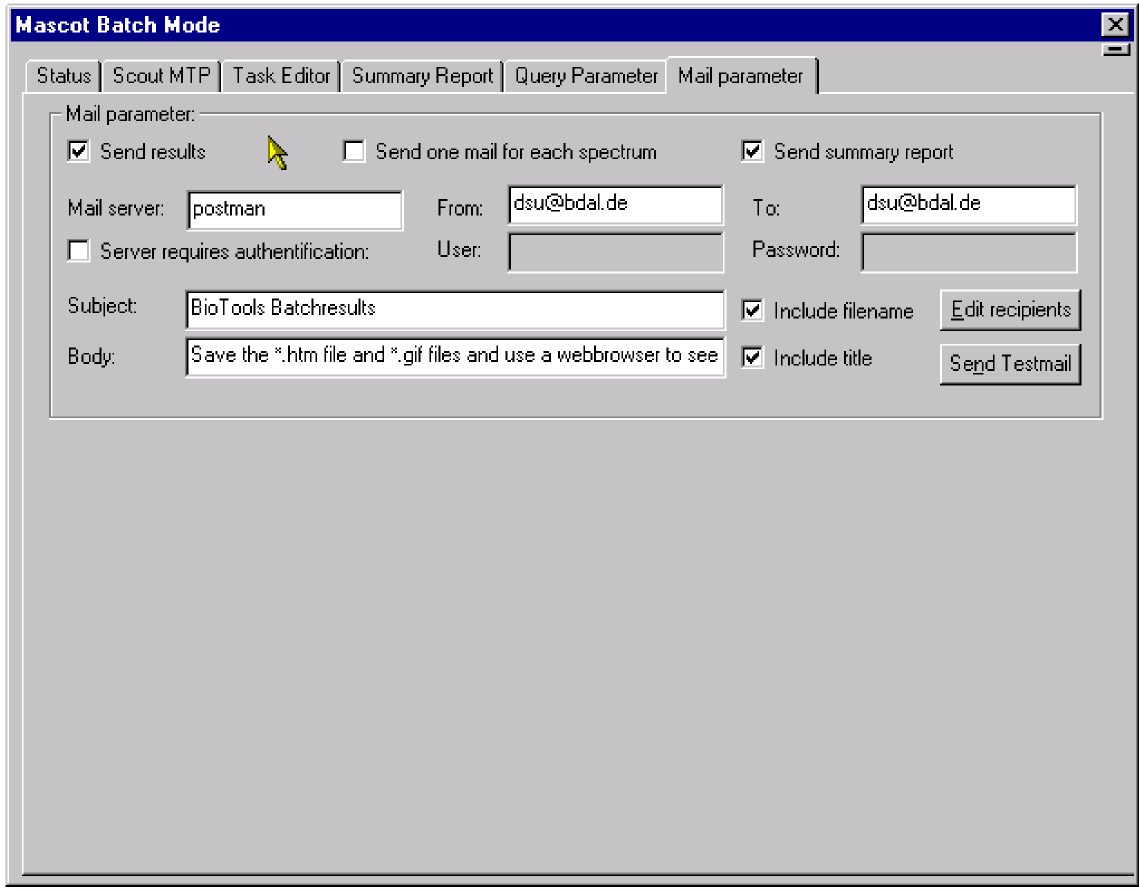


Figure M-11, The Mail Parameter tab

- In the Task Editor tab, All data sets are selected and the proper email recipient is selected. Pressing the Apply recipient button links every data set to email reporting.

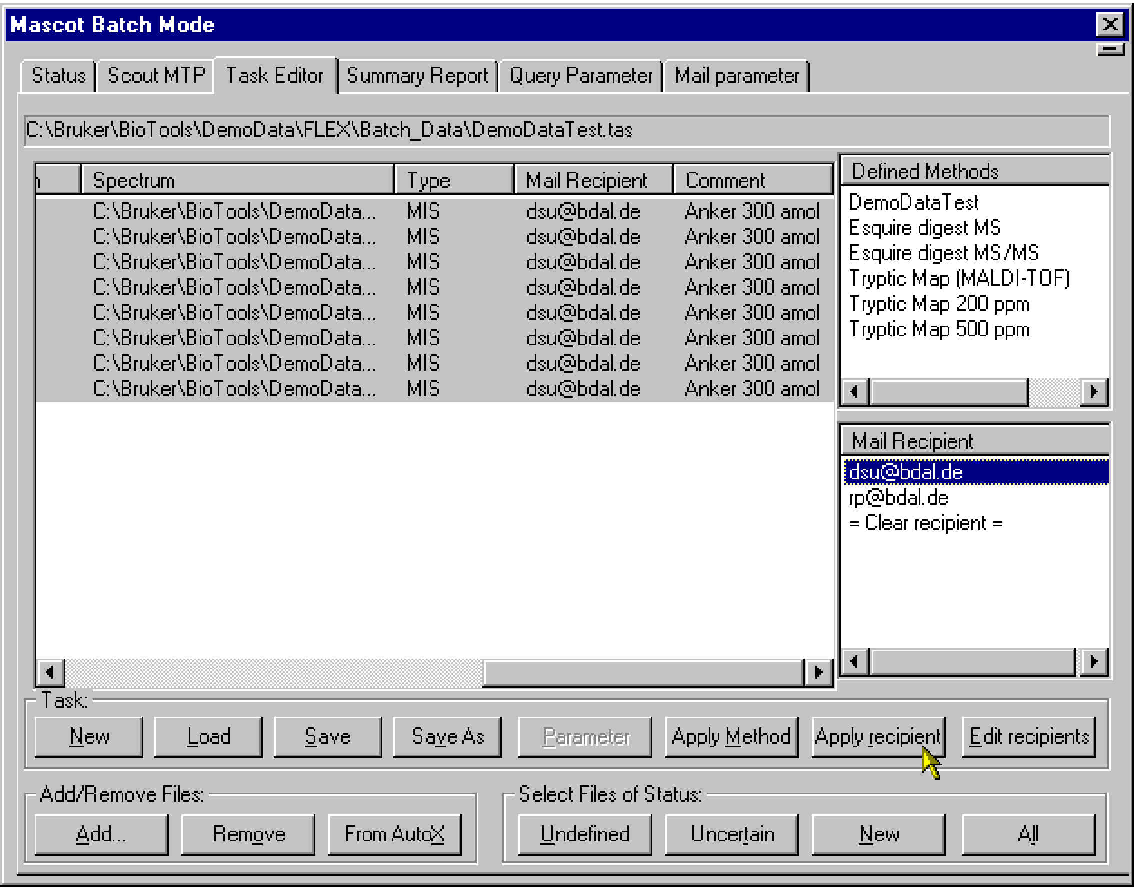


Figure M-12, The Task Editor tab – Apply recipient

- In the Task Editor tab, All data sets are selected and the proper search method *DemoDataTest* is selected. Pressing the Apply Method button links every data set to that method.

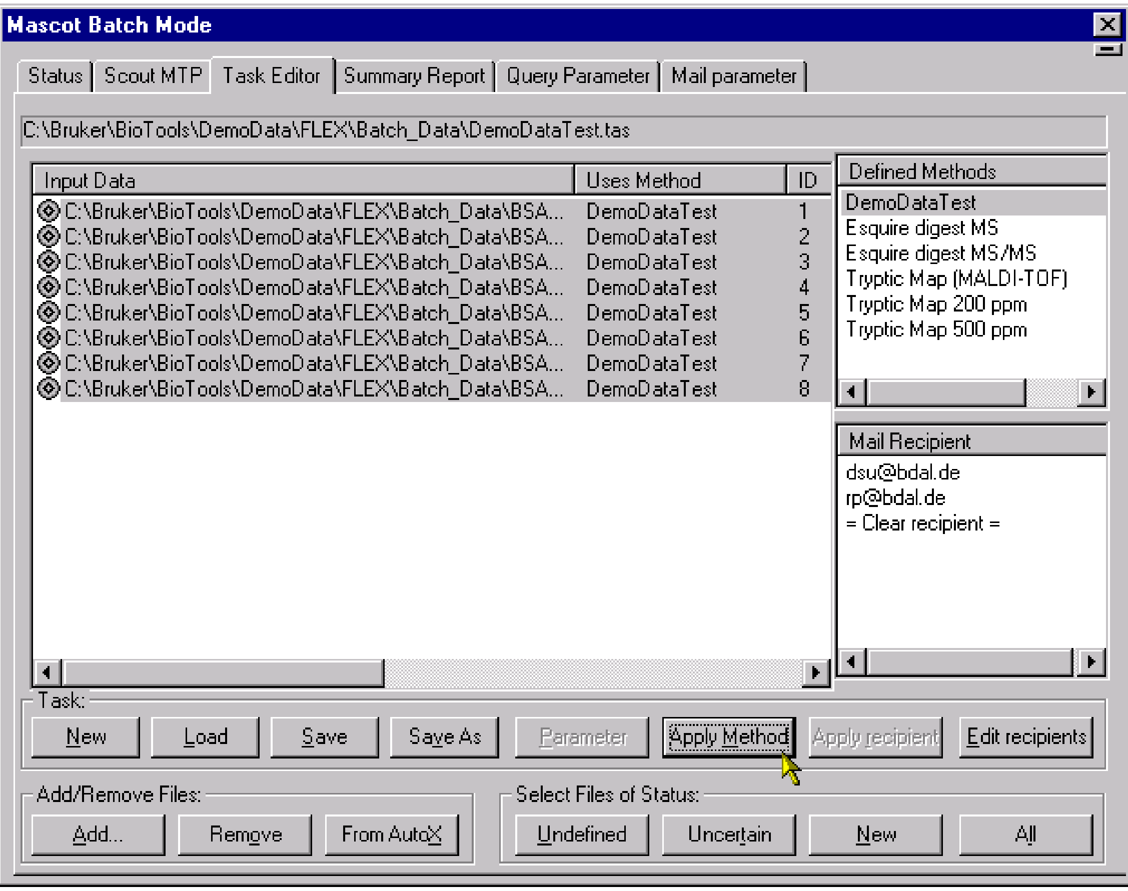


Figure M-13, The Task Editor tab – Apply Method

- In the Status tab select All data sets and Start the task.

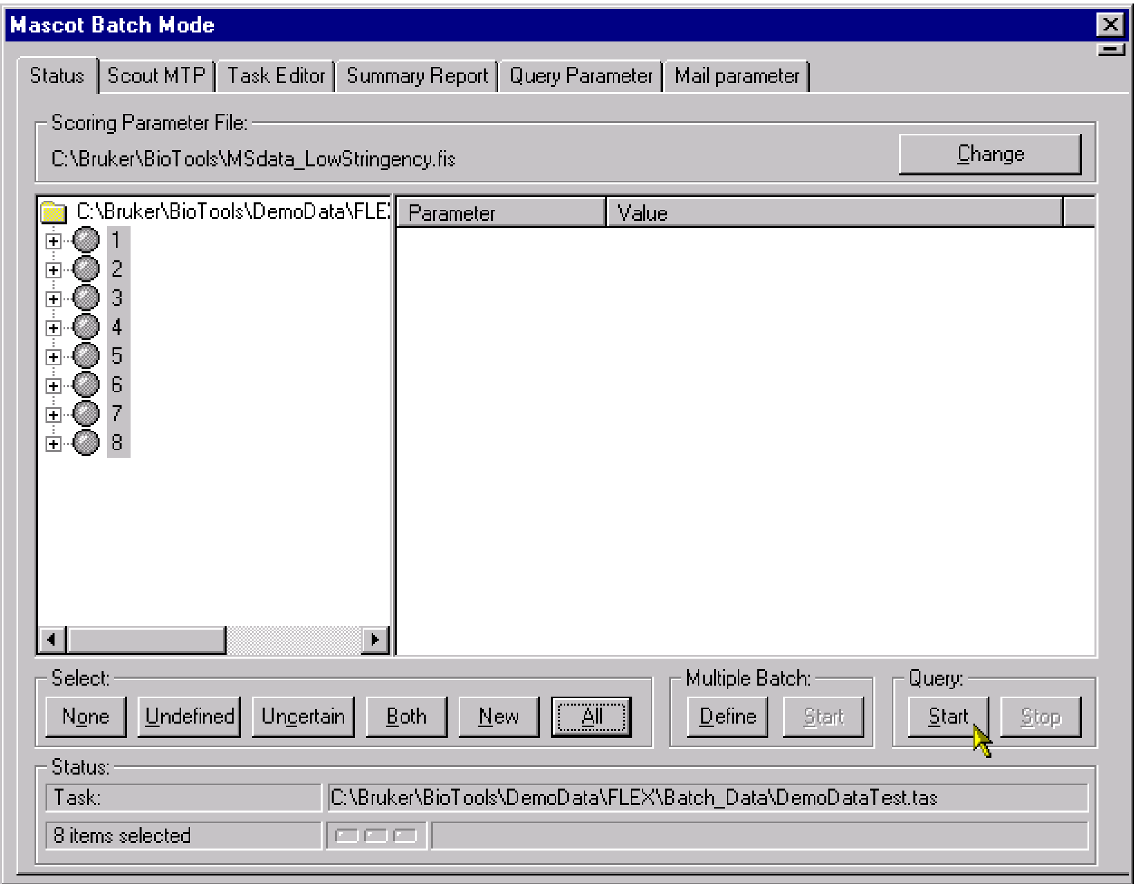


Figure M-14, The Status tab – start

- In the Status tab you can follow the process in real time.

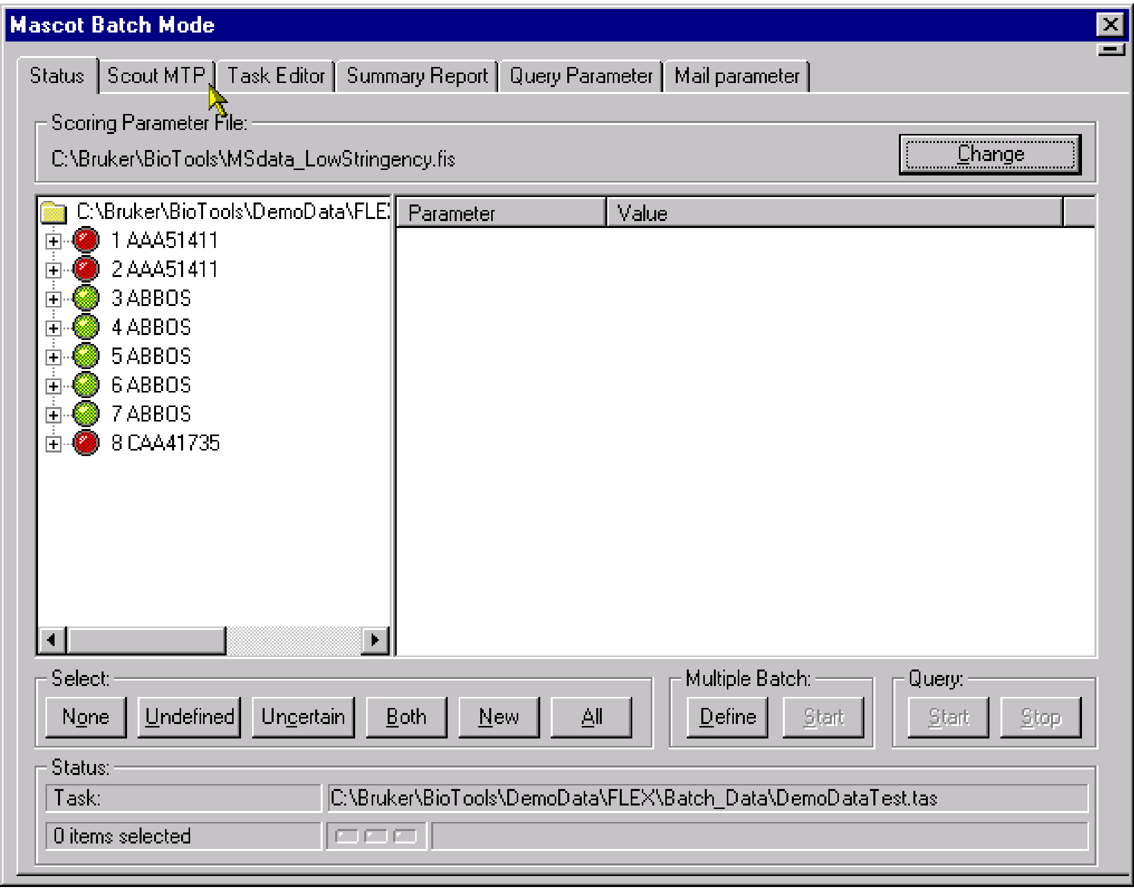


Figure M-15, The Status tab – processing

- Finished searches can be viewed by double-clicking the respective entry while other searches are going on (For a detailed description assessing an individual search result see tutorial Sequence Database Search using a MALDI Mass Fingerprint , chapter L.2.2 MASCOT Search Results). The search result is evaluated in real time by a fuzzy logic algorithm, which can be selected in the header of the tab. Result of this evaluation of the search result is indicated by color-coding, legend can be found in the Scout MTP tab.

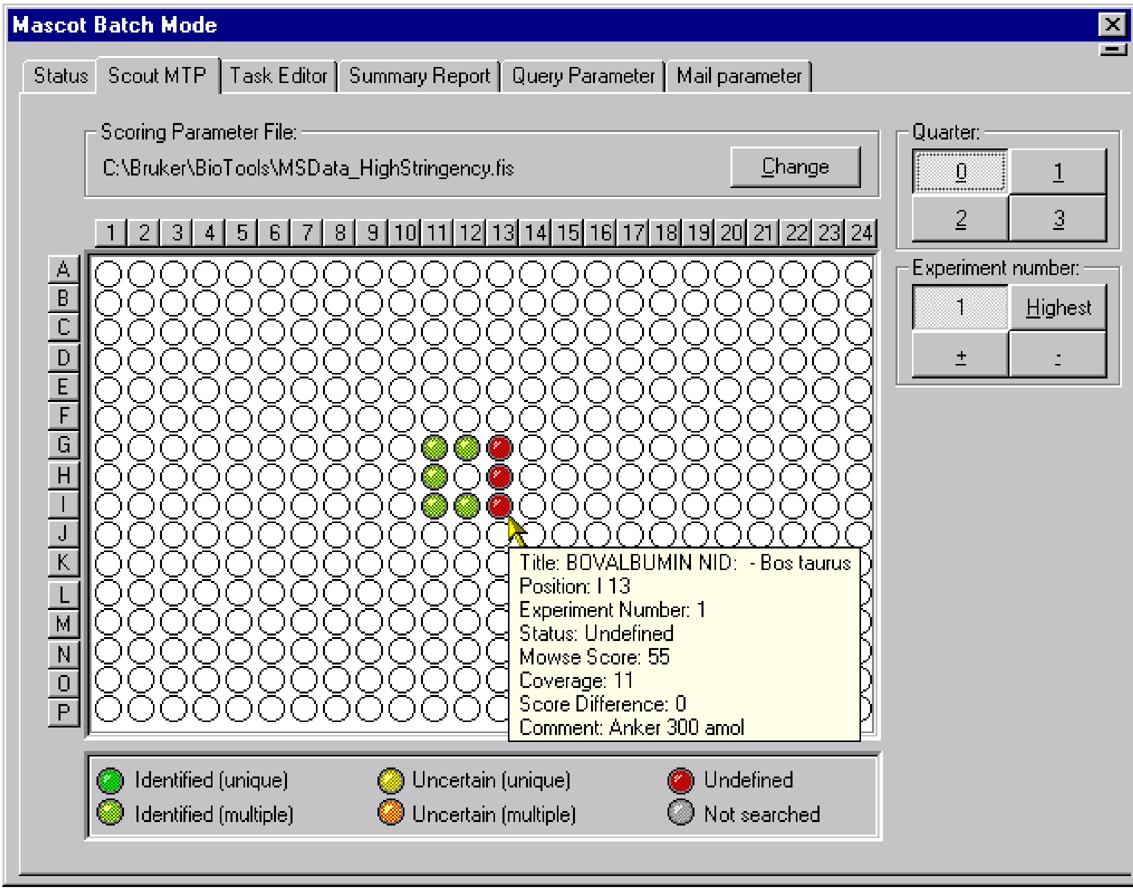


Figure M-16, The Scout MTP tab - results

- After the last search you can switch to the MTP view for an overview of all results in the task. The mouse cursor over the wells causes the "tool tips" to appear after 2 sec with details about the respective sample.
- A left mouse click opens the MASCOT Query results and loads the respective spectrum into BioTools. Get Result(s) loads the matching sequence and peaks into the tree view and allows manual refinement as described under Sequence Database Search using a MALDI Mass Fingerprint.

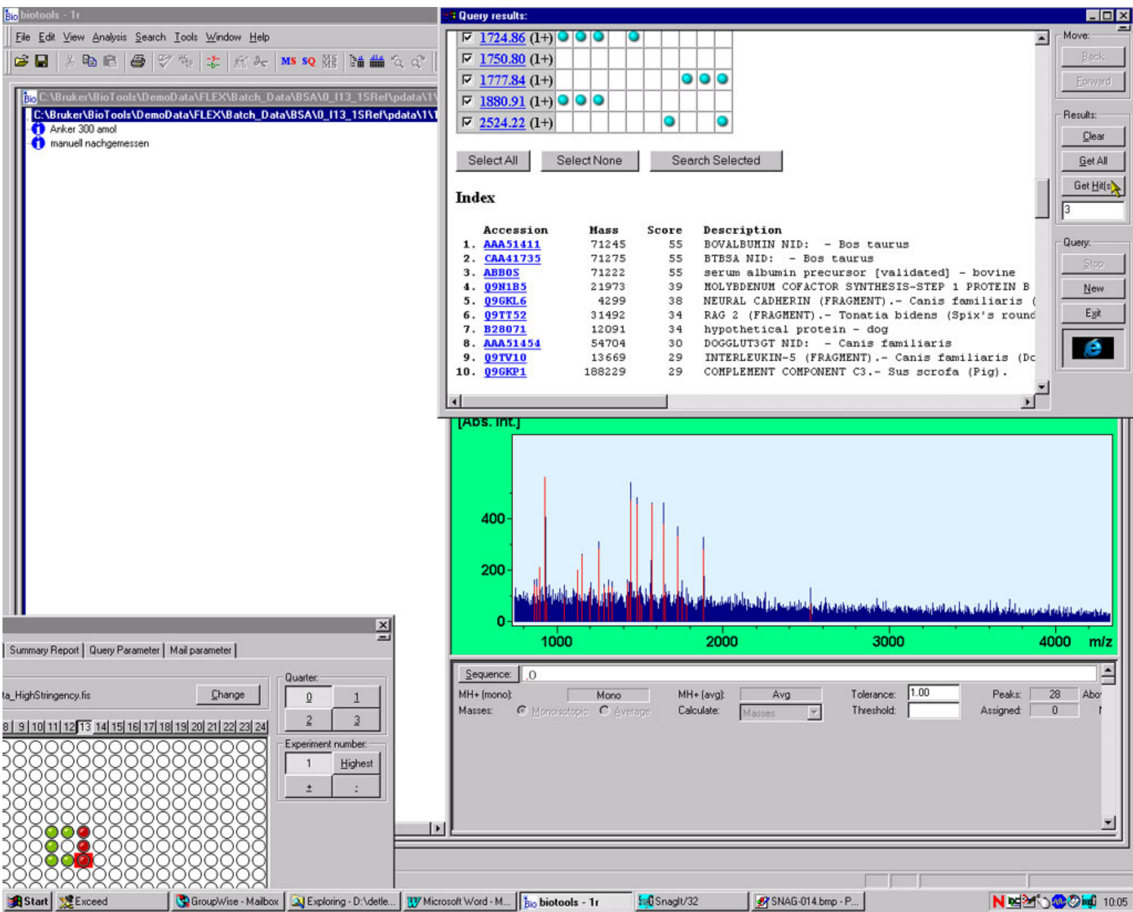


Figure M-17, Mascot Query results

Note: *This Query window as well as the Batch Mode dialog, e.g., with the Scout MTP tab activated, can be shrunk in order not to loose window space while working on a particular data set.*

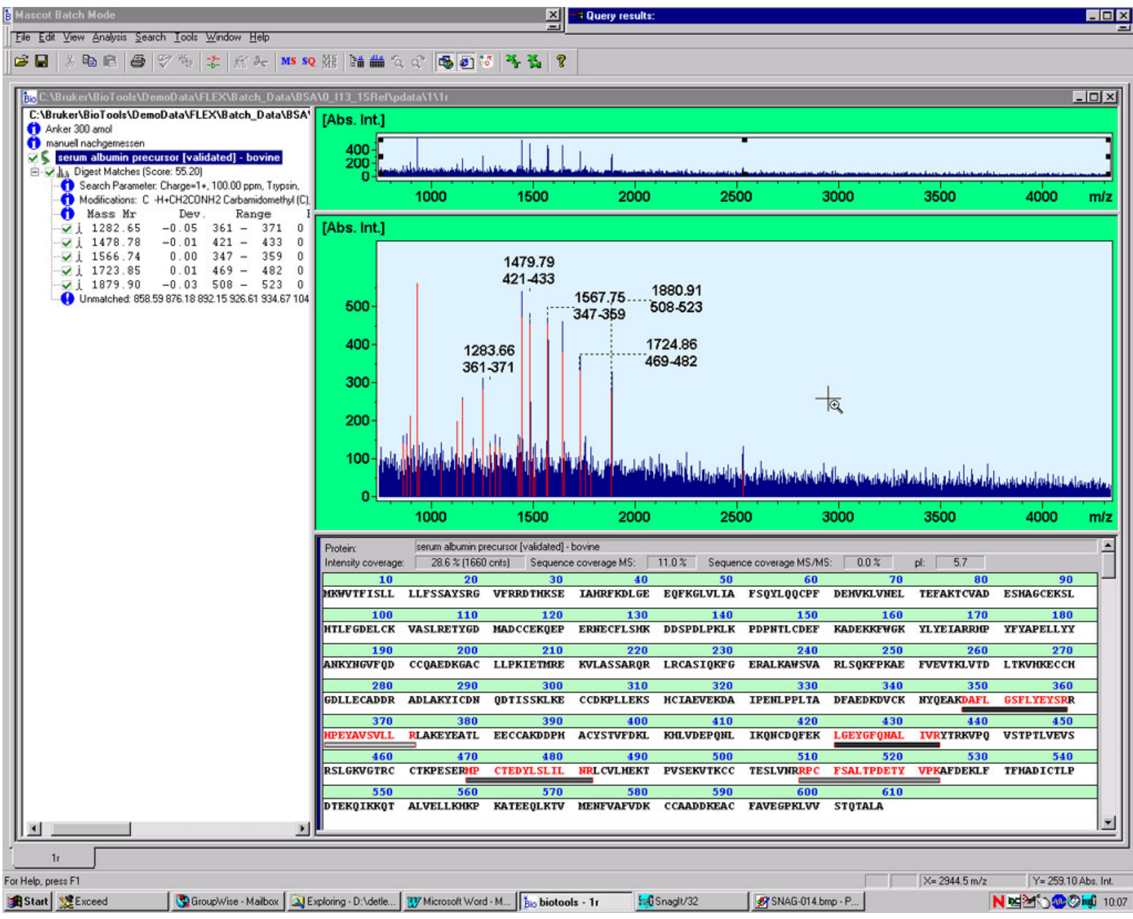


Figure M-18, Results added to spectrum

- Refinements may also include another fingerprint search, which can manually be started as usually, in addition to the shrunk Query results viewer of the batch interface.

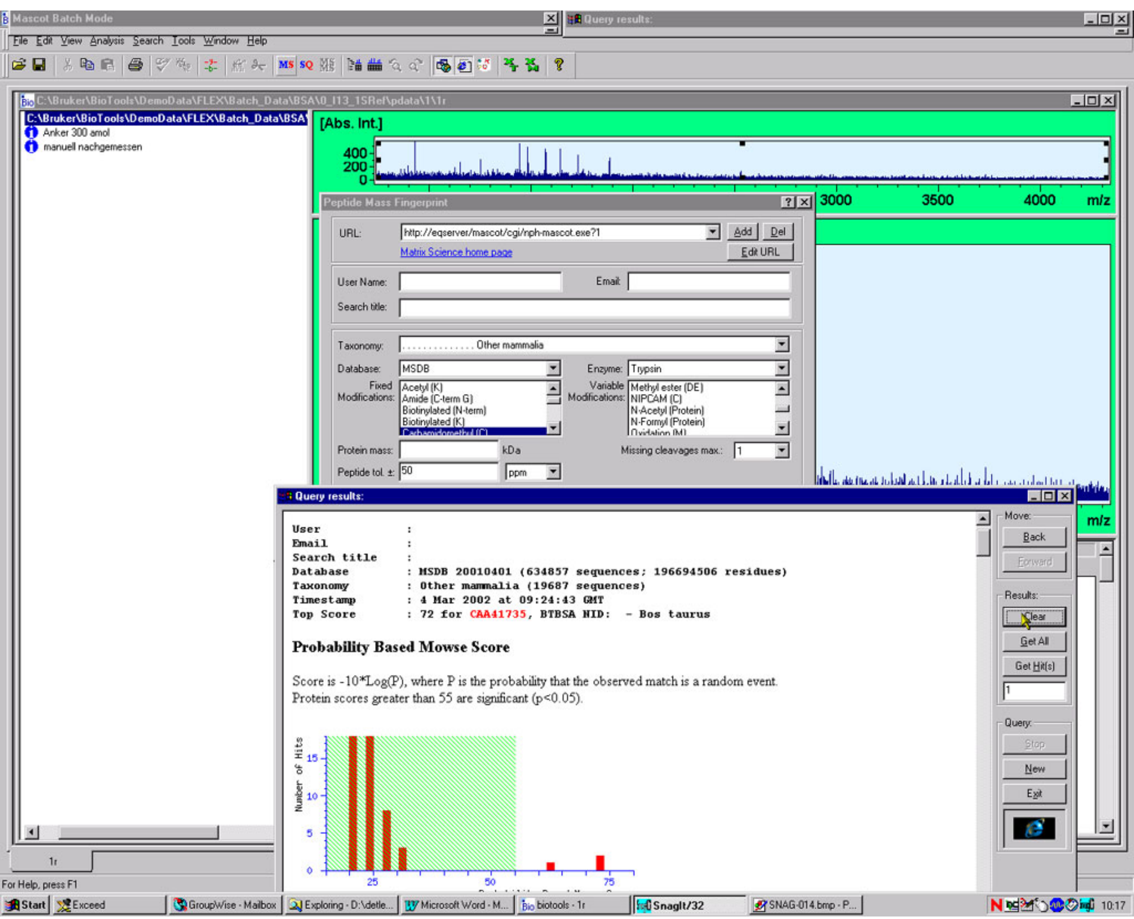


Figure M-19, Refinement of results in spectrum

- Import now the result for the search shown in Figure M-19 with the Get Result(s) button and expand the Batch mode window to display the Scout MTP tab. The batch run result is still shown in the tool tip.

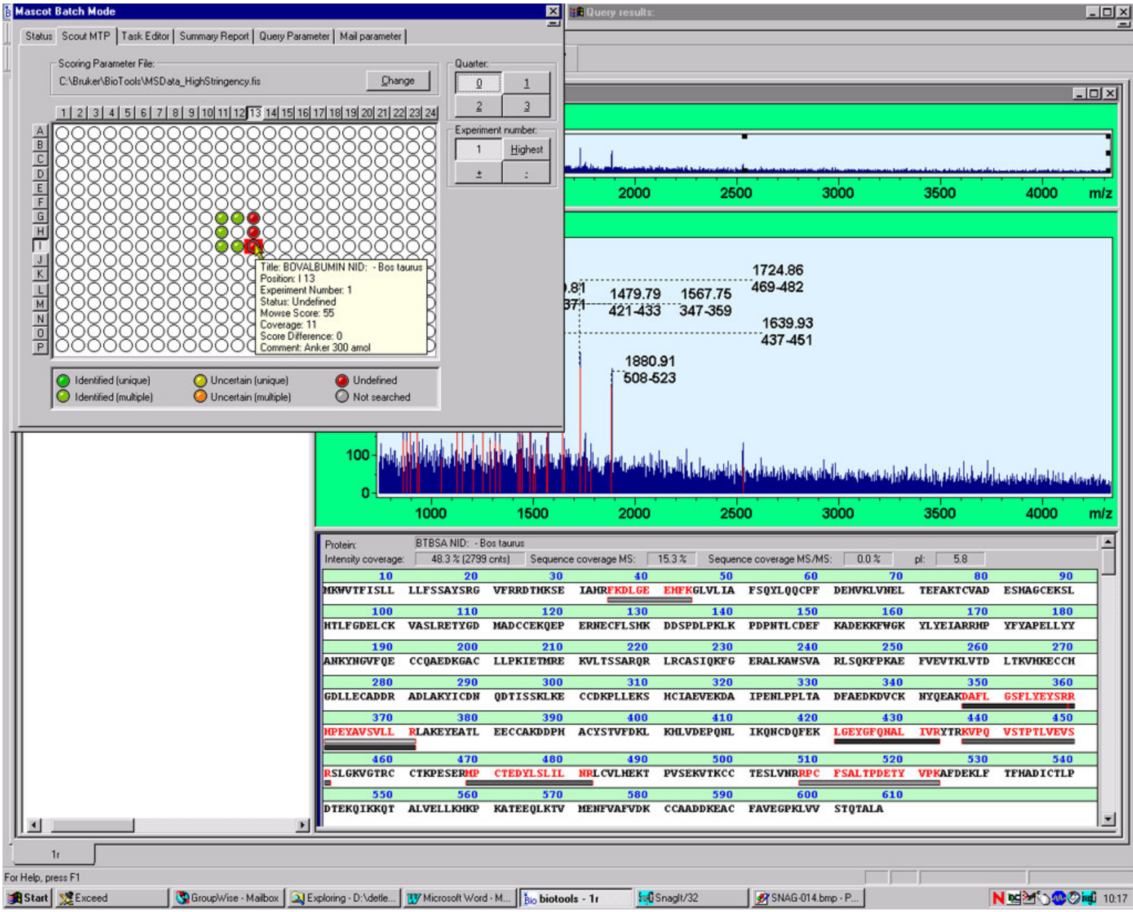


Figure M-20, Batch run result

- Click the right mouse button to Accept the manual result while the mouse is over the respective sample well.

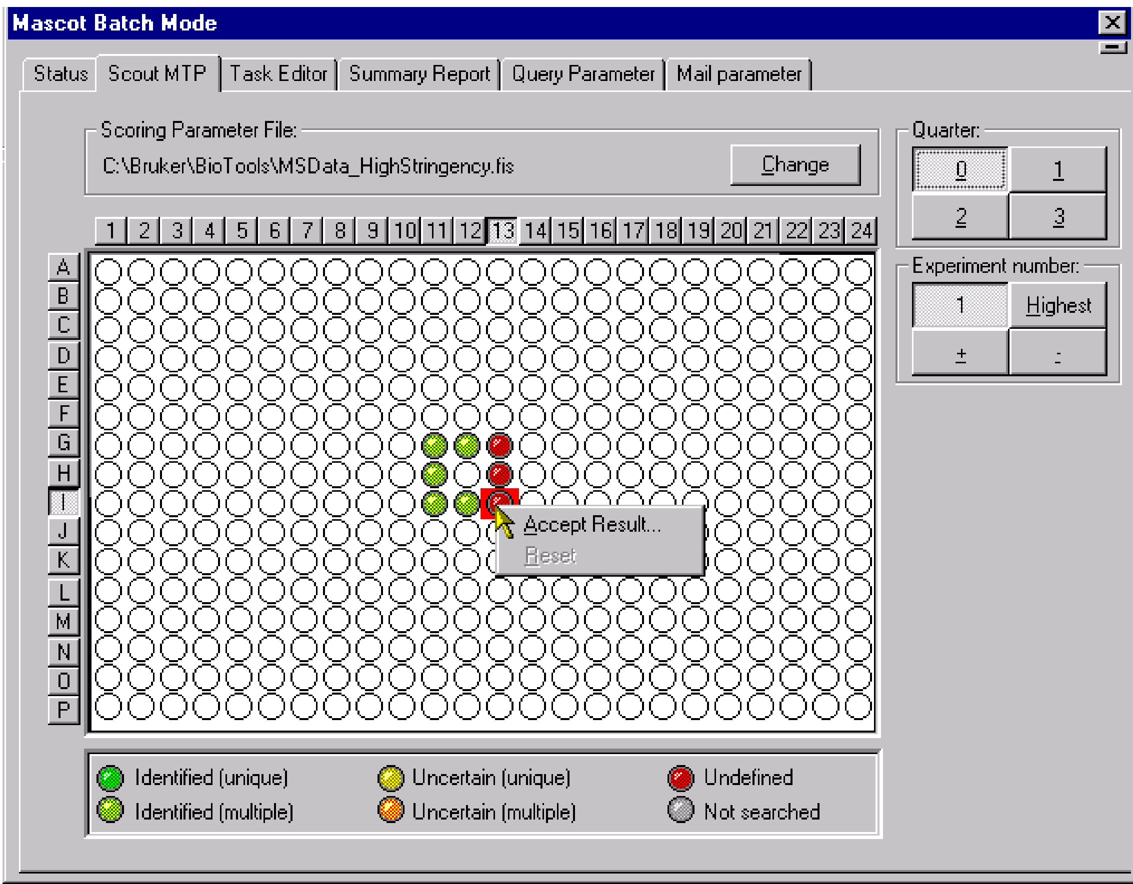


Figure M-21, Accepting the manual Query result

- The list pops up (Figure M-22), which resulted from the manual refinement of the search and you can select the proper sequence, which is now accepted by you. This is shown by a checkmark in the Scout MTP viewer (Figure M-23).

Accept Query Results

Hit	Score	Accession	Mass	Description
1	72.9	CAA41735	71274.64	BTBSA NID: - Bos taurus
2	72.9	AAA51411	71244.63	BOVALBUMIN NID: - Bos taurus
3	60.9	ABBOS	71221.61	serum albumin precursor [validated] - bovine
4	31.3	CAA11411	53505.14	MVLIPOLIP NID: - Mustela vison
5	29.7	Q9MWN3	9613.48	MHC CLASS II DR ANTIGEN BETA SUBUNIT (FR
6	29.6	KIRBA	21738.41	adenylate kinase (EC 2.7.4.3) - rabbit
7	29.4	Q9TTZ1	24306.22	L-GULONO-GAMMA-LACTONE OXIDASE (FRAGN
8	28.1	JC4878	8483.40	alpha endosulfine - bovine
9	27.6	O77852	10869.28	MHC CLASS II DLA-DQ BETA CHAIN B1 DOMAIN
10	27.6	O77854	10855.23	MHC CLASS II DLA-DQ BETA CHAIN B1 DOMAIN
11	27.6	Q9GIW2	10913.27	MHC CLASS II ANTIGEN BETA CHAIN (FRAGMEI
12	26.9	AAA50984	7436.84	LAMBDA-IMMUNOGLOBULIN PRECURSOR (FRA
13	26.3	TMRBB	32930.62	tropomyosin beta chain
14	26.3	O46780	33372.45	MHC CLASS I HEAVY CHAIN (FRAGMENT).- Bos
15	25.9	B28071	12090.94	hypothetical protein - dog
16	25.7	Q9XSU9	8441.66	RIBOSOMAL PROTEIN S14 (FRAGMENT).- Canis
17	25.4	BAA14169	35902.91	BOVPLDH5 NID: - Bos taurus
18	25.2	O19382	9529.58	MHC CLASS II DRB FIRST EXTRACELLULAR DC
19	25.2	O19386	9457.52	MHC CLASS II DRB FIRST EXTRACELLULAR DC
20	25.2	O19435	9515.56	MHC CLASS II DRB (FRAGMENT).- Felis silvestris
21	25.2	O19385	9487.53	MHC CLASS II DRB FIRST EXTRACELLULAR DC
22	25.1	LDHM_BOVIN	36785.43	L-LACTATE DEHYDROGENASE M CHAIN (EC 1.1
23	25.1	Q9TT52	31491.96	RAG 2 (FRAGMENT).- Tonatia bidens (Spix's round
24	25.0	JQ2222	36916.47	L-lactate dehydrogenase (EC 1.1.1.27) chain M - ho

Accept Cancel

Figure M-22, Accept Query result list

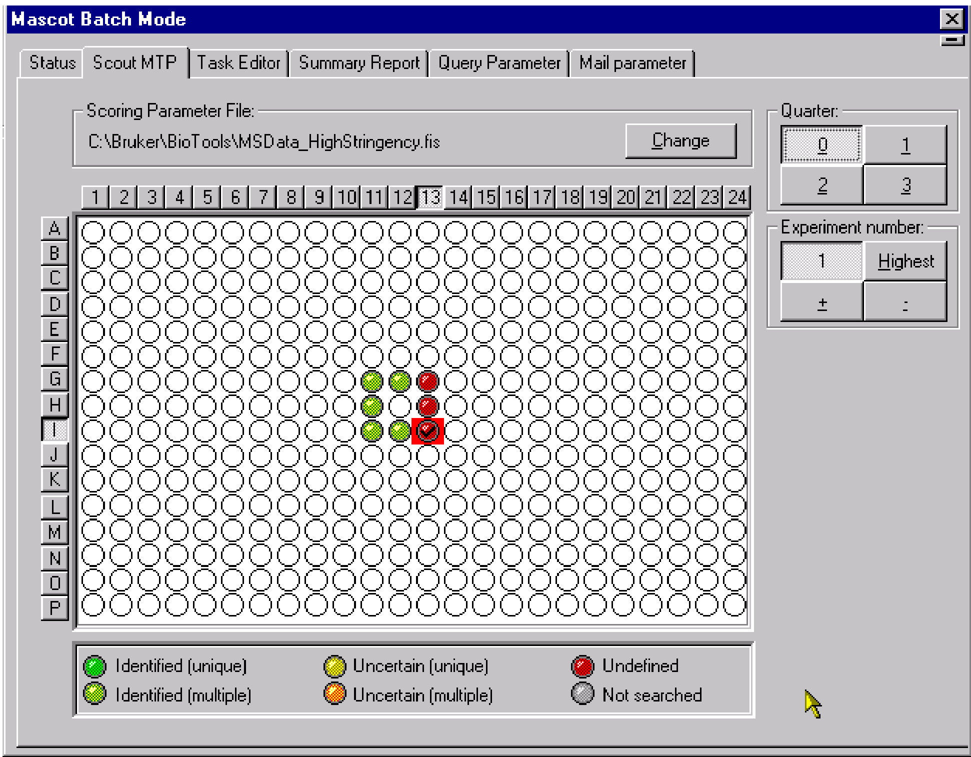
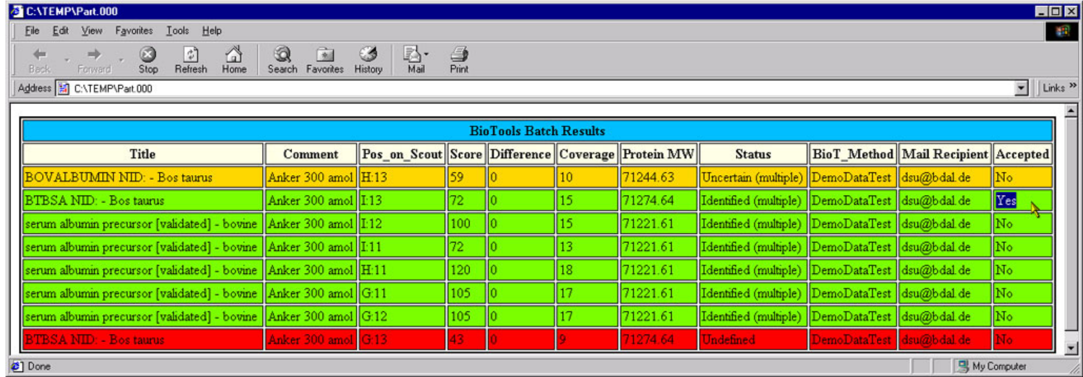


Figure M-23, Mark of accepted results

- Switch to the Summary Report tab and either Export the results to EXCEL or Send it by email to a recipient. The transmitted html document keeps the color-coding and also displays the manual acceptance.
- For a printout of the batch search results press the Print button.



BioTools Batch Results

Title	Comment	Pos_on_Scout	Score	Difference	Coverage	Protein MW	Status	BioT_Method	Mail Recipient	Accepted
BOVALBUMIN NID: - Bos taurus	Anker 300 amol	H13	59	0	10	71244.63	Uncertain (multiple)	DemoDataTest	dsu@bdal.de	No
BTBSA NID: - Bos taurus	Anker 300 amol	I13	72	0	15	71274.64	Identified (multiple)	DemoDataTest	dsu@bdal.de	Yes
serum albumin precursor [validated] - bovine	Anker 300 amol	I12	100	0	15	71221.61	Identified (multiple)	DemoDataTest	dsu@bdal.de	No
serum albumin precursor [validated] - bovine	Anker 300 amol	I11	72	0	13	71221.61	Identified (multiple)	DemoDataTest	dsu@bdal.de	No
serum albumin precursor [validated] - bovine	Anker 300 amol	H11	120	0	18	71221.61	Identified (multiple)	DemoDataTest	dsu@bdal.de	No
serum albumin precursor [validated] - bovine	Anker 300 amol	G11	105	0	17	71221.61	Identified (multiple)	DemoDataTest	dsu@bdal.de	No
serum albumin precursor [validated] - bovine	Anker 300 amol	G12	105	0	17	71221.61	Identified (multiple)	DemoDataTest	dsu@bdal.de	No
BTBSA NID: - Bos taurus	Anker 300 amol	G13	43	0	9	71274.64	Undefined	DemoDataTest	dsu@bdal.de	No

Figure M-24, Exporting the results as html-document

M.5 Batch Processing of AutoLIFT data in BioTools 2.2

On the ultraflex TOF/TOF it is easy to generate many MS/MS data sets in a short time. Multiple LIFT-TOF/TOF MS spectra can be combined and submitted to MASCOT as a single search providing dramatically increased search specificity. Here we use the demo data to show how to do it.

M.5.1. Import Spreadsheets

Click the button to open the batch mode, select the Task editor tab, click the *from AutoX* button and browse to the ..\DemoData\FLEX\BatchData\ autolift_example_10lift.txt file.

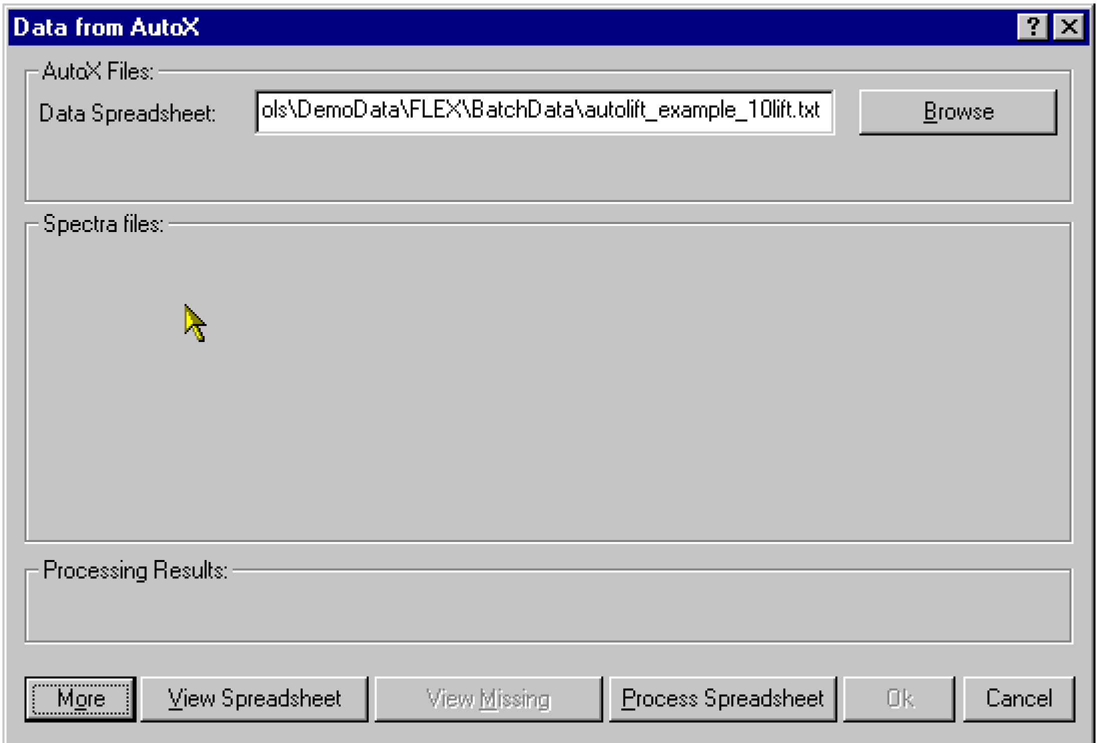


Figure M-25, Select Spreadsheet file in DemoData

The data set containing 8 BSA tryptic digests and 80 LIFT-TOF/TOF spectra was acquired with flexControl 1.2. This must be specified prior to processing of the

spreadsheet, since the data structure, in which MS and MS/MS data are organized, changed with respect to flexControl 1.1.

In Figure M-25 click the *More* button and select all options according to Figure M-26.

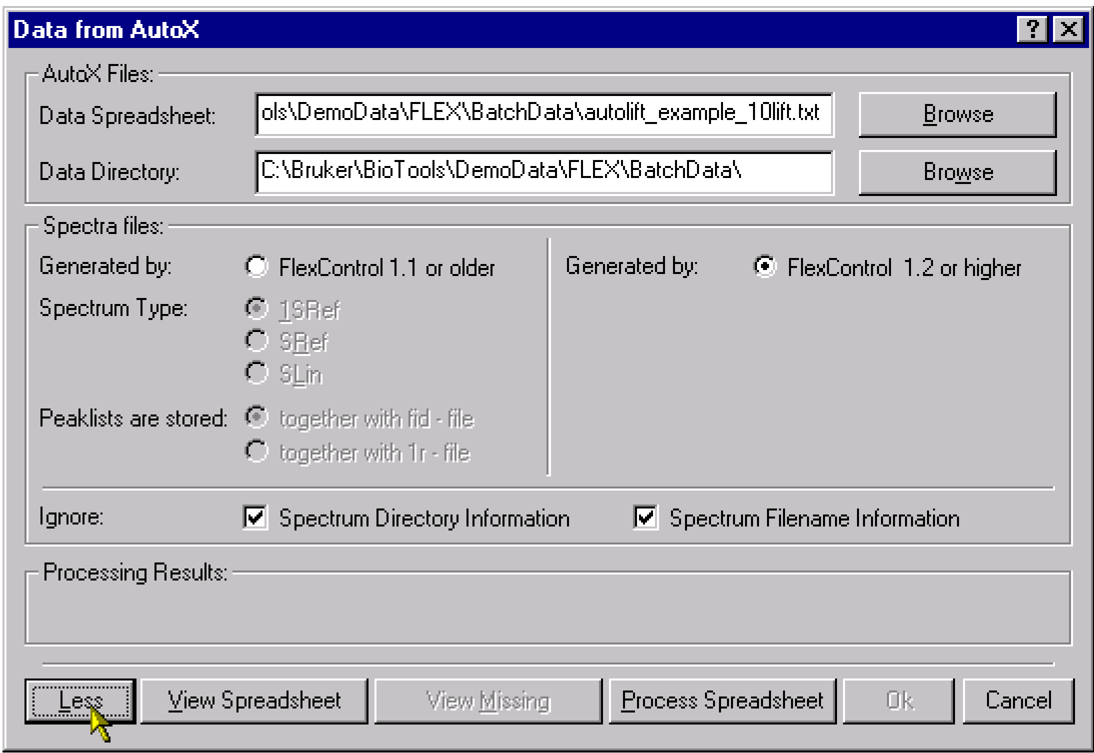


Figure M-26, Select the Options available for flexControl 1.2

After clicking Process *Spreadsheet*, for each position 2 files are imported into the task list: 1) the original digest spectrum, and 2) a mgf file comprising of the peaklists of all the MS/MS data sets acquired from this digest and including the full fingerprint mass list. The Processing Results message should read as "17 Peaklists found, 0 missing", I.e., 8 protein mass fingerprints (PMF), 8 combined data sets and 1 calibrant.

Click *OK* to return to the Task Editor (Figure M-27).

Remove calibrant data from search list. The first sample in the list is the calibrant. It is selected (grayed) so it can be simply removed from the search list with the keyboards *Delete* button. At this point ALL calibrant spectra would be automatically selected for simple removal.

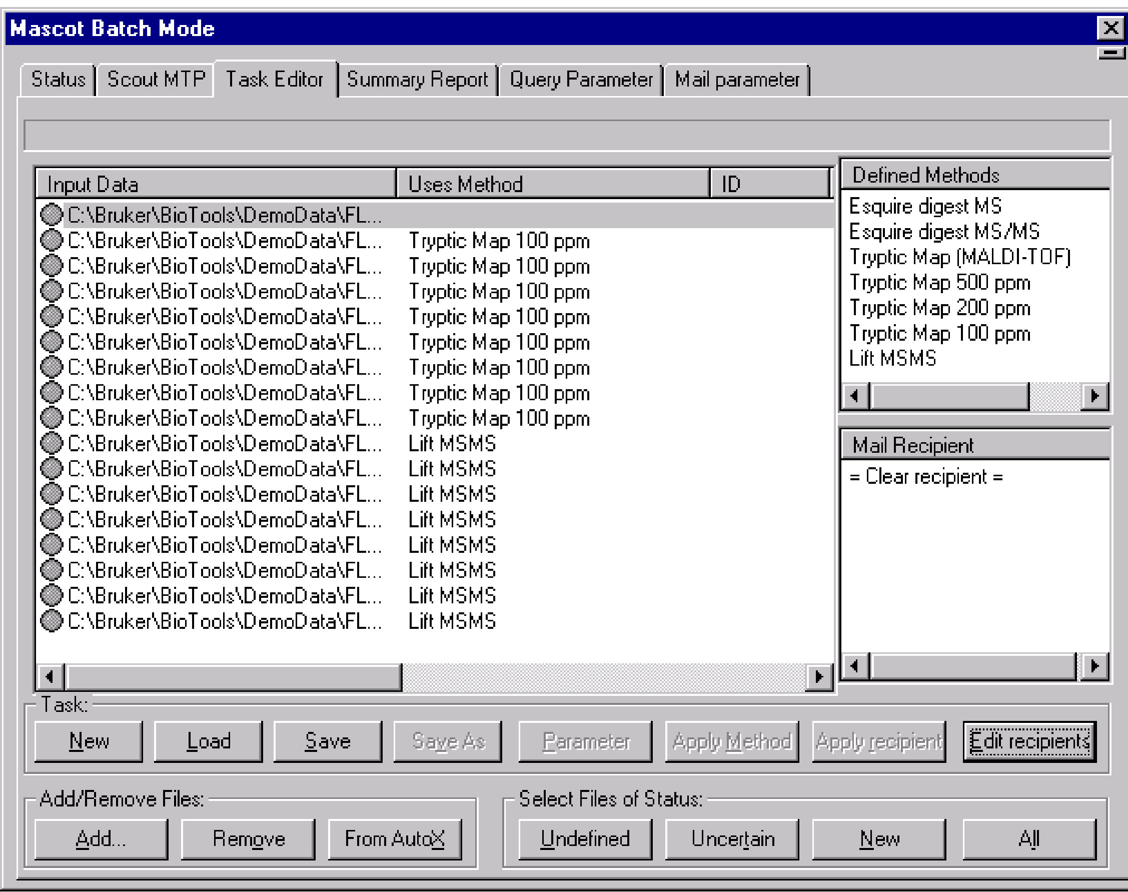


Figure M-27, Ready to delete the calibrant´s peak list from the task

After removal of the calibrants the task must be saved, default using the DemoData is *autolift_example_10lift.tas*.

M.5.2. Define Query Parameters

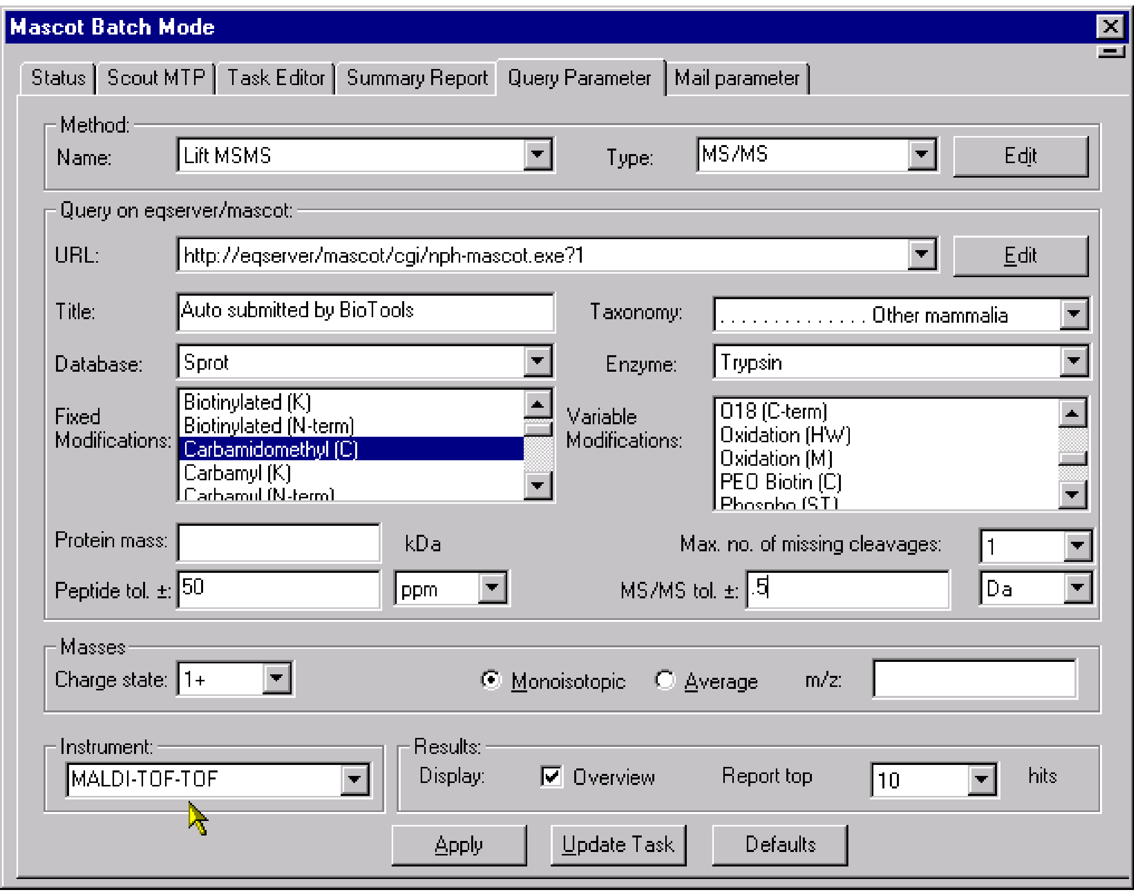


Figure M-28, Select local server URL and appropriate query parameters, including TOF/TOF as the instrument (MASCOT 1.8 and higher)

In the Query Parameter tab verify the correctness of all settings for the defined search methods: *Tryptic Map 100 ppm* and *Lift MSMS*. Also you must select MASCOT on an **INTRANET** server. Good results are obtained with the settings shown in Figure M-28.

Click the *Apply* and *Update Task* buttons to store the settings and make them directly available to the current task (Figure M-29).

M.5.3. Batch Search

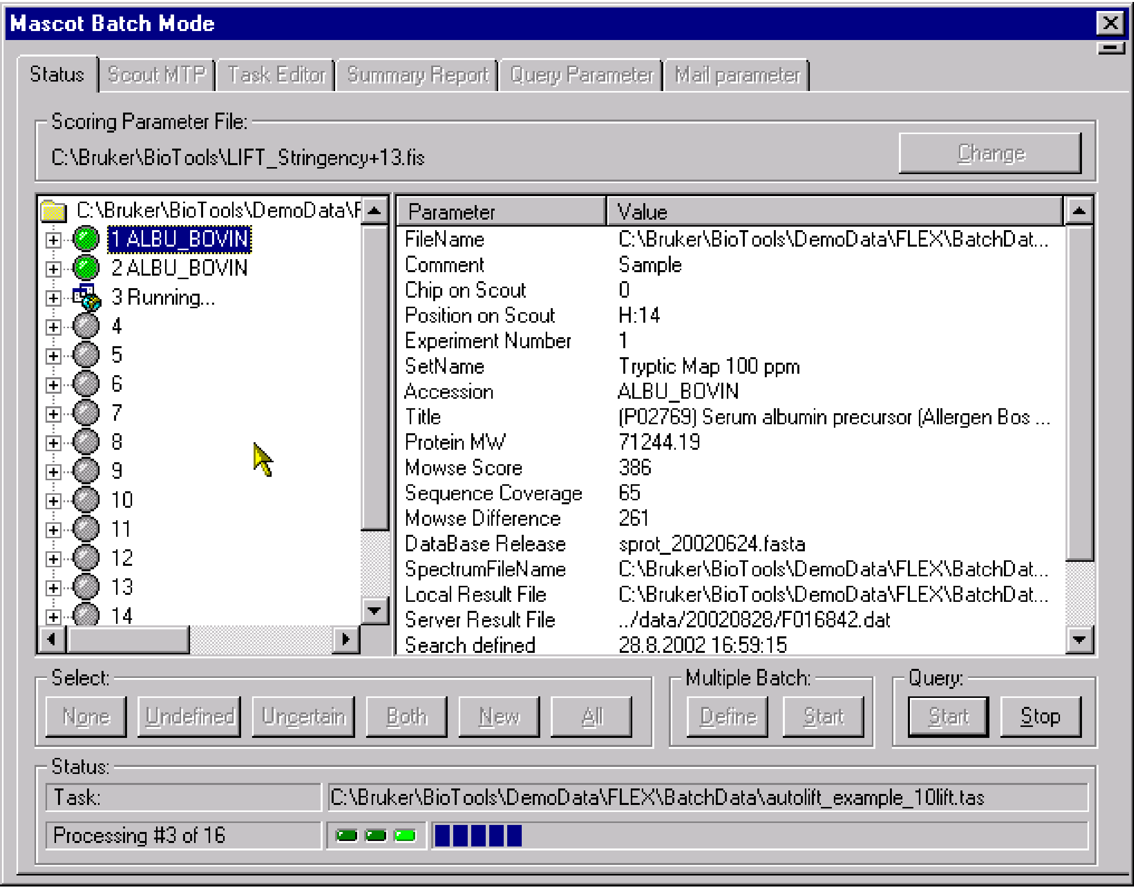


Figure M-29, Start batch processing in the Status tab

Select the Status tab and *Start* the searches.

After the search is completed, switch to the *Scout MTP* tab, which allows you simple access to all of the searched data (Figure M-30). The color-coding comes from fuzzy logic-based judgment engines. The engines suitable for MS/MS data, in particular LIFT spectra are LIFT_Stringency+13.fis, ..+4.fis and ..-11.fis. Stringency +13 means: if for a protein score the 5 % threshold for safe identification is at 65 (end of the green box in the MASCOT result dialog), this fis-file judgment will provide a save identification at scores higher than 65+13=78. This is a conservative score, which is appropriate for combined LIFT data as in this example, but not for single LIFT spectra.

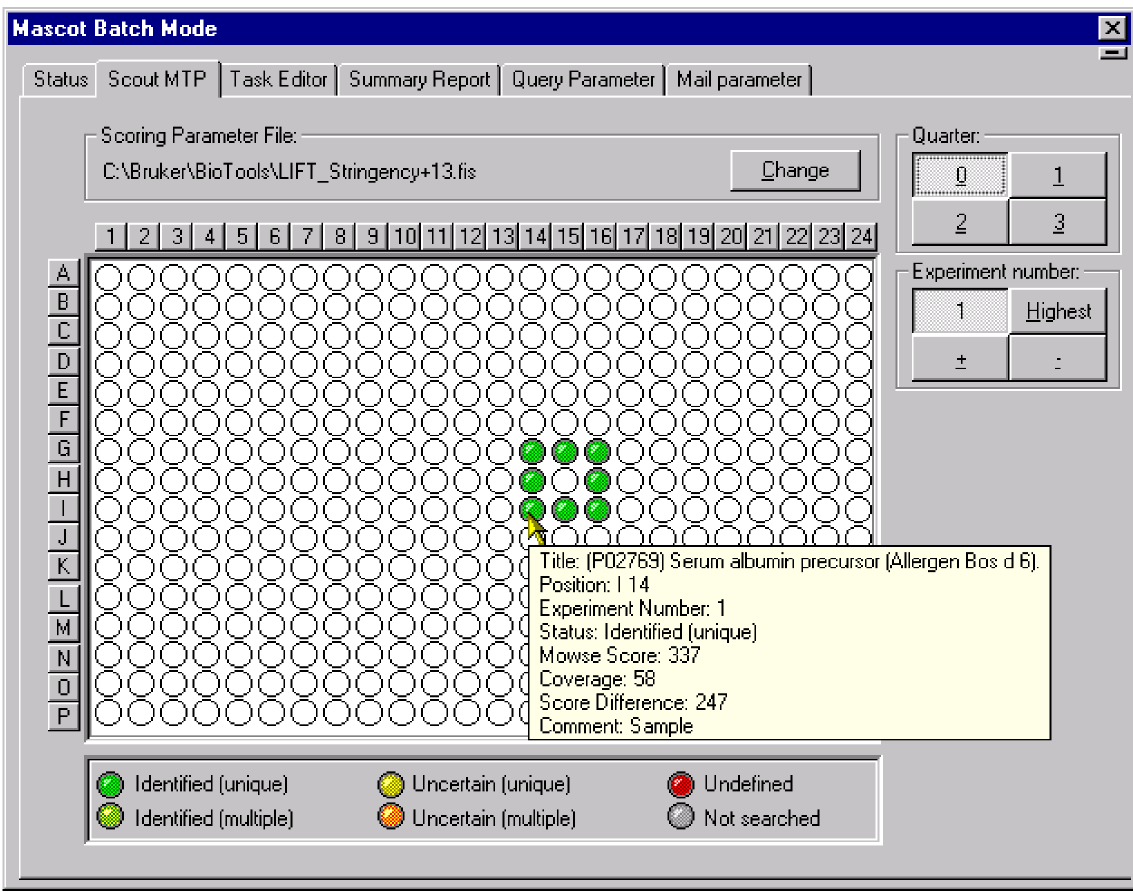


Figure M-30, Search results of the PMFs have *Experiment number* 1

The results of the search for the combined data sets (1 PMF + 10 MS/MS spectra) have Experiment number 2, which is selected by either one click onto the **+** button or by clicking the *Highest* button on the right side of the *Scout MTP* tab. Search results can be directly obtained from the tool tip dialog, which pops up if the mouse cursor is above an MTP spot (Figure M-30 and Figure M-31).

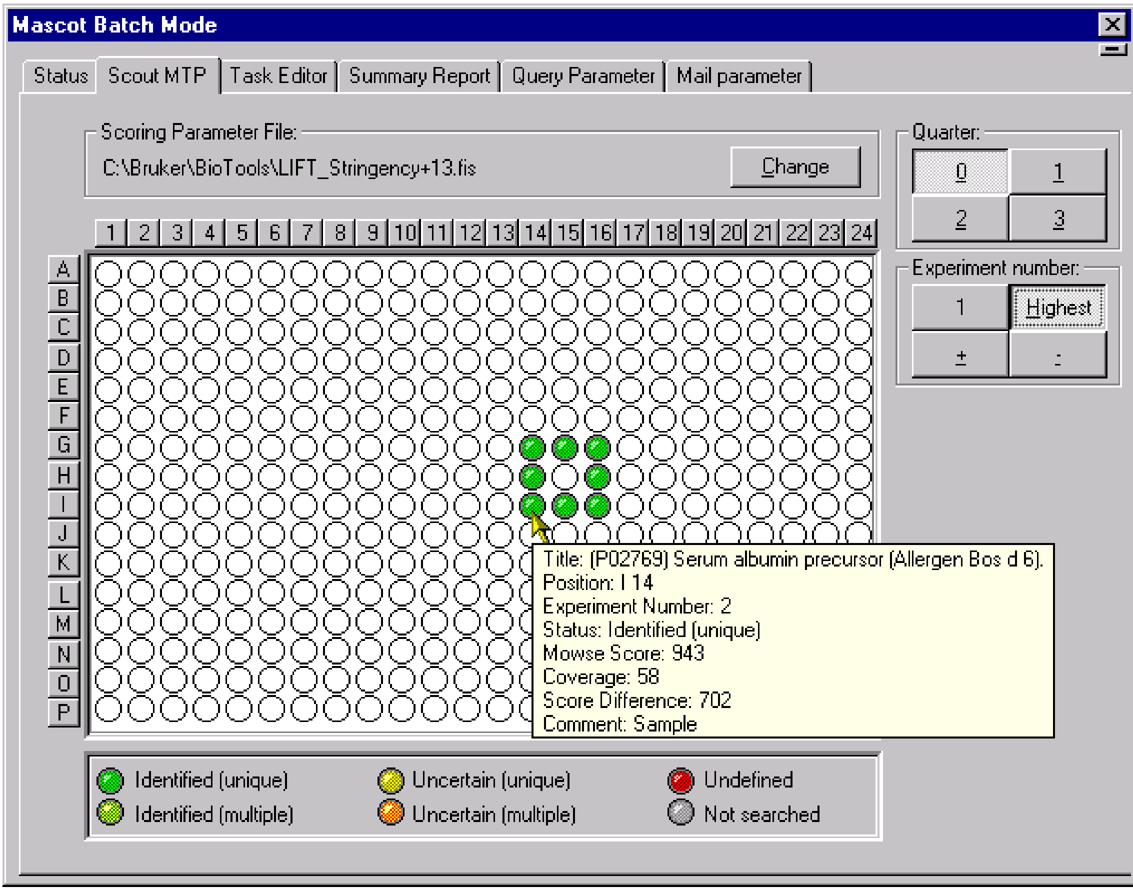


Figure M-31, Search results of the combined data are displayed if *Highest* is selected

A left mouse button click onto such a spot opens the MASCOT search results and a histogram representation of the PMF.

M.5.4. Importing Search Results from Combined Datasets into BioTools

Before you press the Get Hit(s) button make sure the result import mode "Accept all new search results" for combined data sets is the one you need. The mode is selected on the context-sensitive menu (right mouse button) in the tree view, which displays the peak lists (Figure M-32).

There are two options, which differ in their way to select the individual peaks in the imported peak list:

a) Accept all new search results **selected**: is suitable for data searched with *trypsin*. All peaks are check-marked in the list and their labels are displayed in the spectrum and the sequence.

b) Accept all new search results **deselected**: is suitable for complex data depending on MS/MS, such as LC-MS/MS data or results obtained from searches with *no enzyme*.
Only those peaks are check-marked, which were positively identified due to their MS/MS spectrum. A *Delete unchecked sublevels* operation would eliminate all peaks without MS/MS confirmation from the mass list

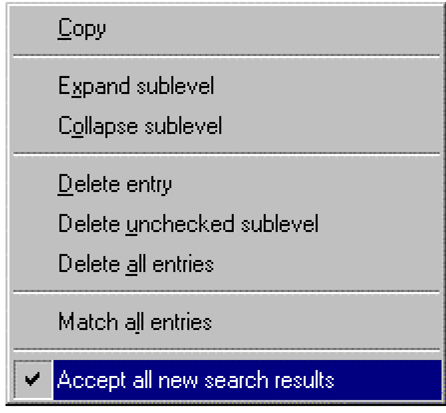


Figure M-32, Select the combined data import mode

After the import, select the top level in the tree view and right mouse click delete unchecked sublevels, in order to remove all peptide mass entries, which are not checked.

Note ***Currently these are all the masses, which were identified by MS/MS, since they occur in the result list twice due to a bug in MASCOT 1.8.***

Result is an annotated histogram spectrum containing all selected peptides. Peptides with a matched MS/MS spectrum contain two rows of red bricks. The upper row indicates matching b-ions the lower row matching y-ions (Figure M-33). From here the original MS or MS/MS data cannot be directly accessed, rather these spectra need to be opened manually if you want to compare them the peaklist.

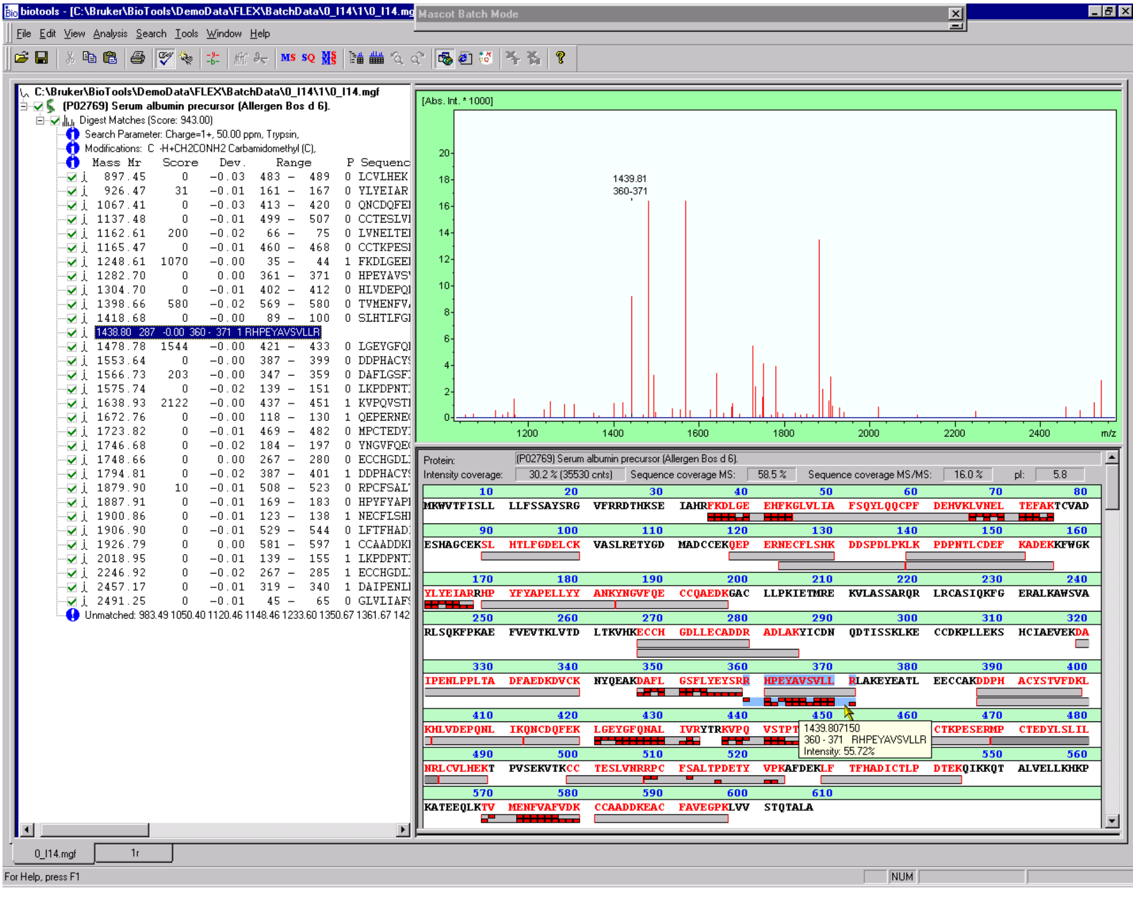


Figure M-33, Result of a combined search, displaying the full PMF information and indicating detected b- and y-ions as red bricks in peptides, which were fragmented and assigned to the sequence

At this point, the respective MS/MS spectra must be manually loaded, if one needs to inspect them directly, there is no link to the original data.

Note ***All MASCOT Scores used and displayed in BioTools are from the Protein Summary Report. However, the correct score to judge MS/MS data including the combined searches are the scores from the Peptide Summary Report. Future releases of MASCOT will account for this.***

M.6 CAF chemistry in BioTools

Chemically assisted fragmentation (CAF) was published by Keough – see also http://www.bdal.com/applications/shared/mt53nterm.pdf and recently commercialized by Amersham Bioscience. The CAF chemistry modifies peptides in 2 steps: 1. All K side chains (ε-amino group) are modified turning Lysine into homoarginine (+42 Da); 2. All free amino groups – now only the N-terminus of each peptide should be left for further modification – are derivatized into 3-sulfonic acid-propionamide (+136 Da) by the CAF-reagent. It´s intention is so simplify MS/MS data since the acidification of the N-terminus tends to suppress N-terminal ion series and to promote the abundant appearance of C-terminal y-ions.

In real life side reaction may interfere with the data analysis following the scheme described above. We recognize 1. incomplete homoarginine formation, 2. CAF-reagent modification of unreacted Lysines

If you use this chemistry on the Bruker mass spectrometers using MASCOT and BioTools , please follow these hints:

M.6.1. MASCOT PMF Searches

On the local MASCOT server make sure the following entries are available in the modifications definition file:

```
Title:Homoarginine (K)
Residues:K 170.11696 170.21468
*
Title:CAF-reagent (N-term)
Nterm:136.991 137.136
*
Title:CAF-reagent (K)
Residues:K 264.07796 264.30268
*
```

These modifications are also contained in BrukerDefault.mod, which must be selected in the SequenceEditor as that affects the defaults used in the BioTools main viewer (spectrum, tree view and sequence).

M.6.2. Search for Masses

In Sequence Editor tryptic digests modified by the CAF-chemistry are handled by the enzyme CAF-Trypsin, which combines the trypsin cleavage with chemical modification.

M.6.3. MASCOT MS/MS Searches

Due to incomplete derivatization and side reactions variable modifications need to be allowed for, if immediate ID failed. Homoarginine (K) or CAF-reagent (K) may be used as variable modification.

M.6.4. Properties of LIFT and PSD Spectra

- N-term CAF — MH^+-136 Da very intense peak
- N-term CAF — i_1 ions not observed
- C-terminal homoarginine (K) — y_1: m/z 189
- C-terminal R — y_1: m/z 175
- Homoarginine close to N-term yield — i: m/z 143
- N-term CAF + side chain-CAF — MH^+ -136 Da, MH^+ -2*136
- N-term CAF + side chain-CAF of K[1] — i: 373
- If 189 or 175 are not clearly visible or clean sequence patterns don not show up, it is likely that multiple side chain modifications occurred.

M.6.5. RapiDeNovo

- *CAF-reagent (N-term)* must be selected as N-term modification in the expanded view of the if neutral loss of 136 Da from the MH^+ is observed.
- *Homoarginine (K)* must be selected as fixed modification and C-terminal K can be specified as sequence hint, if m/z 189 is observed.

Note ***If a C-terminal sequence hint (e.g., K) is given, for that amino acid residue optional modifications must not be specified, they will be ignored!***

M.6.6. References

- Keough T, Youngquist RS, Lacey MP
  A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry
  Proc Natl Acad Sci U S A. 1999 Jun; 22;96(13):7131-6.
- Keough T, Lacey MP, Youngquist RS.
  Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry
  Rapid Commun Mass Spectrom. 2000; 14(24):2348-56.
- Keough T, Lacey MP, Fieno AM, Grant RA, Sun Y, Bauer MD, Begley KB.
  Tandem mass spectrometry methods for definitive protein identification in proteomics research
  Electrophoresis. 2000 Jun; 21(11):2252-65.
- Bauer MD, Sun Y, Keough T, Lacey MP. Related Articles, Links
  Sequencing of sulfonic acid derivatized peptides by electrospray mass spectrometry
  Rapid Commun Mass Spectrom. 2000; 14(10):924-9.

I In Source Decay (ISD) MALDI-TOF Spectra Analysis

I.1 Introduction I-1
I.2 Sample Preparation and Acquisition Protocol I-2
I.2.1. Preconditions I-2
I.2.2. Material I-2
I.2.3. Preparation I-2
I.2.4. Parameter Settings I-2
I.2.5. Acquisition I-2
I.3 Processing in XMASS I-3
I.4 Processing in BioTools I-3
I.4.1. Case 1: The Sequence is known I-3
I.4.2. Case 2: The Sequence is unknown I-5

I.1 Introduction

ISD spectra are used to obtain sequence information directly from intact proteins. ISD-MALDI spectra were introduced by R.S. Brown & J.J. Lennon Anal.Chem. **67** 3990-3999 (*1995*) and reflector ISD by Bruker: D. Suckau and D.S. Cornett Analusis **26 N° 10**, M18-M21 (*1998*). In ISD spectra acquisitions, the molecular ions fragment spontaneously within the ion source and typically c- type fragment ions are formed (and possibly y-). The ions can be observed from the 10-15th N-terminal residue on and can cover large fractions of the whole protein sequence, provided disulfide linkages were reduced and alcylated prior to ISD analysis. In reflector ISD, isotopic resolution and < 0.05 Da mass errors can be expected in the range m/z 1500-3500, which allows to obtain sequence information close to the N-terminus, even in the case of N-terminally blocked proteins not amenable to Edman sequencing. It appears to be a very useful method in the characterization of recombinant proteins

Follow the description of sample preparation and acquisition conditions carefully for good results:

I.2 Sample Preparation and Acquisition Protocol

I.2.1. Preconditions

single purified protein only (**no protein mixture!**).

sample at **10 to 20 pmol/µl** in 0.1 % TFA in water; < 20 % acetonitrile.

I.2.2. Material

Matrix solution 1 (MS1): sinapinic acid (SA) saturated in ethanol.

Matrix solution 2 (MS2): SA saturated in 0.1 % TFA, 30% acetonitrile.

I.2.3. Preparation

Apply 1 µl of MS1 on stainless steel target to create a thin layer of matrix.

Mix sample aliquot 1:1 with MS2 and apply 0.5 to 1 µl onto thin layer and let dry at ambient air.

Prepare standard peptide mix for calibration (1000-4000 m/z) for reflector mode measurements. (Protein mixture for linear ISD with the goal of high sequence coverage up to larger masses.)

I.2.4. Parameter Settings

Choose a parameter setting optimized for peptides using medium delay.

Change digitizer sensitivity to 2 mV and set digital offset to 80 (Channel 1)

Increase reflector detector voltage to 1.8-1.9 kV.

I.2.5. Acquisition

Start analysis by collecting 50 shots each and increase laser power until peptide peaks appear.

To obtain sufficient S/N ratio, several hundreds of shots must be added.

Calibrate with standard peptide mix.

I.3 Processing in XMASS

SNAP the region containing proper isotope-resolved peaks.

Run the *Annotate Mass_Diffs* macro under *Aura_Cmds.*

The sequences obtained here can be used for analysis, alternatively this can be done in BioTools.

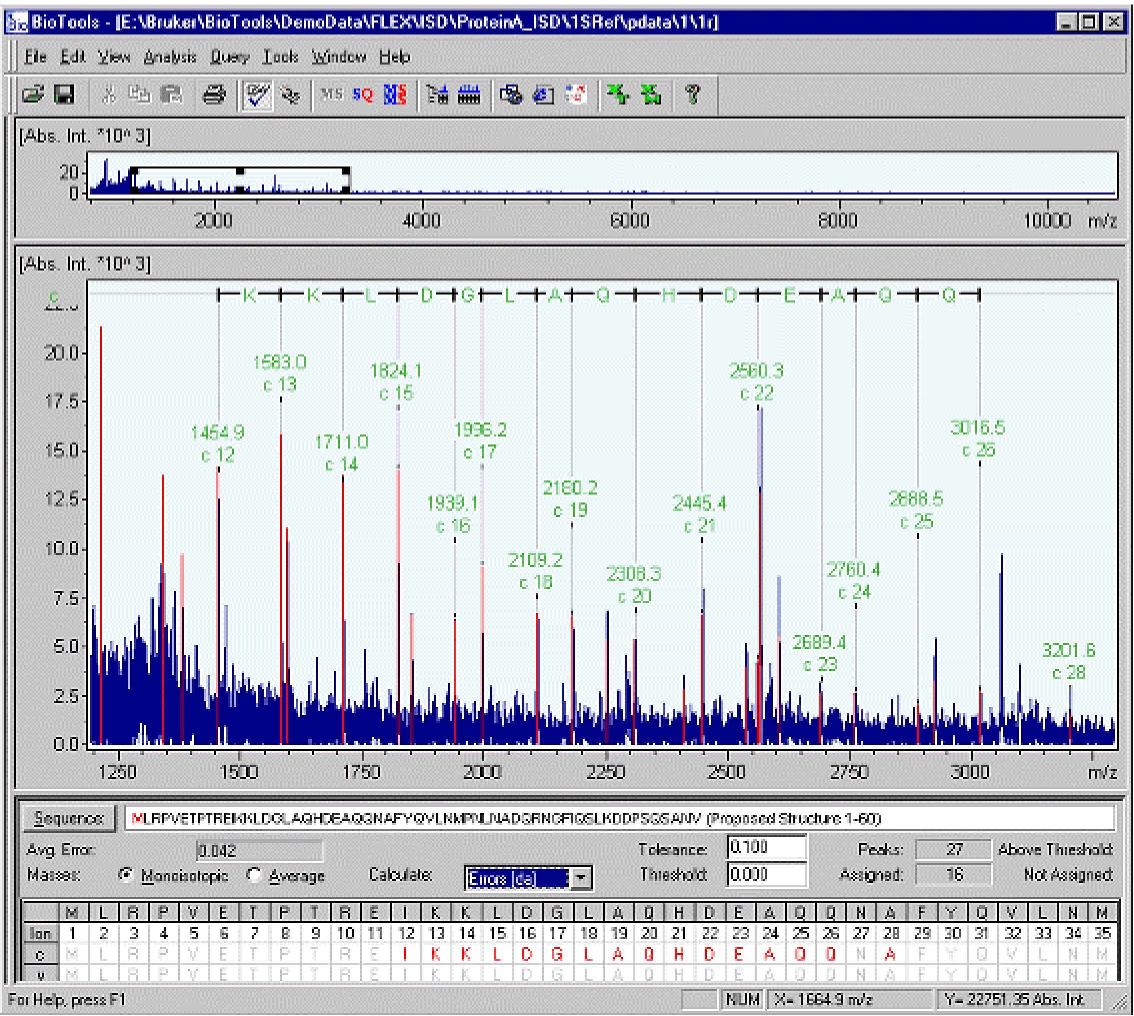
I.4 Processing in BioTools

I.4.1. Case 1: The Sequence is known

(example data set: DemoData/FLEX/ISD/ ProteinA_ISD)

1. Load the spectrum with annotated masses into BioTools, set it to *Check Sequence* mode.
2. Under *Analysis* select *ISD Data* and *Select Ions…*: ISD-TOF.
3. Under Analysis/Annotation/Parameters check the display of each desired fragment ion annotation object.
4. Edit the sequence (including modifications) in SequenceEditor. If you work on masses < 4000 m/z, the N-terminal 50 residues are sufficient. (**Note:** Steps 4 and 5 can be skipped in example dataset, since the sequence is already loaded).
5. Press the BioTools button to transfer the sequence to the spectrum view .
6. Mass Error setting should be set to 0.1 Da, if calibration is OK!

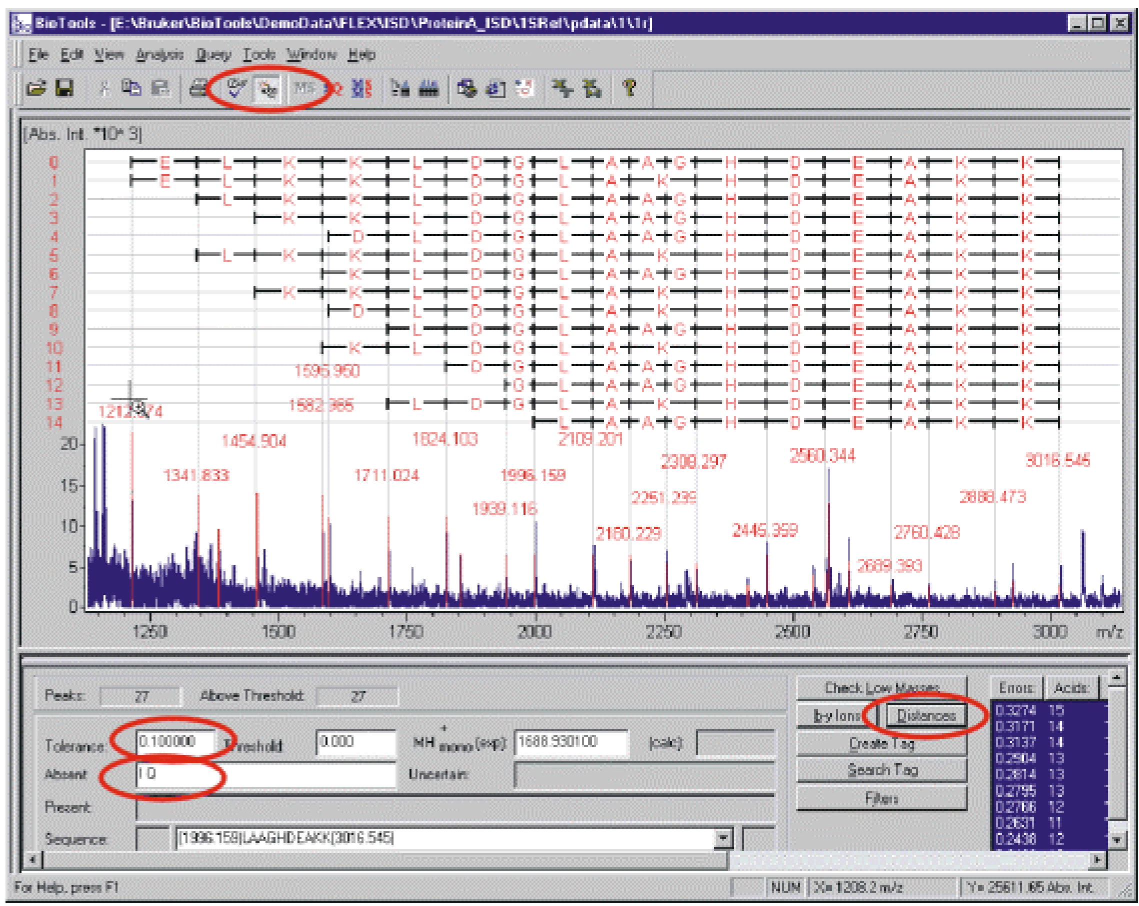
The matching peaks are now displayed in the spectrum.



I.4.2. Case 2: The Sequence is unknown

(example data set: DemoData/FLEX/ISD/ ProteinA_ISD) just do not use the sequence information

1. Load the spectrum with annotated masses into BioTools, set it to *DeNovo Sequencing* mode.
2. Mass Error setting should be set to 0.1 Da, if calibration is OK!
3. Specify **I Q** as residues, which are not considered for calculation.
4. Press the *Distances* button to start DeNovo program based on amino acid residues distances.
5. Evaluate reasonable sequence assignments. Typically, it is easier to obtain correct search results on short sequence tags, e.g., 5-8 residues than on full length sequence tags



Take a sequence or a partial sequence (meaning: write it down and type it into the respective search dialog) and do a sequence-only search at one of the following sites:

PeptideSearch at EMBL:

http://www.mann.embl-heidelberg.de/Services/PeptideSearch/FR_SequenceOnlyFormG4.html

PEPSEA at PROTANA:

http://195.41.108.38/PA_SequenceOnlyForm.html

MS-EDMAN at UCL:

http://falcon.ludwig.ucl.ac.uk/ucsfhtml3.2/msedman.html

Just be aware that in these data you cannot distinguish the isobaric amino acid residues I vs. L and K vs. Q (K and Q can be distinguished by a mass difference of 0.04 Da, so only if your mass accuracy is better than that is it possible). In these search engines you have to specify this fact. Use as search string, e.g., **HDEA(QK)(QK)**.

The correct result for the example data set is SPA2: staphylococcus aureus protein A.

Exercise: The sequence N-terminal to the AQHDEAQQ tag comes from another protein from another species. What is it? Earn experience points for correct solution!

After the full sequence is estimated, follow the procedure under Case 1 to verify the correct structure.

E Using BioTools for esquireSeries Data

E.1 Complete peaklist from a LC-MS/MS run E-2
E.1.1. Protein Identification E-2
E.1.2. Sequence confirmation E-2
E.1.2.1. Loading the mgf-file E-2
E.1.2.2. Loading a protein sequence in the SequenceEditor E-5
E.1.2.3. Starting the Sequence Confirmation E-6
E.1.2.4. The Sequence Confirmation result E-9
E.1.3. Protein Identification with subsequent Sequence Confirmation E-13
E.1.3.1. Loading the mgf-file E-13
E.1.3.2. Starting the Database Search E-16
E.1.3.3. The Database Search result in BioTools E-19
E.1.3.4. Starting the Sequence Confirmation E-22
E.1.3.5. The Sequence Confirmation result E-25
E.2 MGF-files from single LC-MS/MS compounds E-26
E.3 MS/MS-BSC-files E-26
E.3.1. Protein identification E-26
E.3.1.1. Loading the bsc-file E-26
E.3.1.2. Starting the Database Search E-29
E.3.1.3. The Database Search result in BioTools E-32
E.3.2. Refined Search E-35
E.3.2.1. Loading a protein sequence in the SequenceEditor E-35
E.3.2.2. Starting the Refined Search E-36
E.3.2.3. The Refined Search result E-38
E.4 Single profile MS-MGF-file E-39
E.4.1. Protein identification E-39
E.4.1.1. Loading the mgf-file E-39
E.4.1.2. Starting the Database Search E-41
E.4.1.3. The Database Search result in BioTools E-44
E.4.2. Refined Search E-46
E.4.2.1. Starting the Refined Search E-46
E.4.2.2. The Refined Search result E-49

E.1 Complete peaklist from a LC-MS/MS run

For analysing complete LC-MS/MS peaklists BioTools requires an mgf–file (Mascot generic file) from all compounds of the run. A general description of the complete procedure (automatically or interactively) including sample preparation, peptide separation via HyStar 2.2, data acquisition via esquire 5.0, data analysis and export via DataAnalysis 3.0 is given in "Protein identification using esquire3000 + (LC-MS/MS data)". ALP_H_500f_LCMSMS.mgf is the example mgf-file used to show the analysis tools provided by BioTools.

E.1.1. Protein Identification

Please refer to the tutorial "Protein Identification of LC-MS/MS Data".

E.1.2. Sequence confirmation

When opening the LC-MS/MS peaklist (*.mgf) in BioTools, the MS data are combined in one histogram spectrum. The complete LC-MS/MS dataset is used for aligning the experimental data to the protein sequence that has been loaded in the SequenceEditor. All conceivable modifications can be included in the search. The result includes a list of matched peptides which are within the chosen mass deviation, and it shows which part of the protein sequence could be confirmed on MS- and on MS/MS-level.

E.1.2.1. Loading the mgf-file

Use **File | Open** or **File | Find** to open the mgf-file.

When the location of the mgf-file is known choose **File | OpenSpectrum**, select the data type (which is mgf) and the location of the file (Figure 1). Highlight the file with the left mouse button. For loading multiple files use the CONTROL or SHIFT button and the left mouse button. Then **Open** the file.

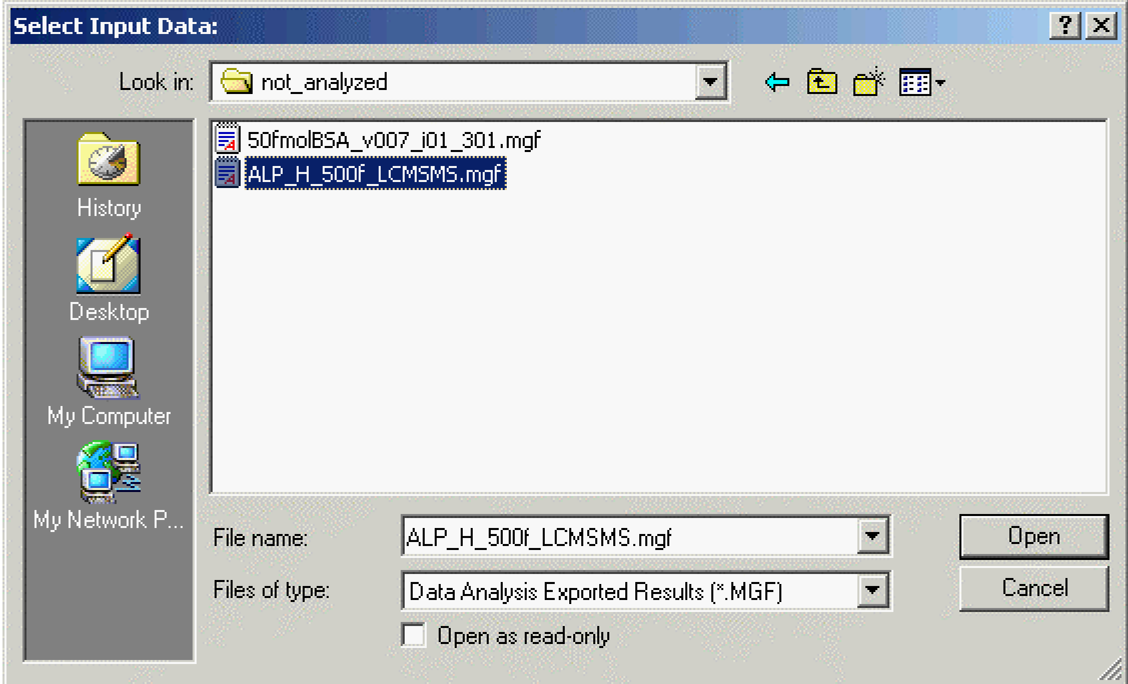


Figure 1

When the location of the mgf-file is not known choose **File | Find**, select the data type (which is mgf) and the folder where the files are located (Figure 2). Upon clicking **FindNow**, this folder is searched for mgf-files which are presented as a list. Highlight the file with the left mouse button. For multiple files use the CONTROL or SHIFT button and the left mouse button. Then, **Open** the file(s).

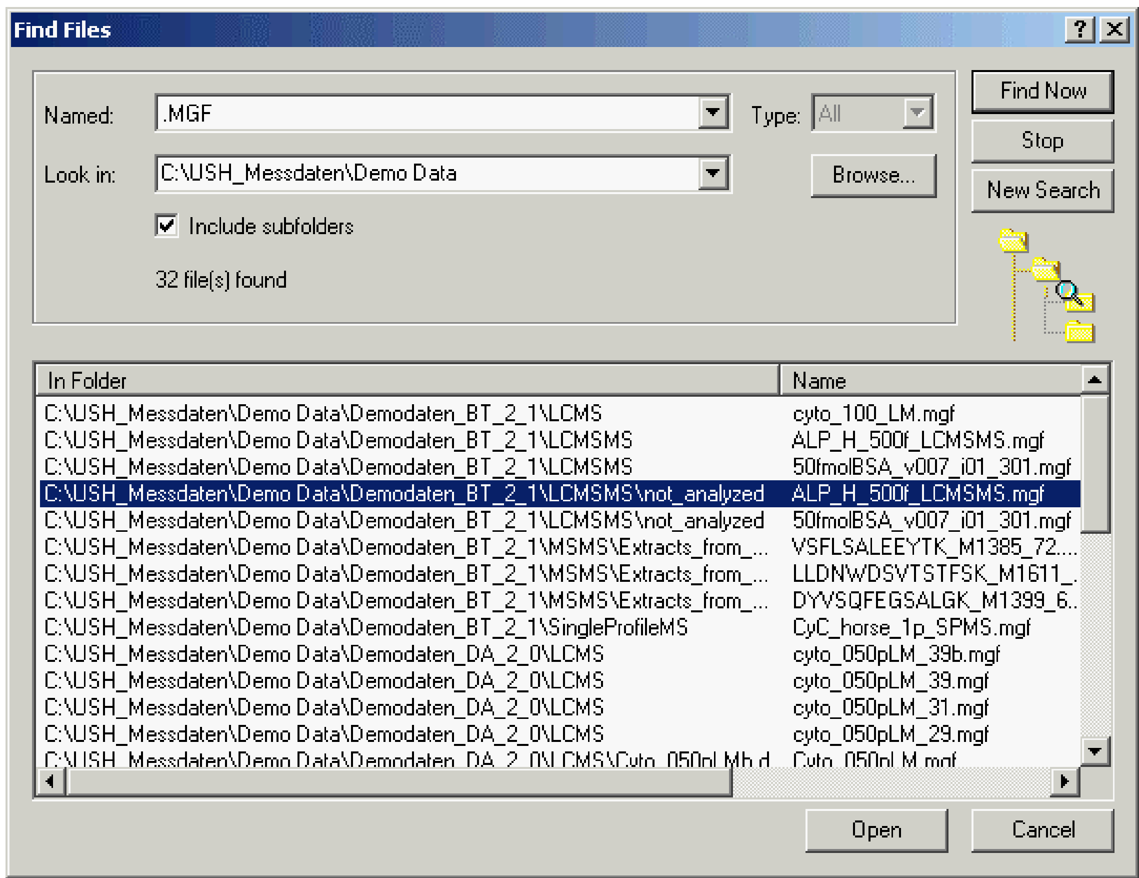


Figure 2

Then, the spectrum is shown (Figure 3).

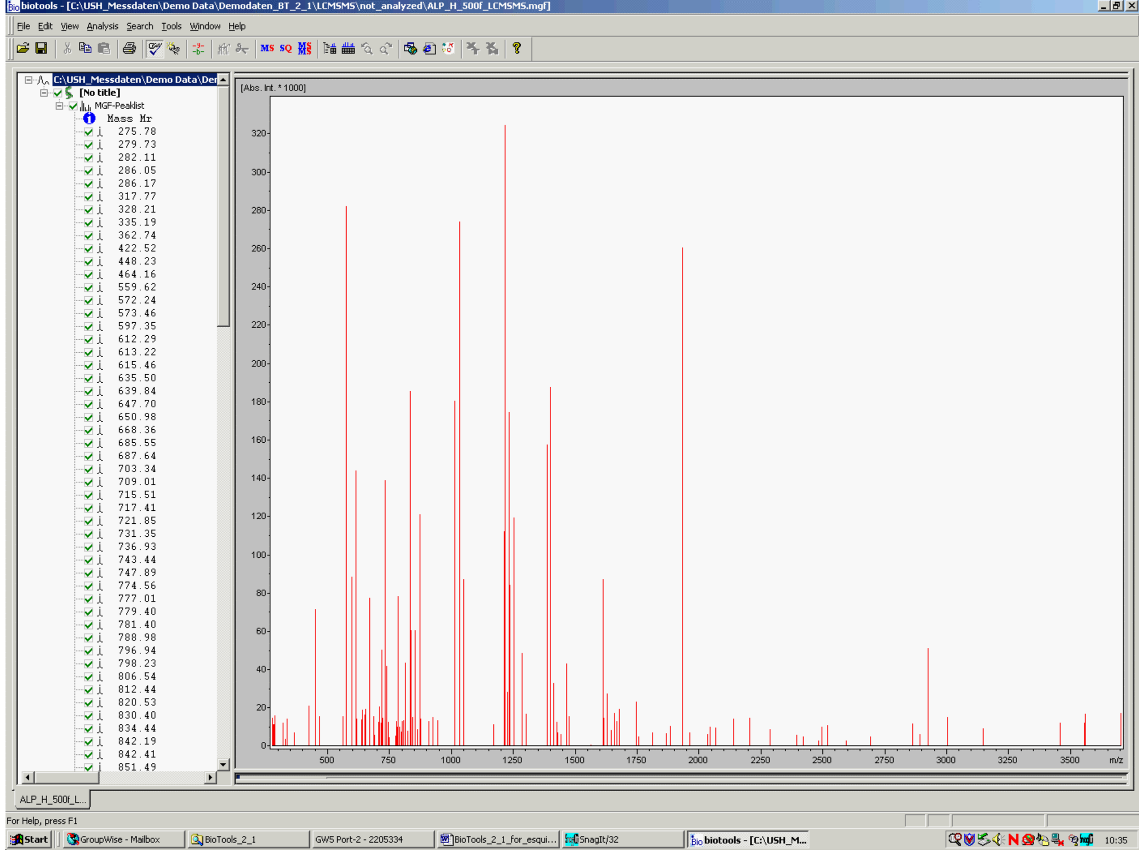


Figure 3

In the treeview on the left side of the screen the file name and the file location are shown. The MGF-peaklist contains the measured neutral masses of the peptides. Those masses are shown as a histogram in the Spectrum window on the upper right side of the screen.

E.1.2.2. Loading a protein sequence in the SequenceEditor

It is necessary to open a protein sequence in the SequenceEditor prior to the Sequence Confirmation procedure. This is done by double-clicking (left mouse button) the sequence file (apolipoproteina1_human_signal_s.sqs) to unfold the sequences in this file, and those are opened by a left mouse button double click (Figure 4).

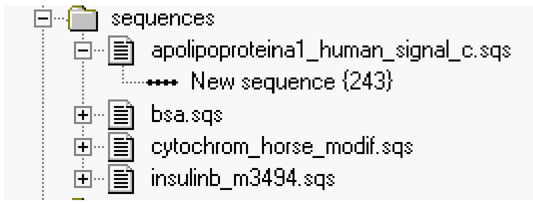


Figure 4

Then, the sequence is opened in the Sequence window (Figure 5).

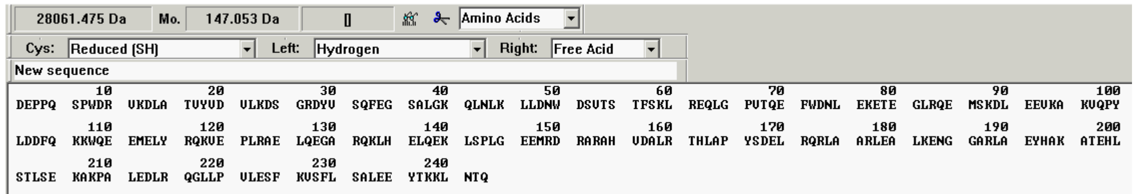


Figure 5

If required new sequences can be introduced in the Sequence Editor by opening a new window (**File | New**) and using copy and paste (**Edit | Paste**) or typing a sequence in the window. The introduction of modifications is described in the Sequence Editor manual.

E.1.2.3. Starting the Sequence Confirmation

In BioTools choosing **Search | SearchForMasses** after activating the MGF-peaklist or double-clicking the MGF-peaklist in the treeview starts the Sequence Confirmation tool. Then, the Search For Masses window (Figure 7) opens. The following parameters can be set:

Parameter	Explanation / Recommendation
Monoisotopic/ Average	Monoisotopic masses should be selected.
Error	Select the unit and the value.
Enzyme	Choose the enzyme that has been used for the digest.
Specific/unspecific cleavages	For both options an unlimited number of missing cleavages (partials) is analyzed.

Modifications	Upon selecting **EditMods** in the Search For Masses window (Figure 7) modifications can be analyzed. Then, the Modification Editor window (Figure 6) is opened. Choose one residue type or "all" and activate the desired modifications (left mouse button). Select the modification type and **Add** the modification to the list on the left side of the window. **Edit type** enables the user to define modifications. **Ok** sends the selected modifications back to the Search for Masses window.
Mass Selection	Select the masses that should be used for the search.

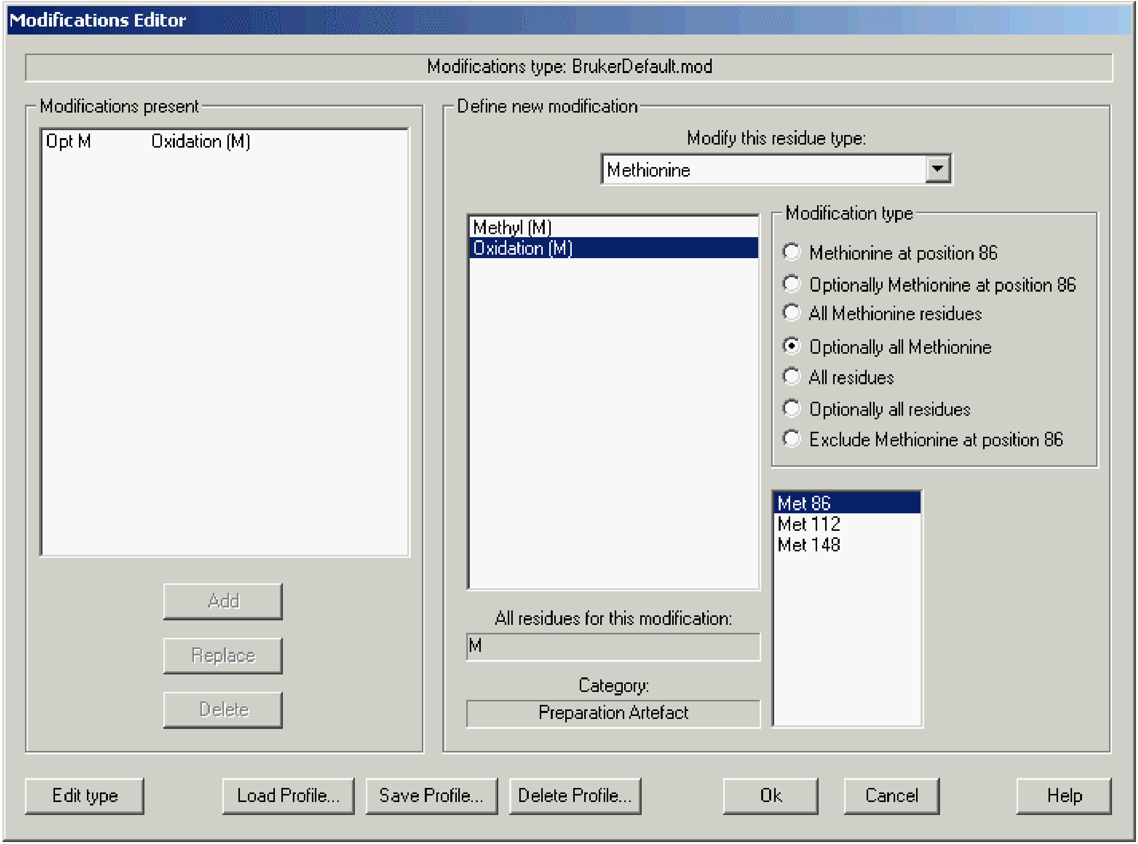


Figure 6

For the file Alp_H_500f_LCMSMS.mgf use the parameters shown in Figure 7.

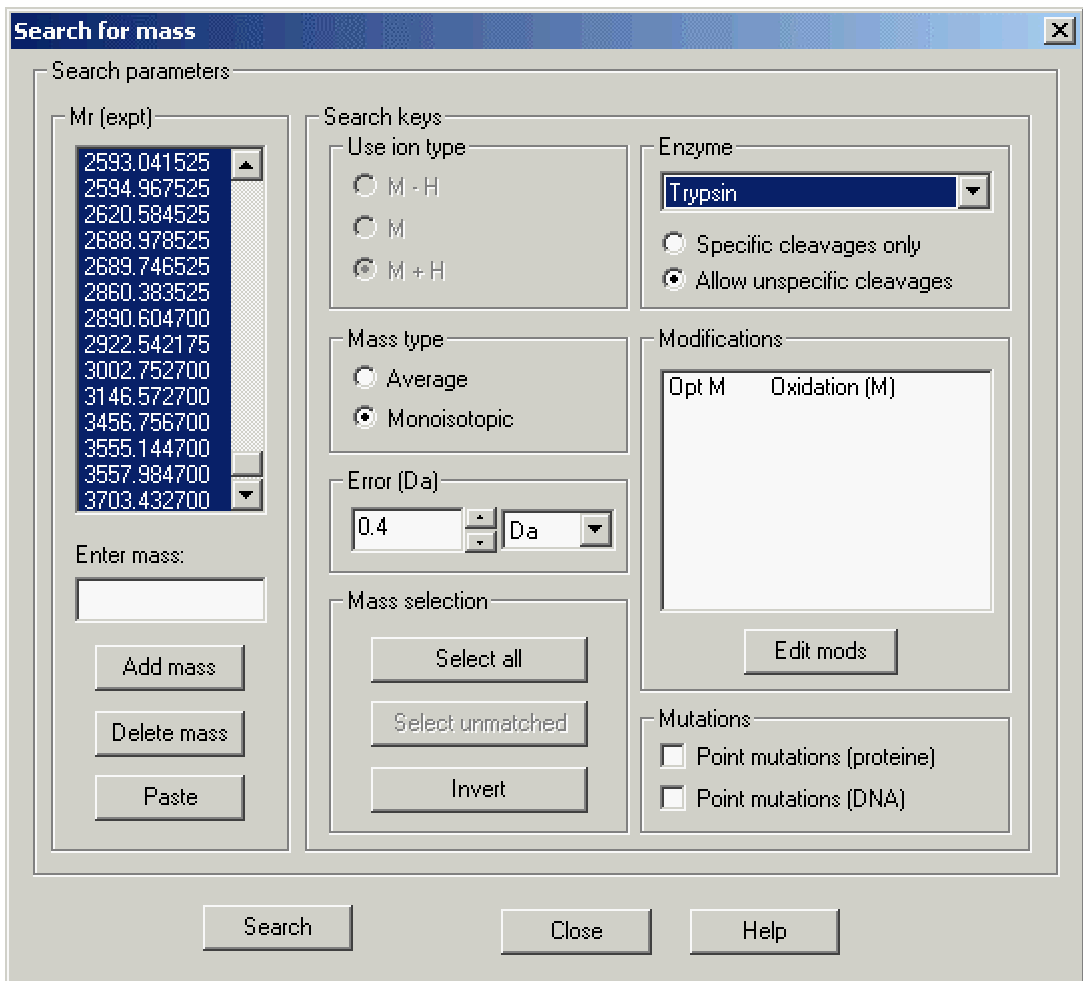


Figure 7

Choose **Search** to start the search. Then, the Mass Search Results window appears (Figure 8). It offers the opportunity for a preliminary match of masses vs. possible peptides sequences.

For the file Alp_H_500f_LCMSMS.mgf trypsin was chosen as enzyme, but unspecific cleavages were allowed in the analysis. Therefore, many matches for unspecific cleavages are shown as indicated by the absence of the blue bracket on either side of the sequence.

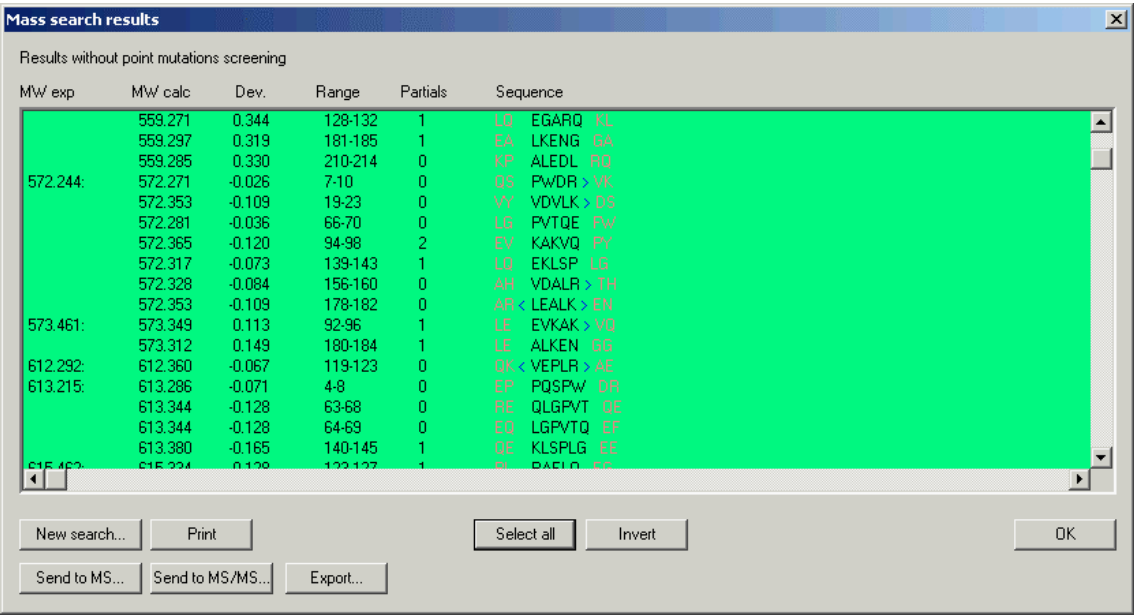


Figure 8

Choose **SelectAll** and **SendToMS…** in the Mass Search Results window to send the result to BioTools.

E.1.2.4. The Sequence Confirmation result

Activate "Mass Search Matches" in the BioTools treeview to see the result (Figure 9).

Treeview

The treeview (on the left side of the screen) now shows additional information: "Mass Search Matches" contains the search parameters, possible modifications and a result table with the following content:

- calculated molecular mass
- BioTools score
- mass deviation
- amino acid range
- partials (missing cleavages)
- peptide sequence.

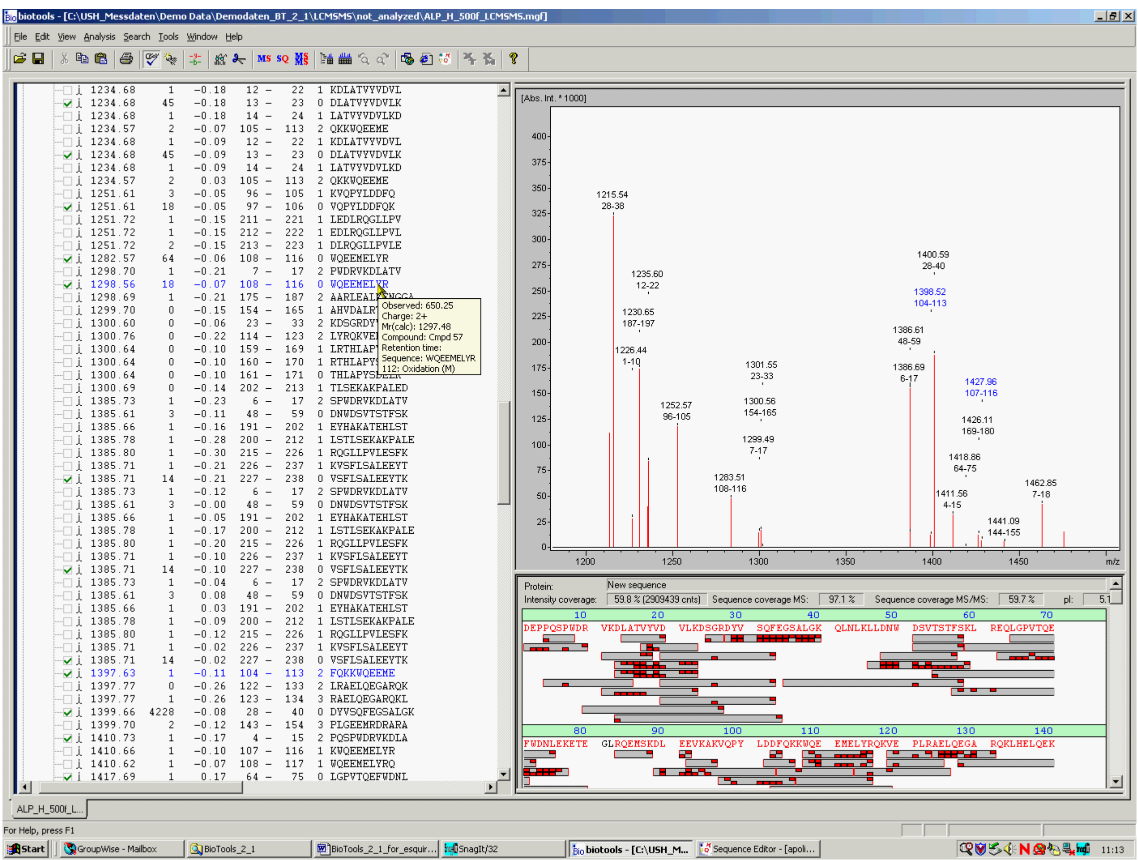


Figure 9

When the cursor is placed on top of any one peptide in the treeview, additional information appears for about 10 seconds which lists the observed mass, the charge state, the compound number and the retention time.

Within the selected mass deviation all possible peptide matches resulting from the digest (specific or unspecific) are listed in the treeview. When multiple matches have been found for one measured mass, BioTools selects the one with the highest BioTools score and labels it with a red check mark. The user can also manually select the match that is deemed appropriate. When all meaningful matches are accepted and checked, the unchecked peptides can be removed irreversibly by clicking "Mass Search Matches" at the top of the treeview with the Right mouse button and selecting **DeleteUncheckedSublevel**.

Spectrum window

The Spectrum window (on the upper right side of the screen) shows the matched and unmatched signals when **View | MatchedAndUnmatchedPeaks** and "Mass Search Matches" are activated.

Sequence Viewer window

The Sequence Viewer window contains three parameters:

- The intensity coverage
- The Sequence coverage for the MS signals
- The Sequence coverage for the MS(n) signals

The fragmentation pattern (Figure 10) shows the amino acid sequence: the red bars in the upper part of the gray bar represent the fragments containing the N-terminus (a- and b-ions) and the lower red bars show the matched fragments with the C-terminus (y-ions).

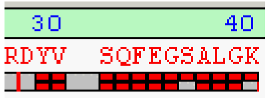


Figure 10

Connection between the result window parts

Upon activating one mass in the treeview the corresponding signal is labeled in the mass spectrum (mass and amino acid range) in the Spectrum window and the sequence is also highlighted green in the Sequence Viewer window. Vice versa, when a peptide is activated by clicking on the gray bar in the Sequence Viewer window it is also labeled in the Spectrum window and highlighted in the treeview.

Printing

For printing the result an appropriate Printing profile is required, and this is created in **Tools | Options**. Choose **AddProfile (**Figure 11) and type in a desired name. Select "Spectrum" and "Matched Sequences" as parameters. Then, the spectrum, the mgf-peaklist and the result list can be printed upon choosing **File | Print**.

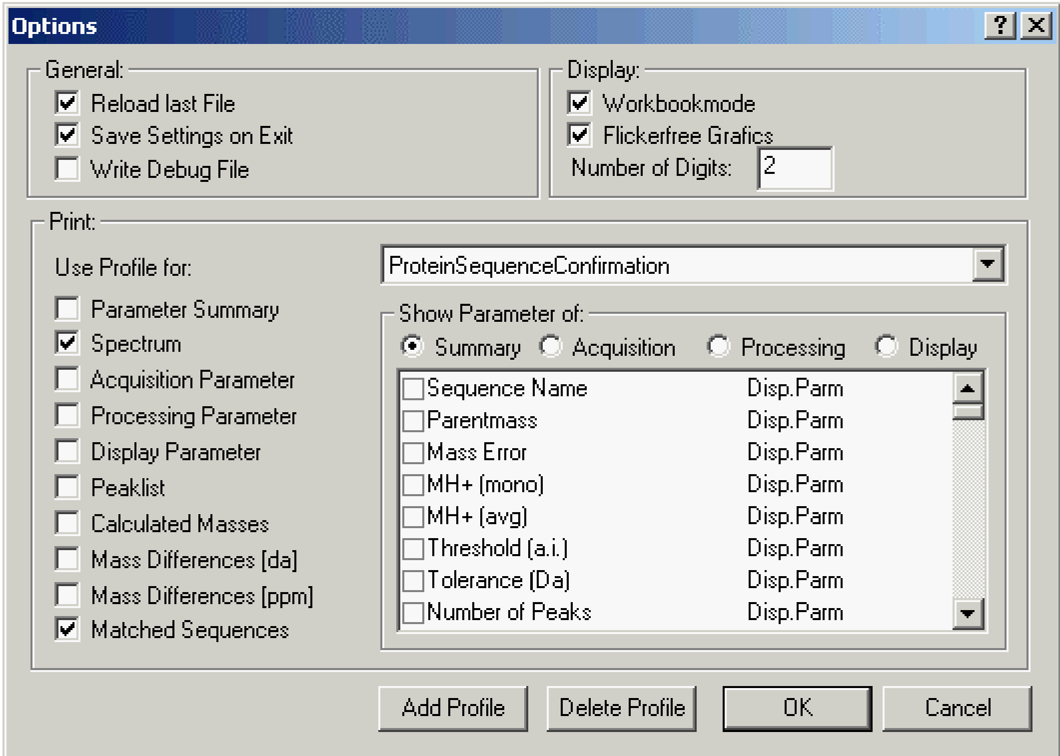


Figure 11

Saving

Upon activating the disk button from the toolbar the Sequence Confirmation result is saved in an additional file (MDI-format).

E.1.3. Protein Identification with subsequent Sequence Confirmation

E.1.3.1. Loading the mgf-file

The tutorial "Protein identification using esquire3000 + (LC-MS/MS data)" describes the export the complete LC-MS/MS-peaklist as an mgf-file in DataAnalysis. In BioTools use **File | Open** or **File | Find** to open the mgf-file.

When the location of the mgf-file is known choose **File | OpenSpectrum**, select the data type (which is mgf) and the location of the file (Figure 12). Highlight the file with the left mouse button. For loading multiple files use the CONTROL or SHIFT button and the left mouse button. Then **Open** the file.

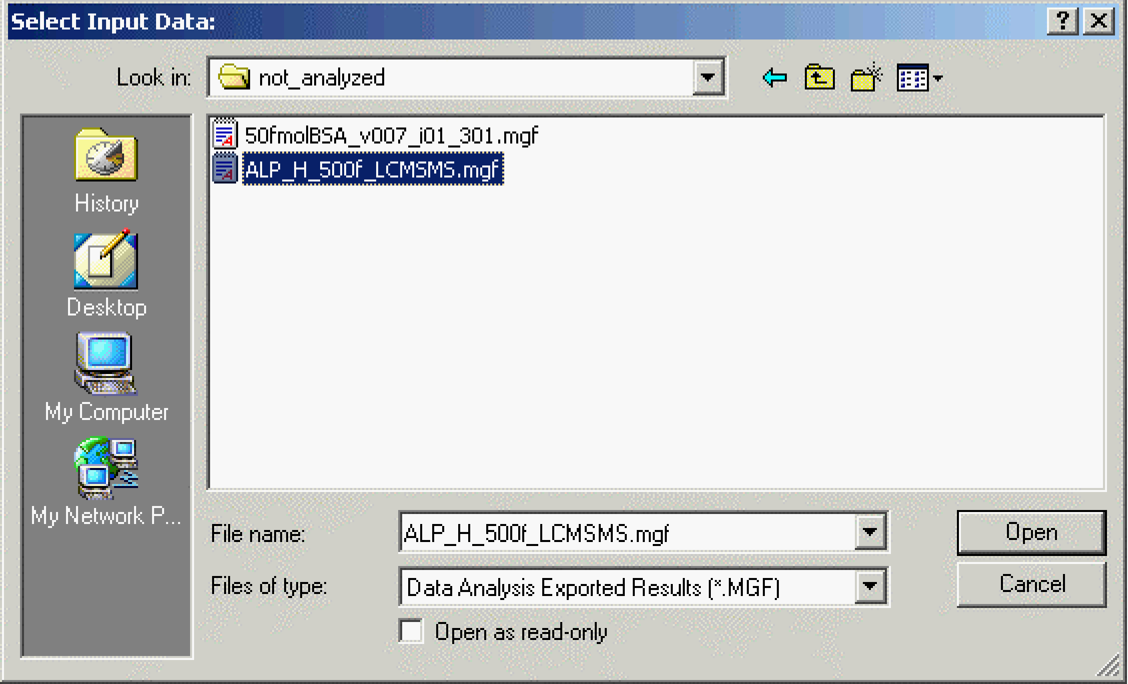


Figure 12

When the location of the mgf-file is not known choose **File | Find**, select the data type (which is mgf) and the folder where the files are located (Figure 13). Upon clicking **FindNow**, this folder is searched for mgf-files which are presented as a list. Highlight

the file with the left mouse button. For loading multiple files use the CONTROL or SHIFT button and the left mouse button. Then, **Open** the file(s).

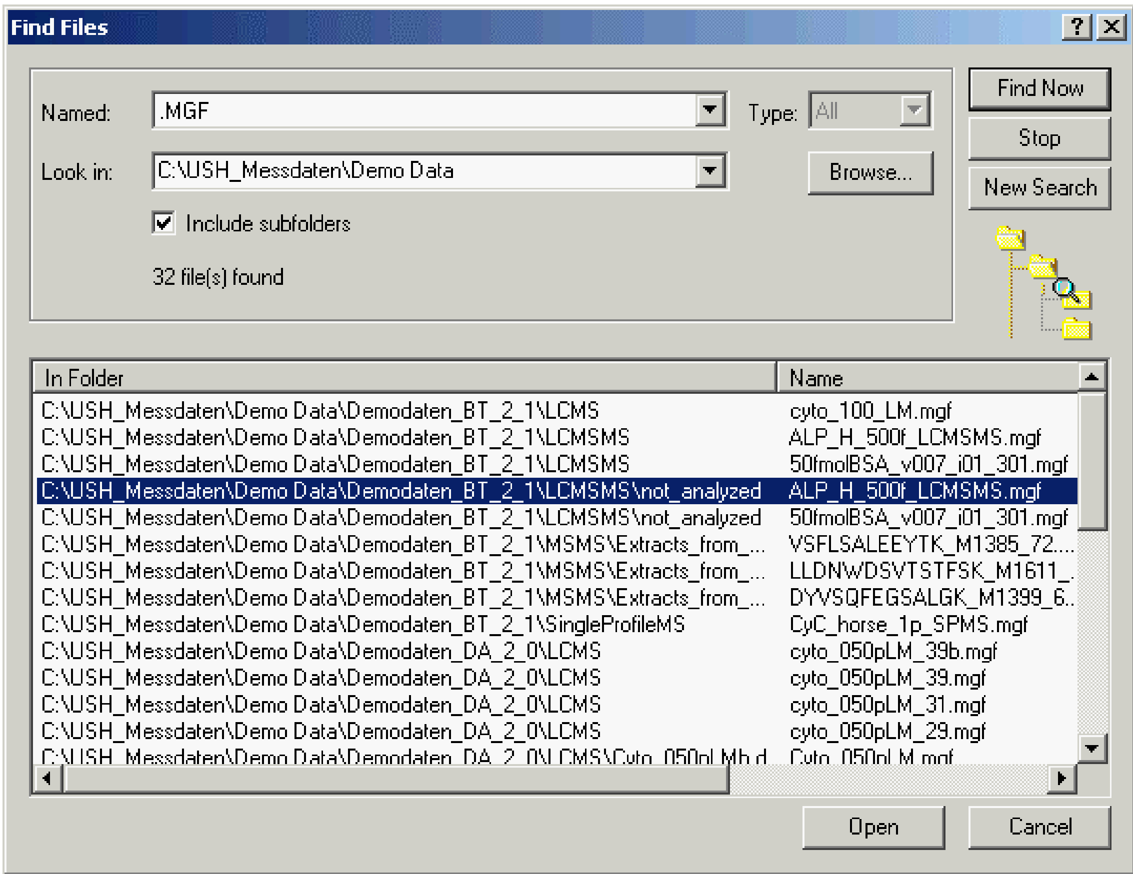


Figure 13

Then, the spectrum is shown (Figure 14).

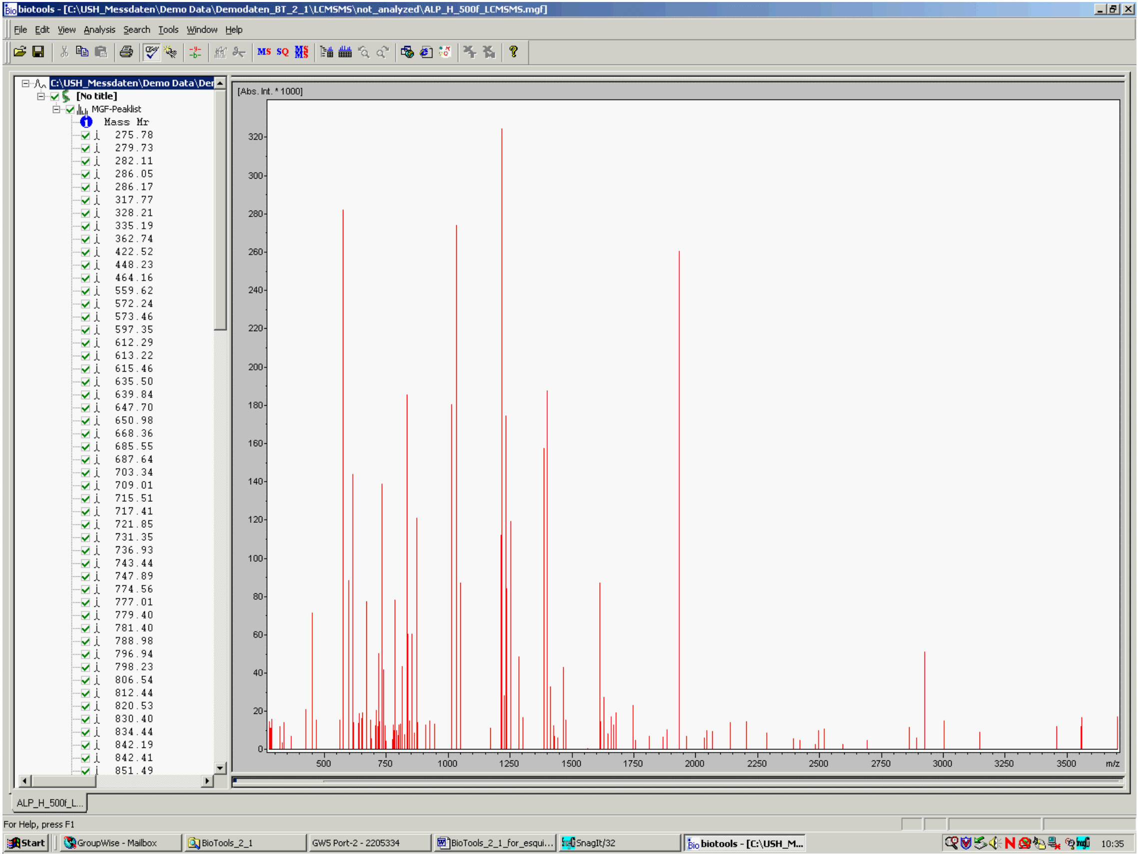


Figure 14

In the treeview on the left side of the screen the file name and the file location are shown. The MGF-peaklist contains the measured neutral masses of the peptides. Those masses are shown as a histogram in the Spectrum window on the upper right side of the screen.

E.1.3.2. Starting the Database Search

In this chapter protein identification via BioTools without the Batch mode (refer to "Protein identification using esquire3000 + (LC-MS/MS data)") is shown. Choose **Search | MascotMS/MSIonSearch** to open the parameter window, and set the parameters as shown in Figure 15.

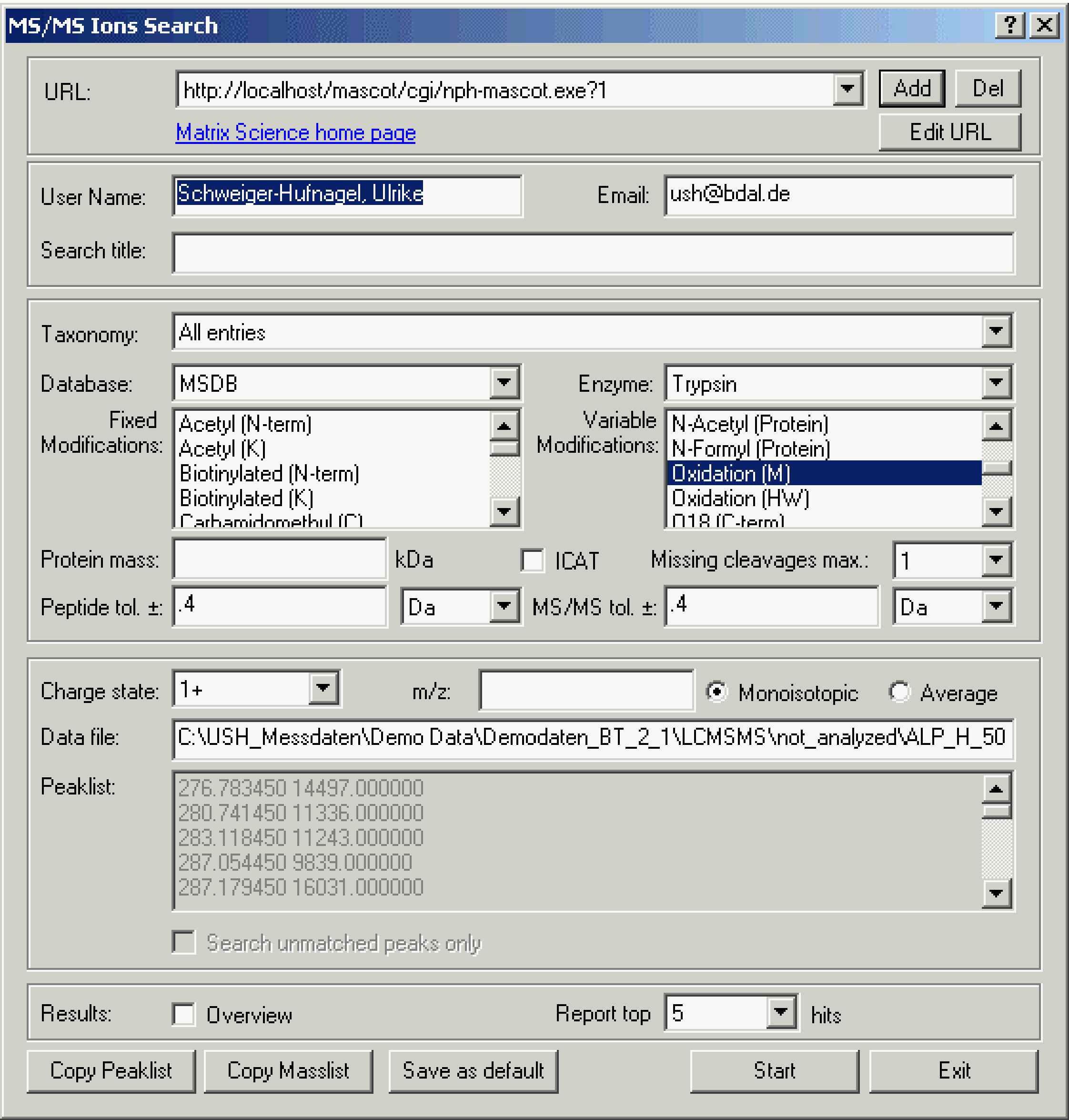


Figure 15

Start is used to start the search. The intermediate window (Figure 16) shows the progress of the search.

{MATRIX SCIENCE} **Mascot Search**

Licensed to: Bruker Daltonik, (2 processors).
Finished uploading search details...
IMPORTANT: If you get disconnected or choose not to wait for your search results
DO NOT RESUBMIT THE SEARCH. Your results will be sent by email when the search is complete
Searching....
.....5% complete
......10% complete
.......15% complete
...

Figure 16

Then comes the search result (Peptide Summary Report, Figure 17).

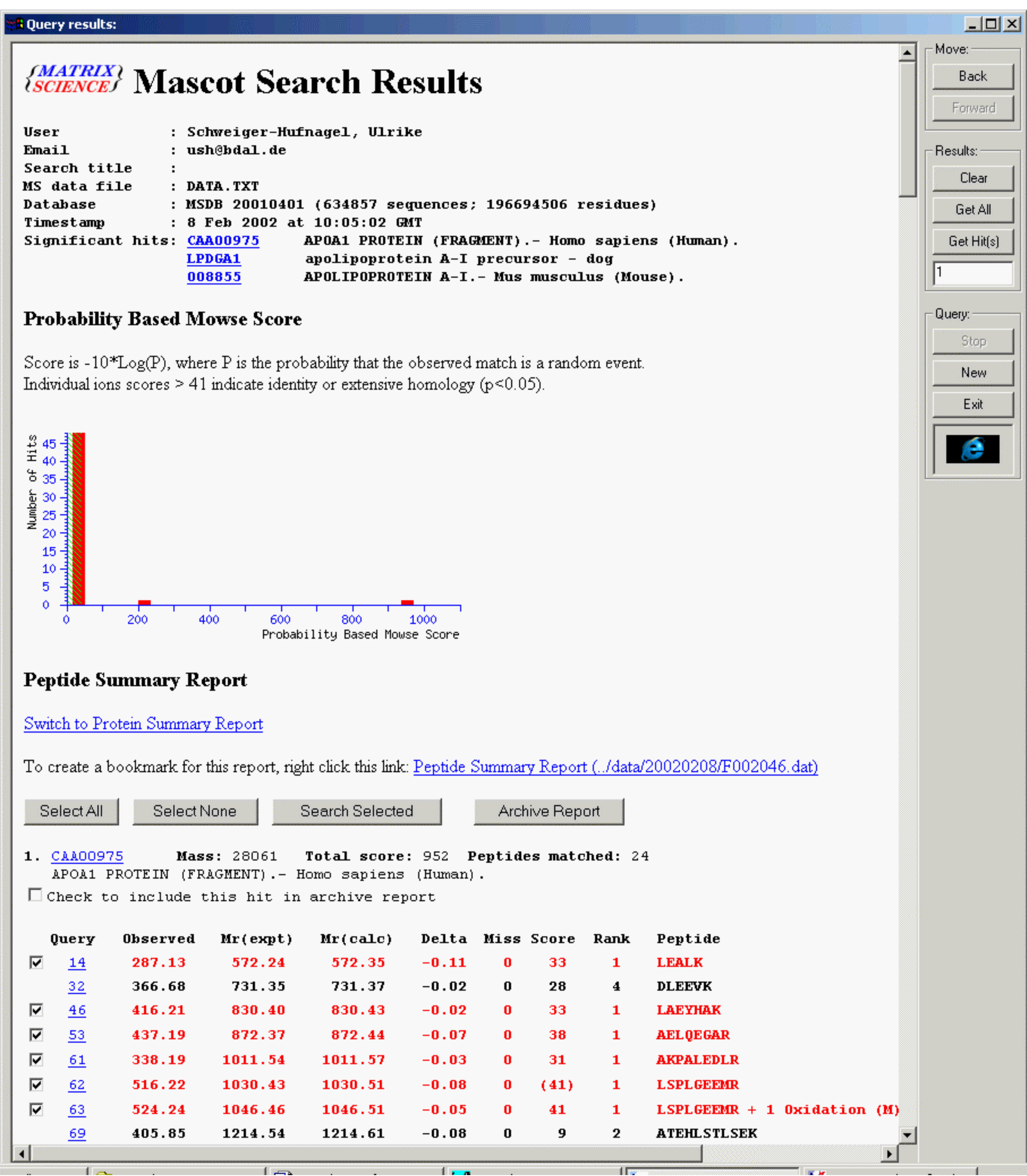


Figure 17

Upon choosing **GetHits** and the desired result number(s), e.g. number 1 or 2 or 1-2, the sequence information is sent to the peaklist in BioTools.

E.1.3.3. The Database Search result in BioTools

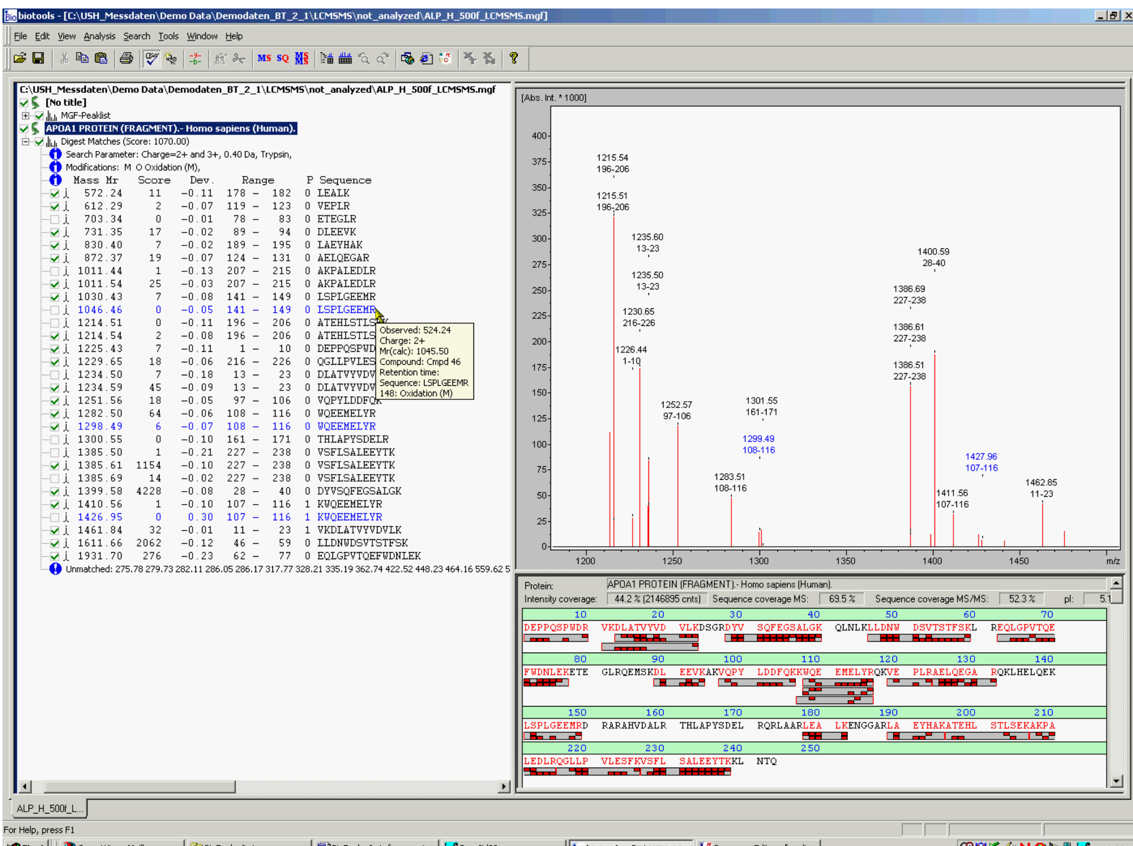


Figure 18

Treeview

The treeview (on the left side of the screen, Figure 18) now shows additional information: the Mascot score, the global charge limitation (DataAnalysis), the search parameters, possible modifications (blue font for modified peptides) and a result table with the following content:

- measured molecular mass
- BioTools score
- mass deviation
- amino acid range
- partials (missing cleavages)
- peptide sequence.

When the cursor is placed on top of any one peptide in the treeview, additional information appears for about 10 seconds which lists the fragmented m/z value, the charge, the calculated neutral mass, the compound number, the retention time and modification information for modified peptides.

Within the selected mass deviation all measured masses which can be explained by the theoretical protein sequence are listed in the treeview. All masses with an BioTools score of > 1 are labeled with a green check mark per default. This can be changed by the user. When all meaningful matches are accepted and checked, the unchecked peptides can be removed irreversibly by clicking "Digest Masses" at the top of the treeview with the Right mouse button and selecting **DeleteUncheckedSublevel**.

Spectrum window

The Spectrum window (on the upper right side of the screen) shows the matched and unmatched signals when **View | MatchedAndUnmatchedPeaks** and the "Digest Masses" are activated.

Sequence Viewer window

The Sequence Viewer window contains four parameters:

- The intensity coverage
- The Sequence coverage for the MS signals
- The Sequence coverage for the MS(n) signals
- The isoelectric point of the protein.

The fragmentation pattern (Figure 19) the amino acid sequence: the red bars in the upper part of the gray bar represent the fragments containing the N-terminus (a- and b-ions) and the lower red bars show the matched fragments with the C-terminus (y-ions).

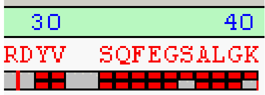


Figure 19

Connection between the result window parts

Upon highlighting one mass in the treeview the corresponding signal is labeled in the mass spectrum (mass and amino acid range) in the Spectrum window and the sequence is also highlighted green in the Sequence Viewer window. Vice versa, when

a peptide is activated by clicking on the gray bar in the Sequence Viewer window it is also labeled in the Spectrum window and highlighted in the treeview.

Printing

For printing the result an appropriate Printing profile is required, and this is created in **Tools | Options**. Choose **AddProfile** and type in a desired name. Select "Spectrum" and "Matched Sequences" as parameters (Figure 20). Then, the spectrum, the mgf-peaklist and the result list can be printed upon choosing **File | Print**.

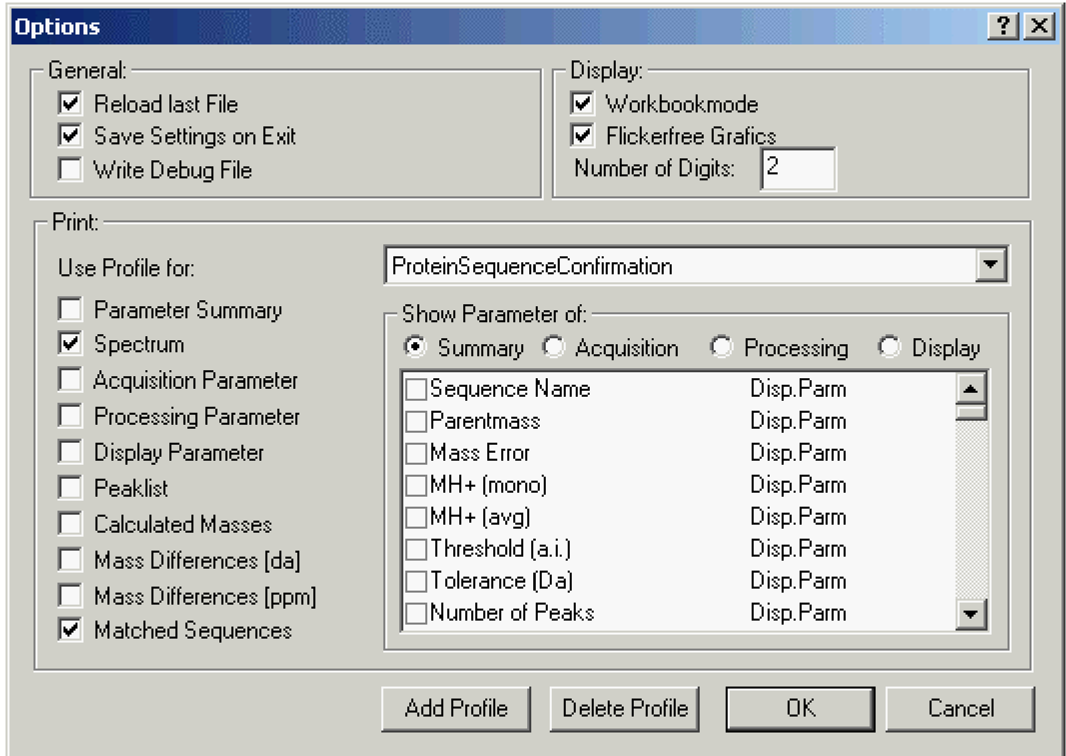


Figure 20

Saving

Upon activating the disk button from the toolbar the Sequence Confirmation result is saved in an additional file (MDI-format).

E.1.3.4. Starting the Sequence Confirmation

In SequenceEditor choosing **Search | SearchForMasses** after activating "digest masses" or double-clicking "digest masses" in the treeview starts the Sequence Confirmation tool. Then, the Search For Masses window (Figure 21) opens.

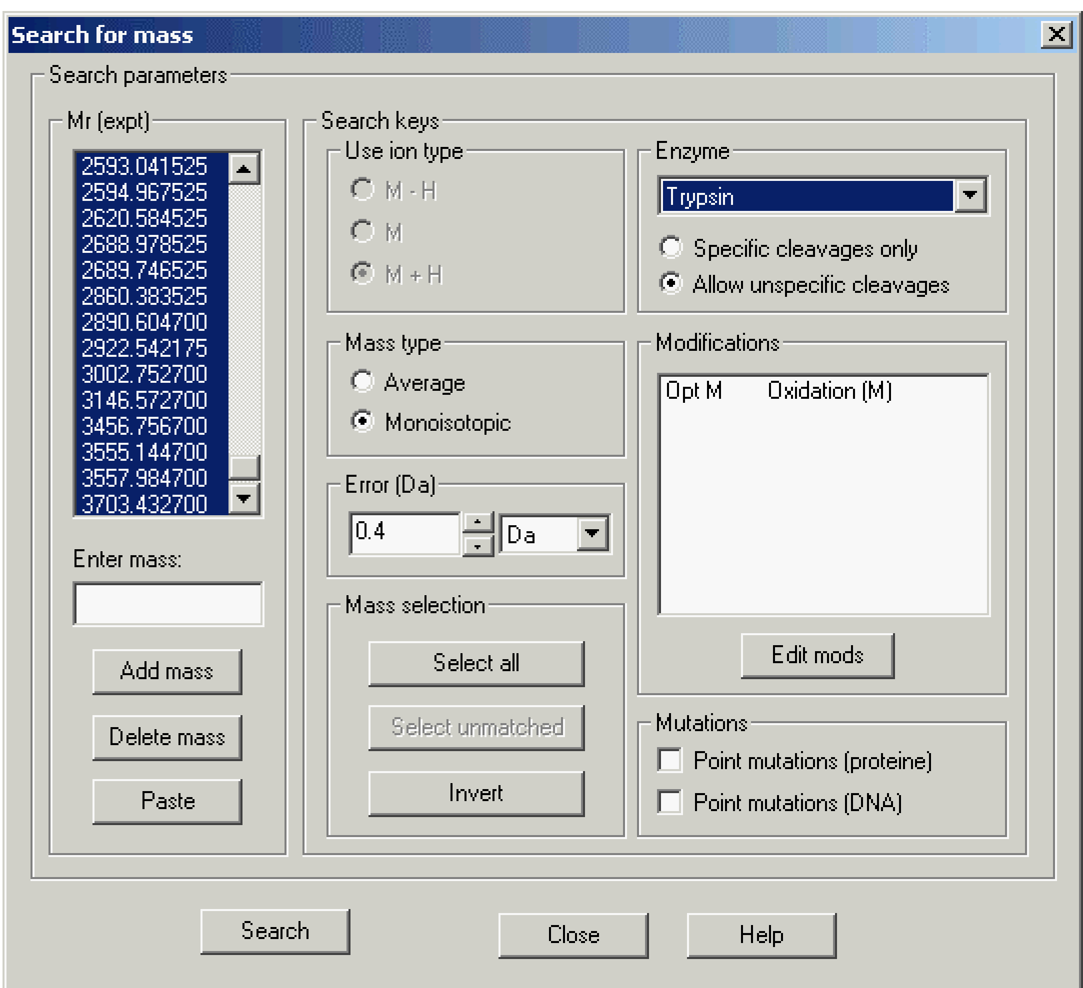


Figure 21

The following parameters can be set:

Parameter	Explanation / Recommendation
Monoisotopic/ Average	Monoisotopic masses should be selected.

Error	Select the unit and the value.
Enzyme	Choose the enzyme that has been used for the digest.
Specific/unspecific cleavages	For both options an unlimited number of missing cleavages (partials) is analyzed.
Modifications	Upon selecting **EditMods** in the Search For Masses window (Figure 21) modifications can be analyzed. Then, the Modification Editor window (Figure 22) is opened. Choose one residue type or "all" and activate the desired modifications (left mouse button). Select the modification type and **Add** the modification to the list on the left side of the window. **Edit type** enables the user to define modifications. **Ok** sends the selected modifications back to the Search for Masses window.
Mass Selection	Select the masses that should be used for the search.

For the file Alp_H_500f_LCMSMS.mgf use the parameters shown in Figure 22.

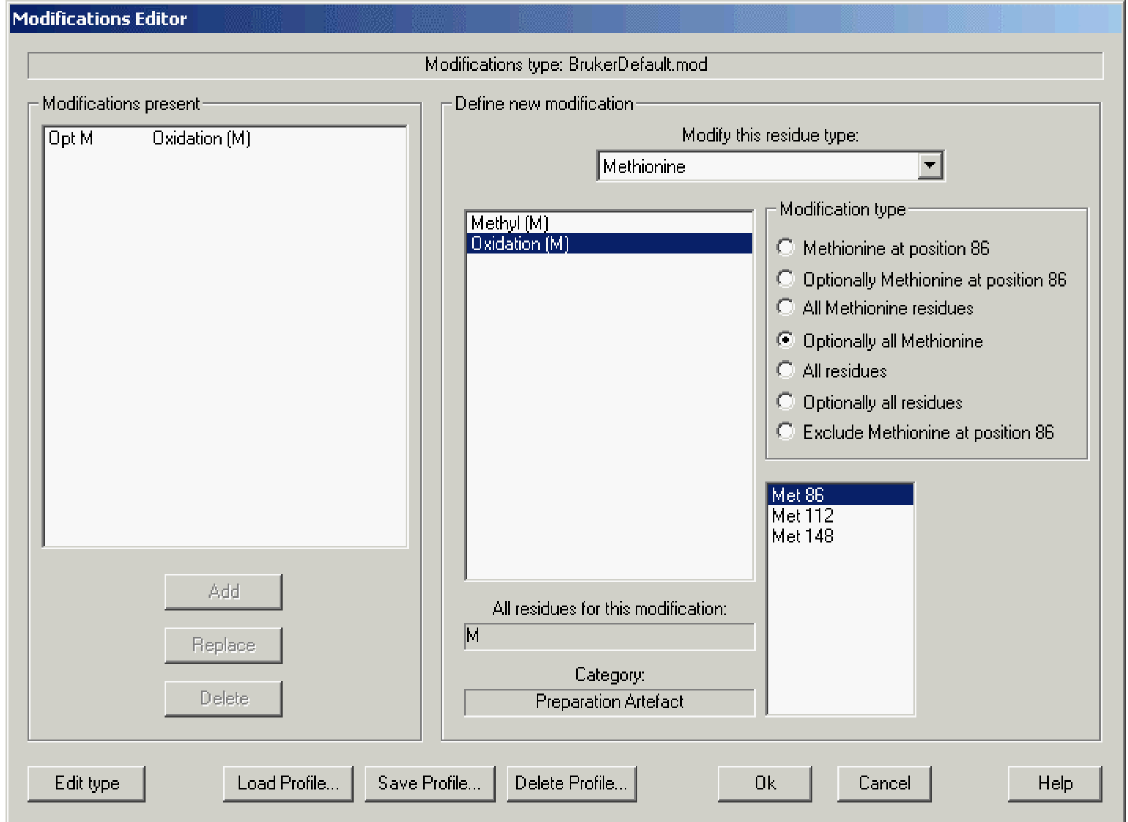


Figure 22

Choose **Search** to start the search. Then, the Mass Search Results window appears (Figure 23). It offers the opportunity for a preliminary match of masses vs. possible peptides sequences.

For the file Alp_H_500f_LCMSMS.mgf no enzyme was selected (option "trypsin" combined with "allow unspecific cleavages". Therefore, many matches for unspecific cleavages are shown as indicated by the absence of the blue bracket on either side of the sequence.

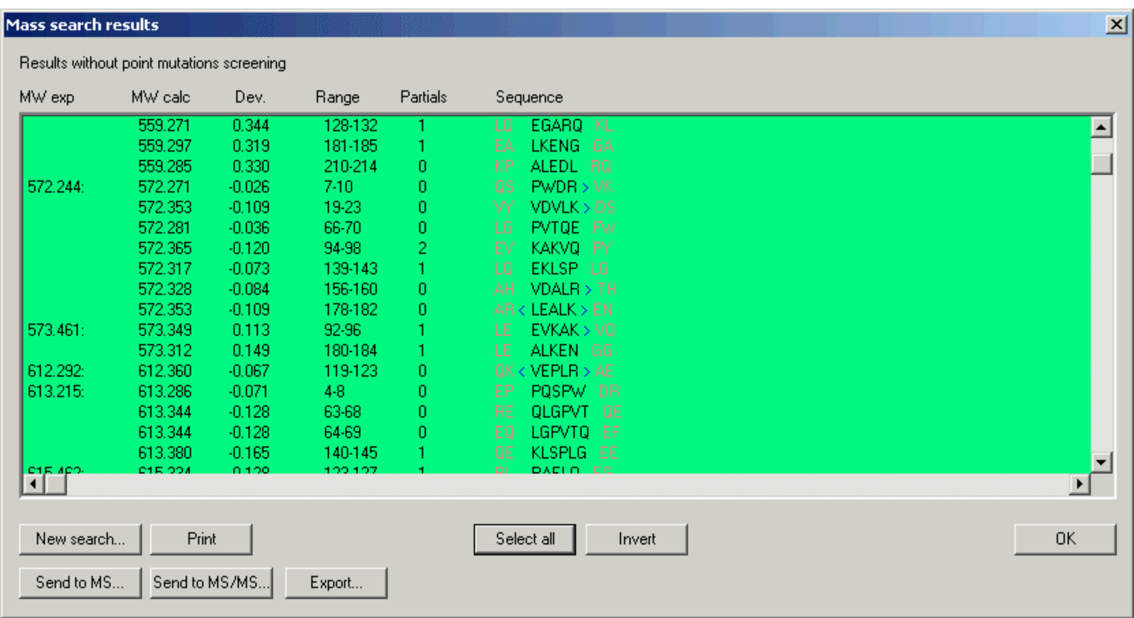


Figure 23

Choose **SelectAll** and **SendToMS** in the Mass Search Results window to send the result to BioTools.

E.1.3.5. The Sequence Confirmation result

The treeview, the Spectrum window, the Sequence Viewer (Figure 24), printing and saving have been described above. But one difference exists to the treeview described before (E.1.3.3): here, in the protein sequence confirmation procedure, the treeview shows the calculated neutral molecular masses and not the measured ones!

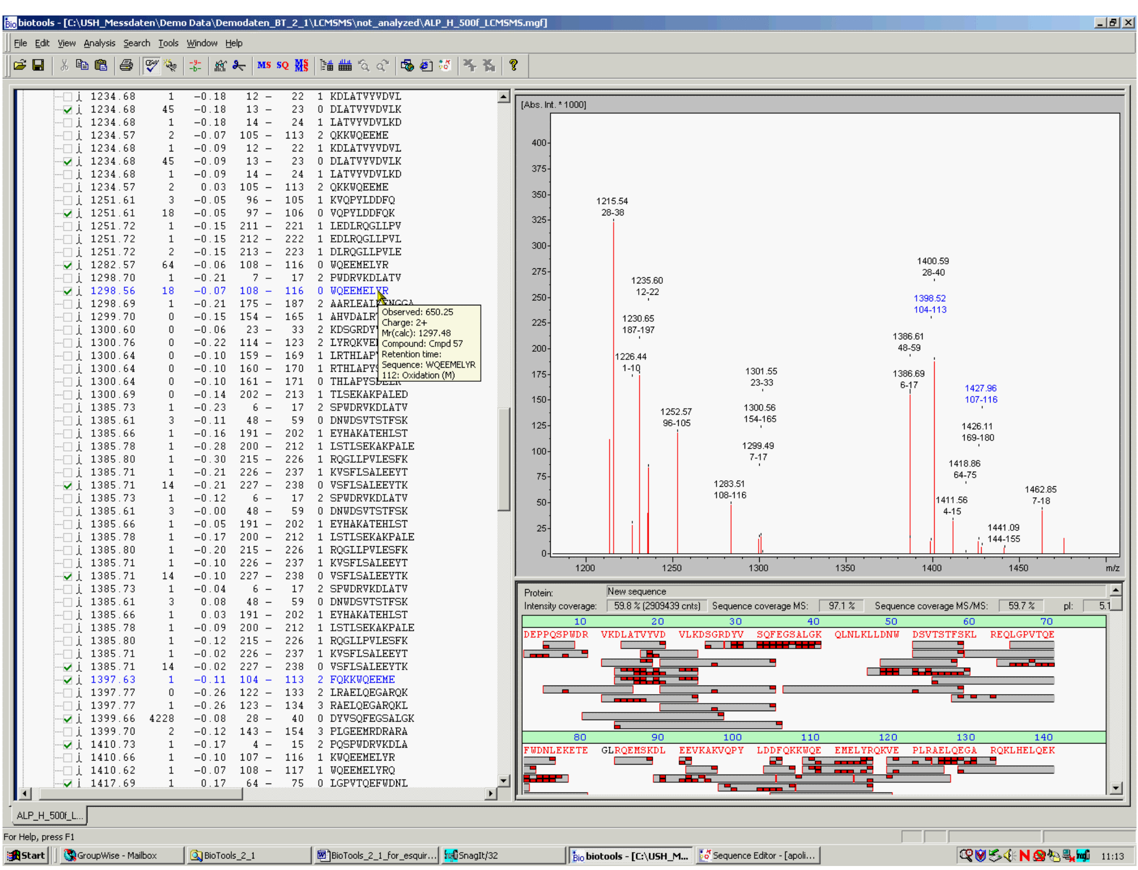


Figure 24

Within the selected mass deviation all possible peptide matches resulting from the digest (specific or unspecific) are listed in the treeview. When multiple matches have been found for one measured mass, BioTools selects the one with the highest BioTools score and labels it with a red check mark. The user can also manually select the match that is deemed appropriate. When all meaningful matches are accepted and checked, the unchecked peptides can be removed irreversibly by activating "Mass Search Matches" at the top of the treeview with the Right mouse button and selecting **DeleteUncheckedSublevel**.

E.2 MGF-files from single LC-MS/MS compounds

The handling of mgf-files from single LC-MSMS compounds is the same as for all compounds of a LC-MS/MS dataset (see before).

E.3 MS/MS-BSC-files

DYVSQFEGSALGK_M1399_67.bsc, extracted from human apolipoprotein and analysed by LC-MSMS will serve as an example for the use of MS/MS-BSC-files in BioTools. A general description of sample preparation, peptide separation via Hystar, data acquisition via esquire 5.0, and data analysis is given in "Protein identification using esquire3000 + (LC-MS/MS data)".

E.3.1. Protein identification

E.3.1.1. Loading the bsc-file

Either from a LC-MS/MS run or from a Single profile MS/MS spectrum the MS/MS peaklist is transferred to BioTools. Two ways exist to do this:

1. Activate the MS/MS-spectrum in DataAnalysis (left mouse button) and select **Tools| BioTools**.

2. Or export the MS/MS-peaklist as a bsc-file (**File | Export | MassSpectrum)** and open it in BioTools via **File | Open** or **File | Find**.

When the location of the bsc-file is known, choose **File | OpenSpectrum**, select the data type (which is bsc) and the location of the file (Figure 25). Choose the file with the left mouse button, For loading multiple files use the CONTROL or SHIFT button and the left mouse button. Then **Open** the file.

When the location of the bsc-file is not known choose **File | Find**, select the data type (which is bsc) and the folder where the files are located (Figure 26). Upon selecting **FindNow**, this folder is searched for bsc-files which are presented as a list. Choose the file with the left mouse button, for multiple files use the CONTROL or SHIFT button and the left mouse button. Then, **Open** the file(s).

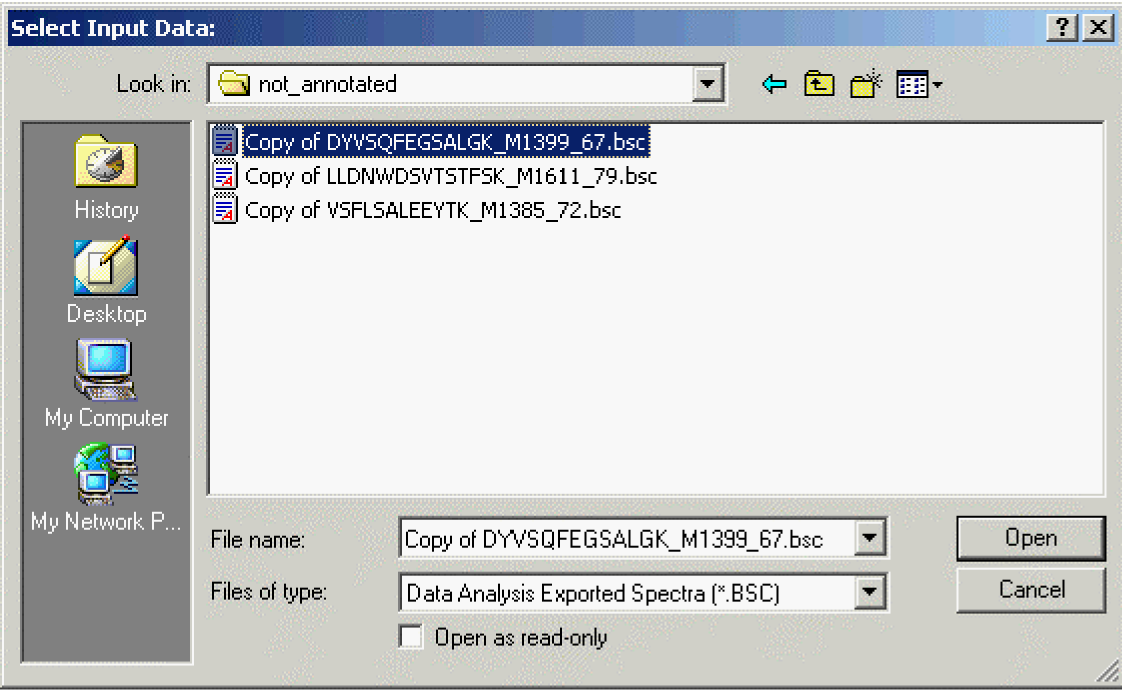


Figure 25

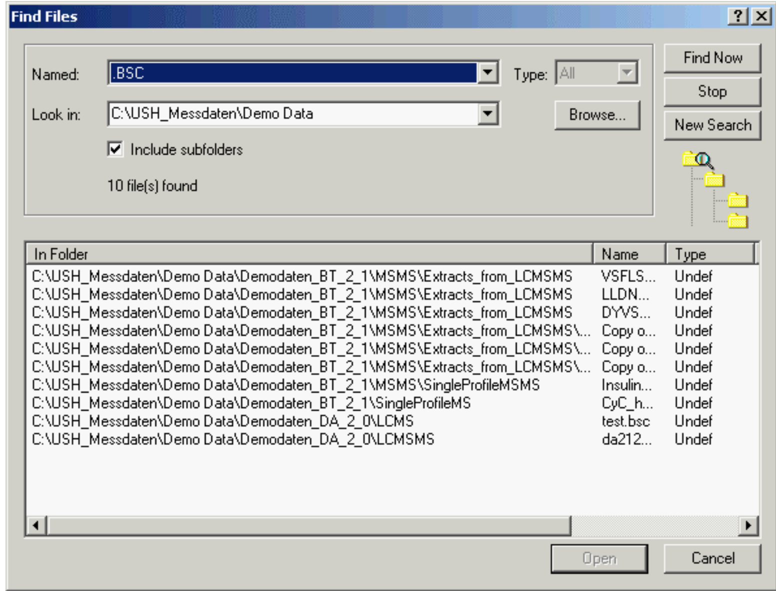


Figure 26

In any case, the spectrum is shown (Figure 27).

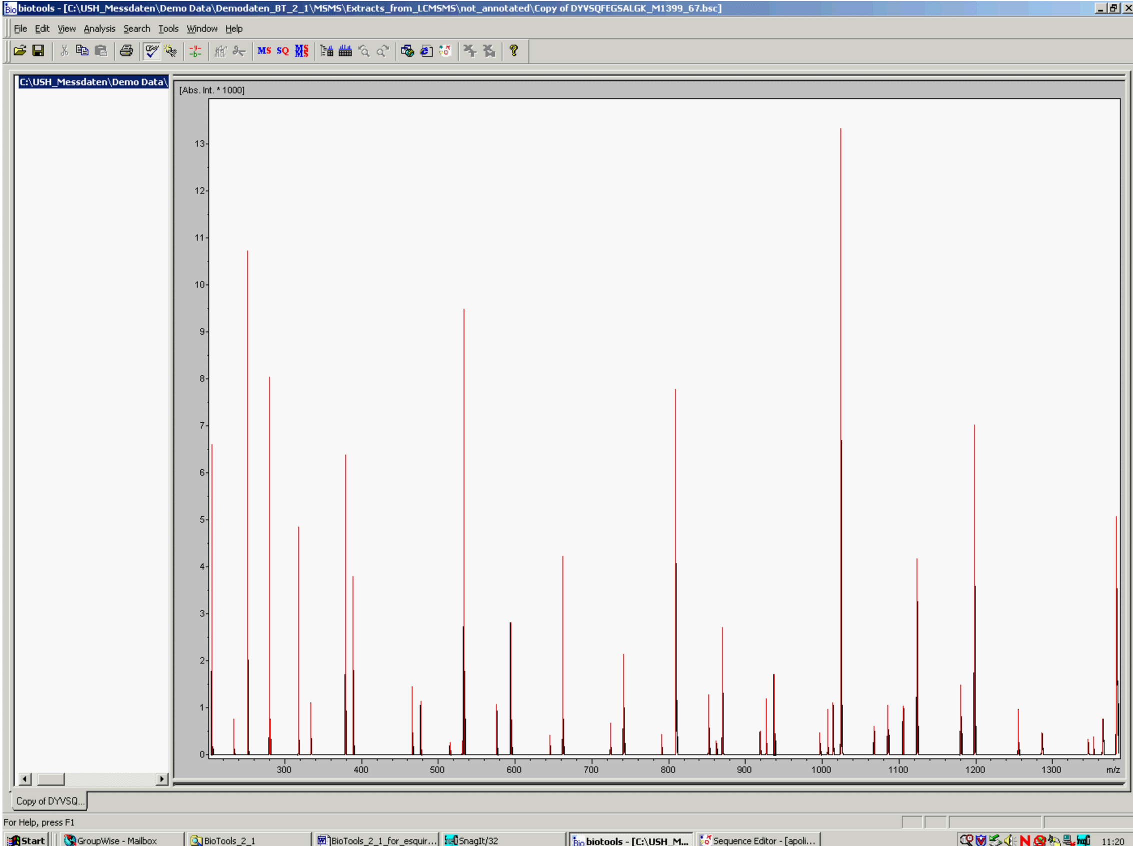


Figure 27

In the treeview on the left side of the screen the file name and the location are shown.

E.3.1.2. Starting the Database Search

In this chapter protein identification via BioTools without the Batch mode (refer to "Protein identification using esquire3000 + (LC-MS/MS data)") is shown. Choose **Search | MascotMS/MSIonSearch** to open the parameter window, and set the parameters as shown in Figure 28.

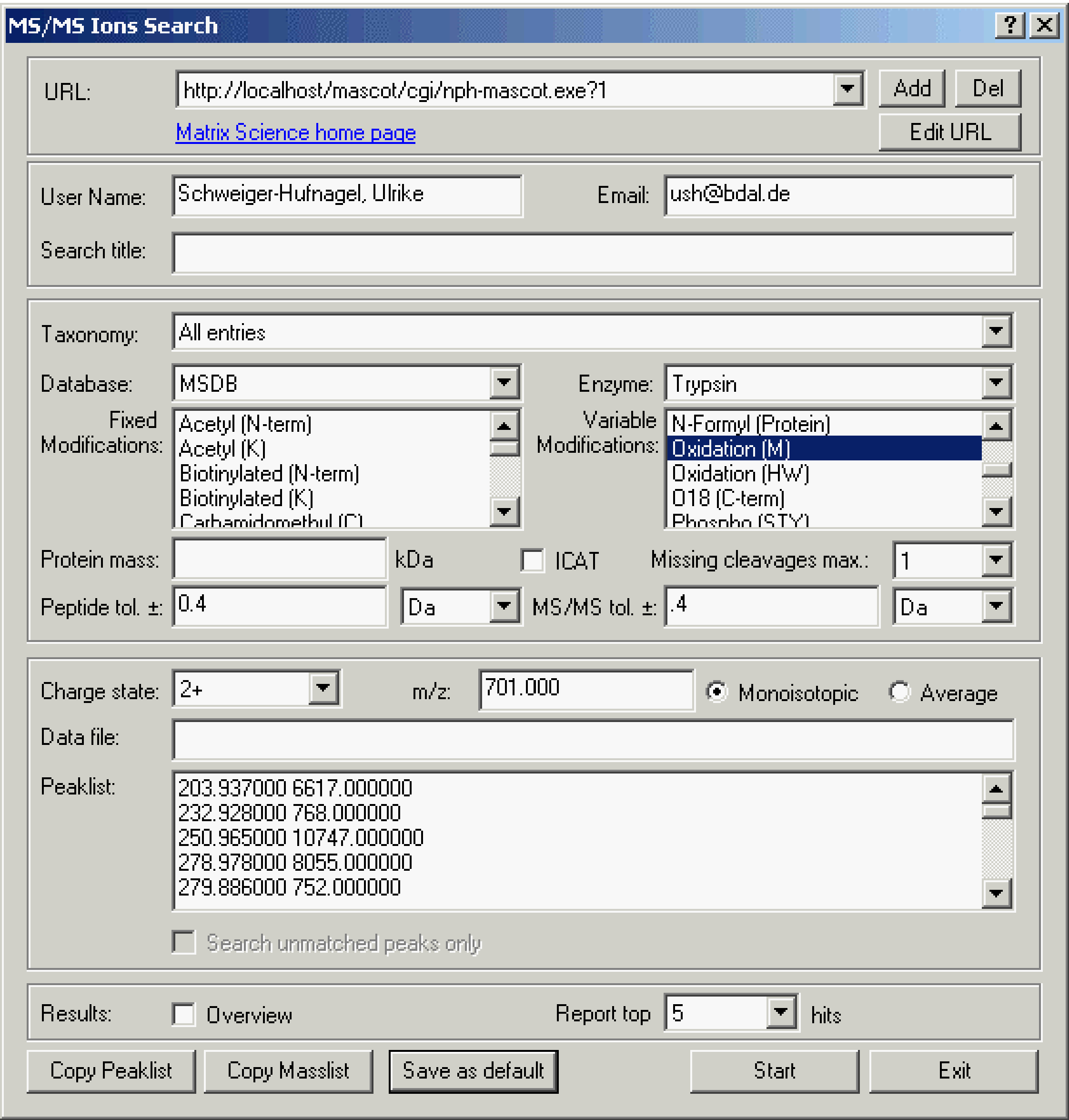


Figure 28

Start is used to start the search. The intermediate window (Figure 29) shows the progress of the search.

{MATRIX SCIENCE} **Mascot Search**

Licensed to: Bruker Daltonik, (2 processors).
Finished uploading search details...
IMPORTANT: If you get disconnected or choose not to wait for your search results
DO NOT RESUBMIT THE SEARCH. Your results will be sent by email when the search is complete
Searching....
.....5% complete
......10% complete
.......15% complete
...

Figure 29

Then comes the search result (Peptide Summary Report, Figure 30).

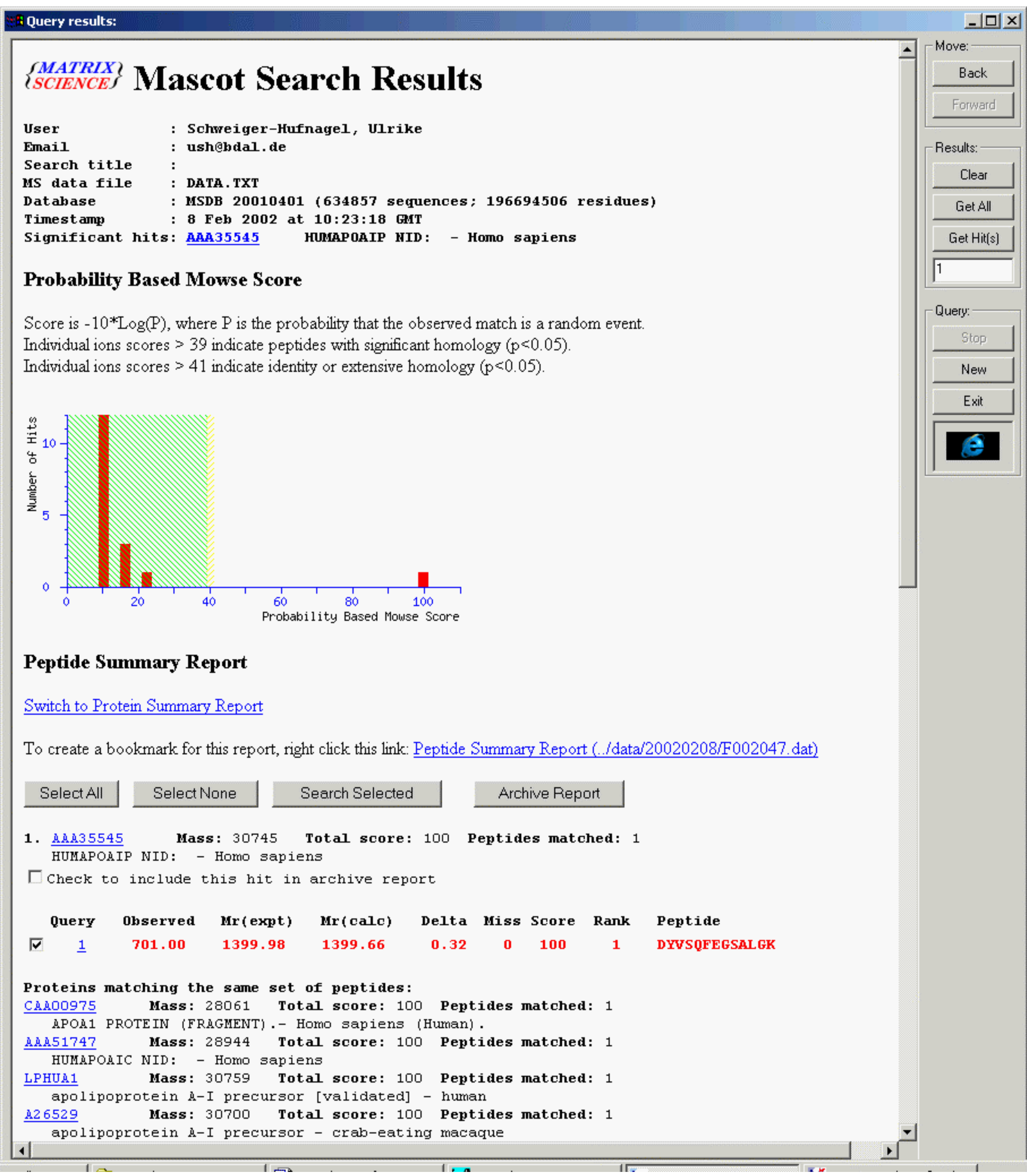


Figure 30

Upon choosing **GetHits** and the desired result number(s), e.g. number 1 or 2 or 1-2, the sequence information is sent to the peaklist in BioTools.

E.3.1.3. The Database Search result in BioTools

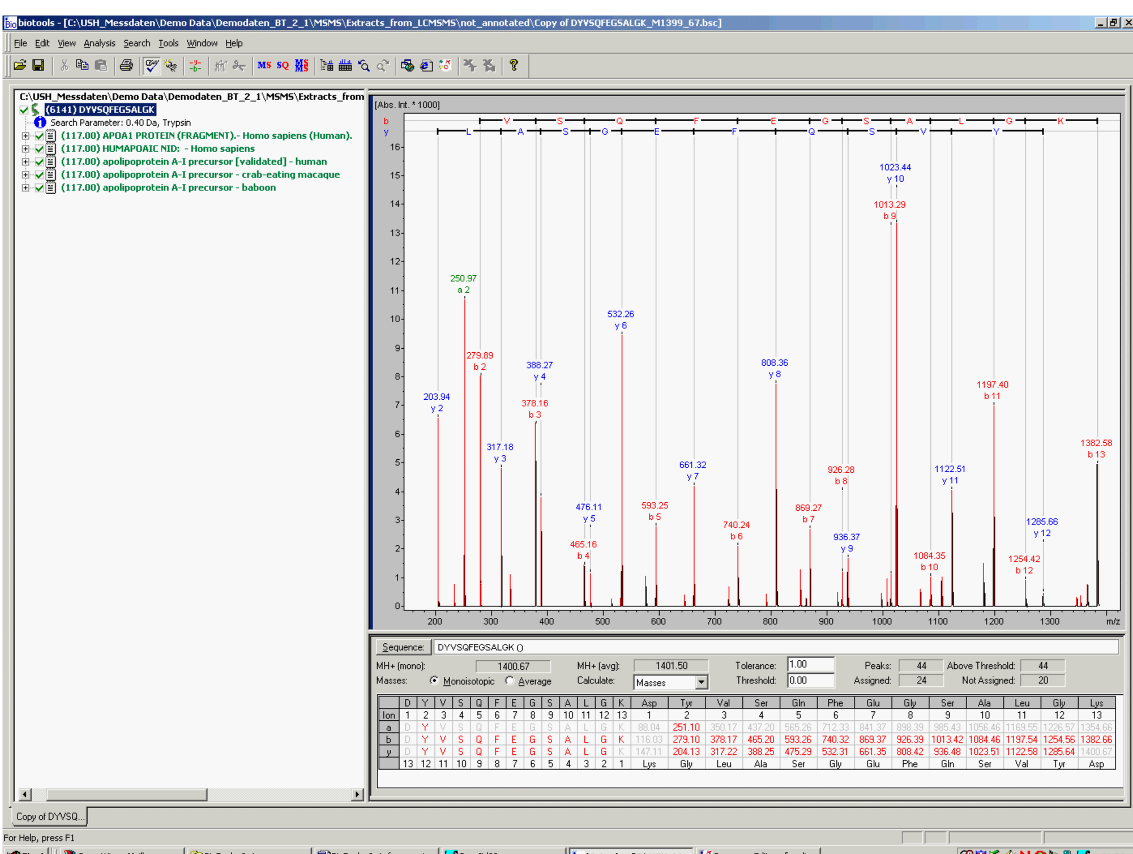


Figure 31

Treeview

The treeview (on the left side of the screen, Figure 31) now shows additional information: the BioTools score, the Mascot score, the search parameter, the resulting protein name(s) and possible modifications. Undesired results can be removed irreversibly by clicking them with the Right mouse button and selecting **DeleteUncheckedSublevel.**

Spectrum window

The Spectrum window (on the upper right side of the screen) shows the annotated MS/MS fragments. The type and the colors of the annotated ions can be changed in the Annotation Options window (Figure 32) which is opened via the menu **Analysis | AnnotationParameters**.

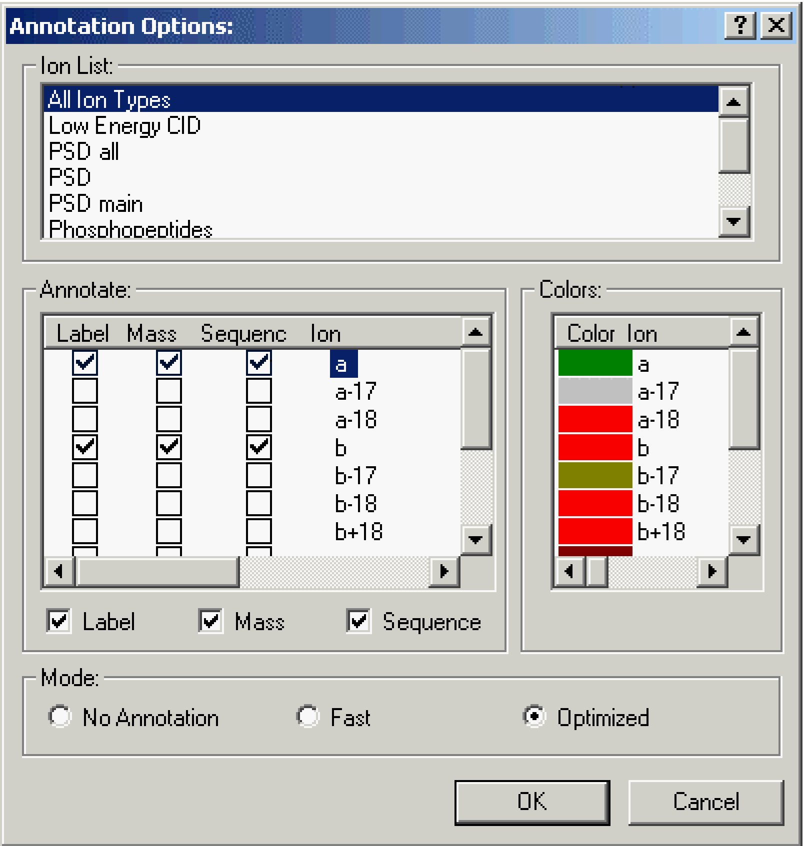
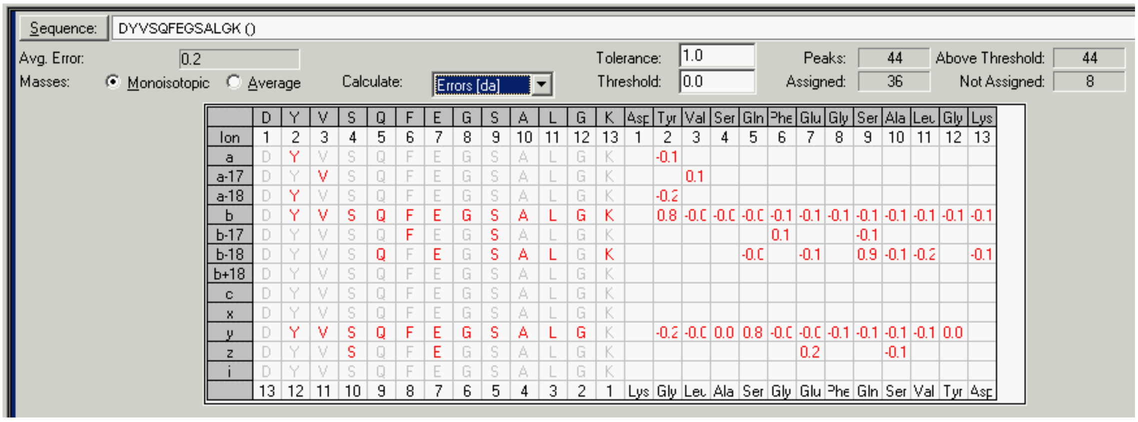


Figure 32

Fragment table

The fragment table (Figure 33) either shows the theoretical masses or the mass differences. The number of the annotated fragments can be influenced in three ways:

- via the threshold. Choose **Analysis | SetThreshold** and set the value by using the left mouse button or by setting the value in the fragment table.
- via the tolerance. It is set in the fragment table window.
- via the admitted fragments. Choose **Analysis | SelectIons** and select the fragments to be shown.



Sequence: DYVSQFEGSALGK ()
Avg. Error: 0.2
Masses: Monoisotopic Average
Calculate: Errors (da)
Tolerance: 1.0
Threshold: 0.0
Peaks: 44
Assigned: 36
Above Threshold: 44
Not Assigned: 8

	D	Y	V	S	Q	F	E	G	S	A	L	G	K	Asp	Tyr	Val	Ser	Gln	Phe	Glu	Gly	Ser	Ala	Leu	Gly	Lys
Ion	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
a	D	Y	V	S	Q	F	E	G	S	A	L	G	K		-0.1											
a-17	D	Y	V	S	Q	F	E	G	S	A	L	G	K			0.1										
a-18	D	Y	V	S	Q	F	E	G	S	A	L	G	K		-0.2											
b	D	Y	V	S	Q	F	E	G	S	A	L	G	K		0.8	-0.0	-0.0	-0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
b-17	D	Y	V	S	Q	F	E	G	S	A	L	G	K						0.1			-0.1				
b-18	D	Y	V	S	Q	F	E	G	S	A	L	G	K					-0.0		-0.1		0.9	-0.1	-0.2		-0.1
b+18	D	Y	V	S	Q	F	E	G	S	A	L	G	K													
c	D	Y	V	S	Q	F	E	G	S	A	L	G	K													
x	D	Y	V	S	Q	F	E	G	S	A	L	G	K													
y	D	Y	V	S	Q	F	E	G	S	A	L	G	K		-0.2	-0.0	0.0	0.8	-0.0	-0.0	-0.1	-0.1	-0.1	-0.1	0.0	
z	D	Y	V	S	Q	F	E	G	S	A	L	G	K							0.2			-0.1			
i	D	Y	V	S	Q	F	E	G	S	A	L	G	K													
	13	12	11	10	9	8	7	6	5	4	3	2	1	Lys	Gly	Leu	Ala	Ser	Gly	Glu	Phe	Gln	Ser	Val	Tyr	Asp

Sequence: DYVSQFEGSALGK ()
MH+ (mono): 1400.7
MH+ (avg): 1401.5
Masses: Monoisotopic Average
Calculate: Masses
Tolerance: 1.0
Threshold: 0.0
Peaks: 44
Assigned: 36
Above Threshold: 44
Not Assigned: 8

	D	Y	V	S	Q	F	E	G	S	A	L	G	K	Asp	Tyr	Val	Ser	Gln	Phe	Glu	Gly	Ser	Ala	Leu	Gly	Lys
Ion	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
a	D	Y	V	S	Q	F	E	G	S	A	L	G	K	88.0	251.1	350.2	437.2	565.3	712.3	841.4	898.4	985.4	1056.5	1169.5	1226.6	1354.7
a-17	D	Y	V	S	Q	F	E	G	S	A	L	G	K	71.0	234.1	333.1	420.2	548.2	695.3	824.3	881.4	968.4	1039.4	1152.5	1209.5	1337.6
a-18	D	Y	V	S	Q	F	E	G	S	A	L	G	K	70.0	233.1	332.2	419.2	547.3	694.3	823.4	880.4	967.4	1038.5	1151.5	1208.6	1336.7
b	D	Y	V	S	Q	F	E	G	S	A	L	G	K	116.0	279.1	378.2	465.2	593.3	740.3	869.4	926.4	1013.4	1084.5	1197.5	1254.6	1382.7
b-17	D	Y	V	S	Q	F	E	G	S	A	L	G	K	99.0	262.1	361.1	448.2	576.2	723.3	852.3	909.4	996.4	1067.4	1180.5	1237.5	1365.6
b-18	D	Y	V	S	Q	F	E	G	S	A	L	G	K	98.0	261.1	360.2	447.2	575.2	722.3	851.4	908.4	995.4	1066.4	1179.5	1236.6	1364.6
b+18	D	Y	V	S	Q	F	E	G	S	A	L	G	K	134.0	297.1	396.2	483.2	611.3	758.3	887.4	944.4	1031.4	1102.5	1215.6	1272.6	1400.7
c	D	Y	V	S	Q	F	E	G	S	A	L	G	K	133.1	296.1	395.2	482.2	610.3	757.4	886.4	943.4	1030.4	1101.5	1214.6	1271.6	1399.7
x	D	Y	V	S	Q	F	E	G	S	A	L	G	K	173.1	230.1	343.2	414.2	501.3	558.3	687.3	834.4	962.5	1049.5	1148.6	1311.6	1426.6
y	D	Y	V	S	Q	F	E	G	S	A	L	G	K	147.1	204.1	317.2	388.3	475.3	532.3	661.4	808.4	936.5	1023.5	1122.6	1285.6	1400.7
z	D	Y	V	S	Q	F	E	G	S	A	L	G	K	130.1	187.1	300.2	371.2	458.3	515.3	644.3	791.4	919.5	1006.5	1105.6	1268.6	1383.6
i	D	Y	V	S	Q	F	E	G	S	A	L	G	K	88.0	136.1	72.1	60.0	101.1	120.1	102.1	30.0	60.0	44.0	86.1	30.0	118.1
	13	12	11	10	9	8	7	6	5	4	3	2	1	Lys	Gly	Leu	Ala	Ser	Gly	Glu	Phe	Gln	Ser	Val	Tyr	Asp

Figure 33

Printing

Use **File | Print** for printing. If a Printing profile (Figure 34) is desired which is not provided this is created in **Tools | Options**. Choose **AddProfile,** type a name and select the desired parameters.

Saving

Upon activating the disk button from the toolbar the Sequence Confirmation result is saved in an additional file (MDI-format).

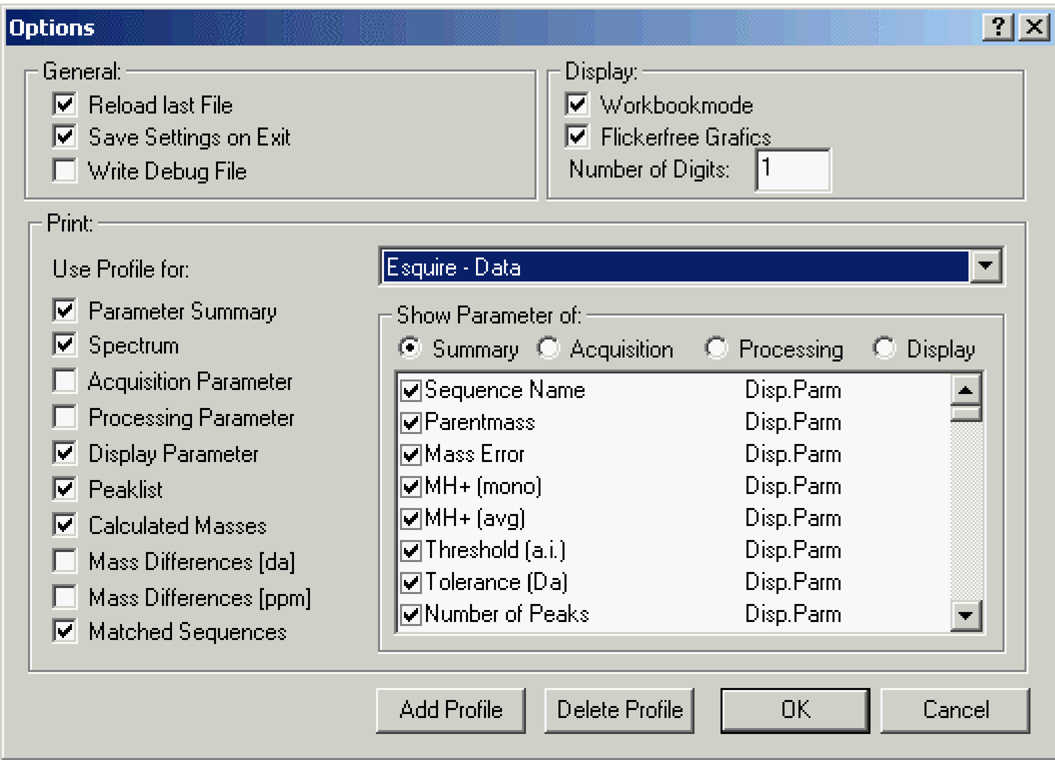


Figure 34

E.3.2. Refined Search

E.3.2.1. Loading a protein sequence in the SequenceEditor

Sequences are opened in the treeview on the left side of the screen. A left mouse button click on the desired filename (apolipoproteina1_human_signal_s.sqs) unfolds the sequences in this file, and those are opened by a left mouse button double click (Figure 35).

If required new sequences can be introduced in the Sequence Editor by opening a new window (**File | New**) and using copy (e.g. from the Database Search result) and paste (**Edit | Paste**) or typing a sequence in the window. The introduction of modifications is described in the Sequence Editor manual.

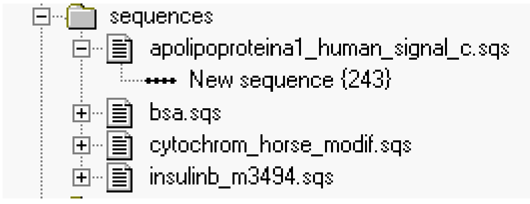


Figure 35

E.3.2.2. Starting the Refined Search

Search for mass
Search parameters
Mr (expt)
1399.600000
Enter mass:
Add mass
Delete mass
Paste
Search keys
Use ion type
M - H
M
M + H
Enzyme
Trypsin
Specific cleavages only
Allow unspecific cleavages
Mass type
Average
Monoisotopic
Modifications
Opt M Oxidation (M)
Error (Da)
0.4
Da
Edit mods
Mass selection
Select all
Select unmatched
Invert
Mutations
Point mutations (proteine)
Point mutations (DNA)
Search
Close
Help

Figure 36

Choose **Sequence | MassSearch** in the Sequence Editor to open the Search for Masses dialogue (Figure 36). If necessary delete the masses in the mass list and type those you are interested in. Set the other parameters as shown in Figure 36 (description in E.1.3.4 Starting the Database Search). In this example, the aim is to search the data for mutations, and therefore, the option "point mutations" is activated. Only one mass can be analysed each time.

Choosing **Search** starts the search, and the calculated proposals are shown in the Mass Search Result window (Figure 37). After **select**ing **all** masses the window will look like this:

Choose **SendToMS/MS** in the Mass Search Result window to send the result to BioTools.

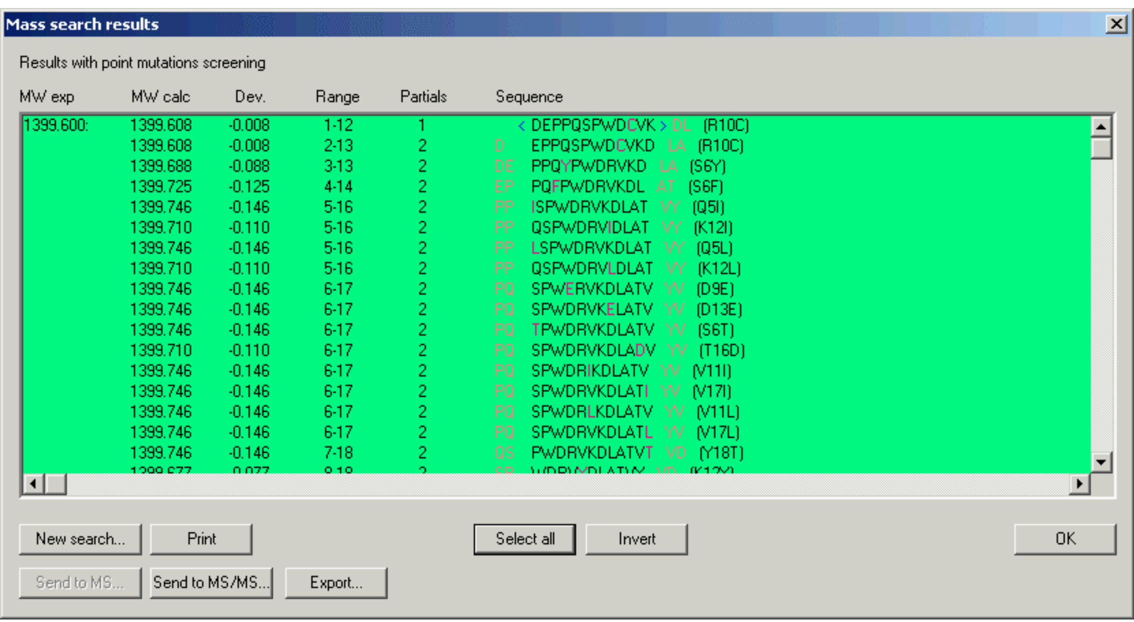


Figure 37

E.3.2.3. The Refined Search result

The treeview (Figure 38) shows all putative mutated peptides which, for one mass, can be derived from the protein sequence. All of them differ from the original sequence by one exchanged amino acid. For all possible peptides the BioTools score is calculated.

In the present example only one mutated peptide with a high BioTools score results. It contains a lysine (K) instead of a glutamine (Q), which cannot be distinguished by the esquire.

Figure 38

The Spectrum window, the Fragment table, printing and saving are described in E.3.1.3 The Database Search result in BioTools.

E.4 Single profile MS-MGF-file

CyC_horse_1p_SPMS.mgf contains the peaklist from CyC_horse_1p_SPMS.d, a single profile file generated horse cytochrome c by offline nanospray. The mgf-file serves as an example for the Database Search and a subsequent Refined Search in BioTools. A general description of the sample preparation, the peptide separation via Hystar, the data acquisition via esquire 5.0, and the data analysis is given in "Protein identification using esquire3000 + (LC-MS/MS data)".

E.4.1. Protein identification

E.4.1.1. Loading the mgf-file

The tutorial "Protein identification using esquire3000 + (LC-MS/MS data)" describes the export of a single profile MS-peaklist as an mgf-file in DataAnalysis. In BioTools use **File | Open** or **File | Find** to open the mgf-file.

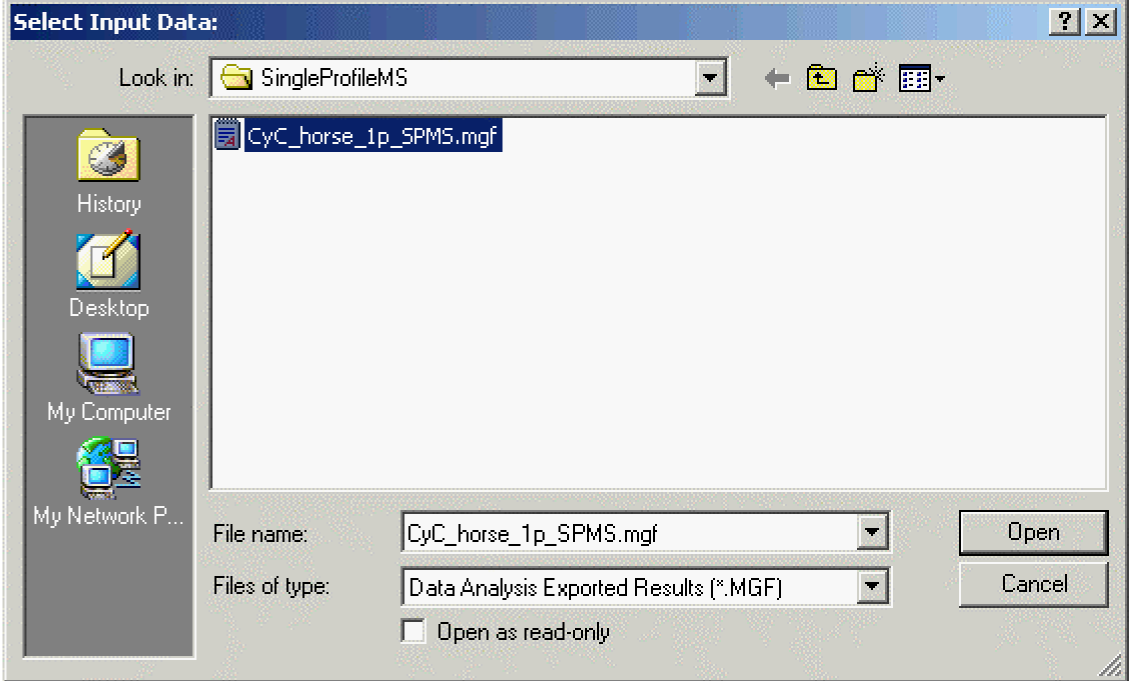


Figure 39

When the location of the mgf-file is known, choose **File | OpenSpectrum**, select the data type (which is mgf) and the location of the file (Figure 39). Choose the file with the left mouse button, For loading multiple files use the CONTROL or SHIFT button and the left mouse button. Then **Open** the file.

When the location of the mgf-file is not known choose **File | Find**, select the data type (which is mgf) and the folder where the files are located (Figure 40). Upon selecting **FindNow**, this folder is searched for mgf-files which are presented as a list. Choose the file with the left mouse button, for multiple files use the CONTROL or SHIFT button and the left mouse button. Then, **Open** the file(s).

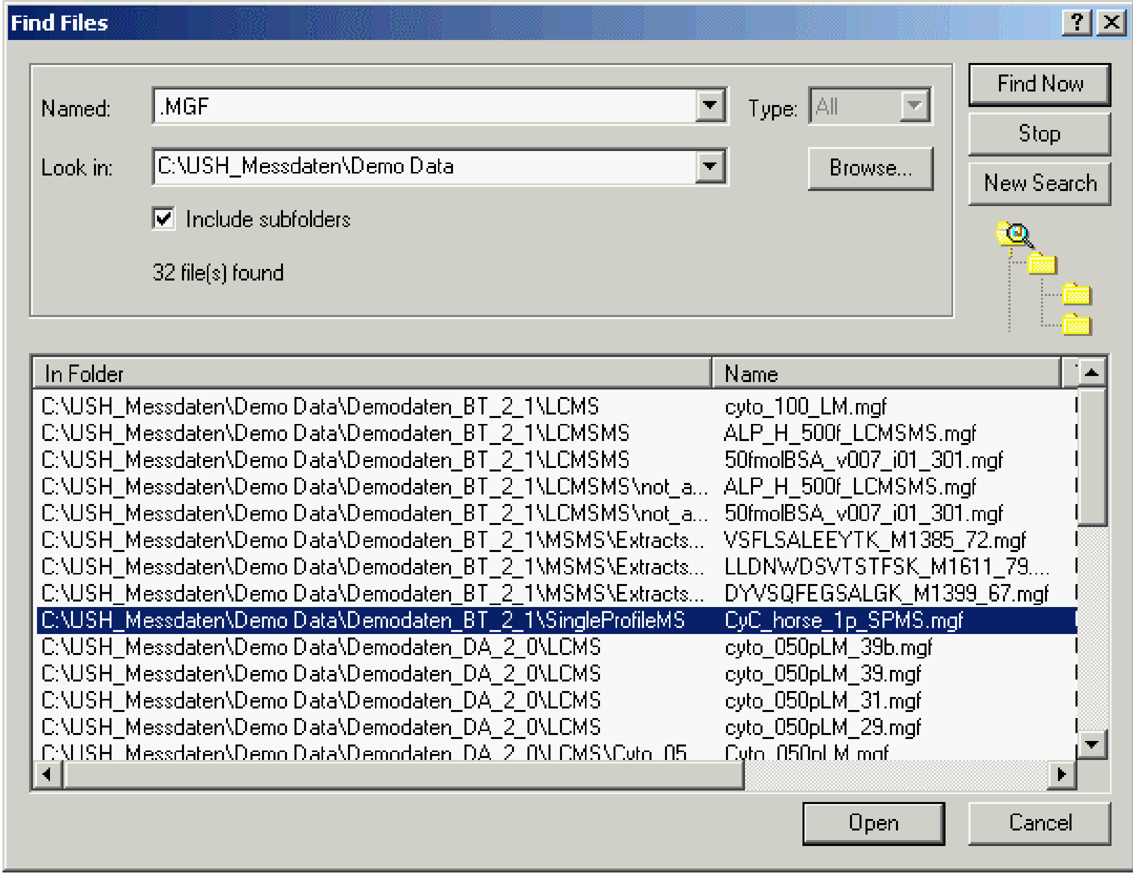


Figure 40

In any case, the spectrum is shown (Figure 41).

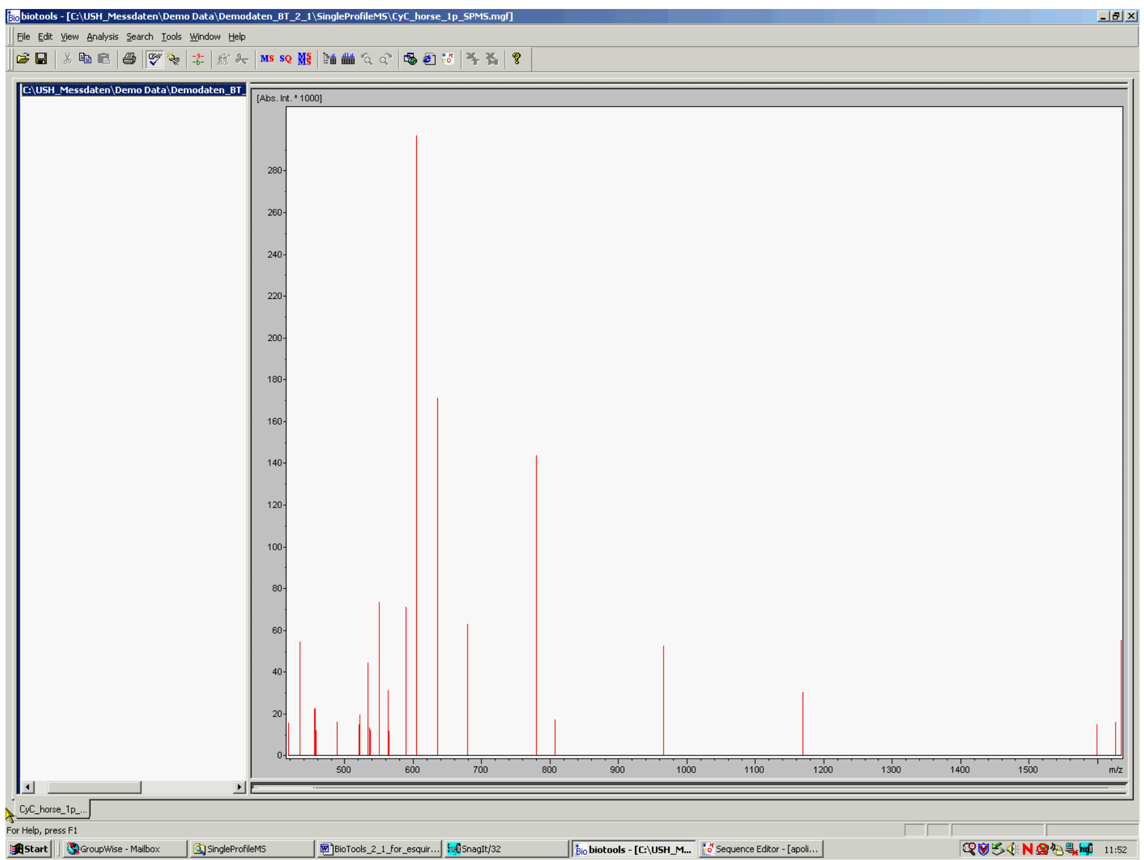


Figure 41

In the treeview on the left side of the screen the file name and the location are shown.

E.4.1.2. Starting the Database Search

In this chapter protein identification via BioTools without the Batch mode (refer to "Protein identification using esquire3000 + (LC-MS/MS data)") is shown. Choose **Search | MascotPeptideMassFingerprint** to open the Parameter window, and set the parameters as shown in Figure 42.

Start is used to start the search. The intermediate window (Figure 43) shows the progress of the search.

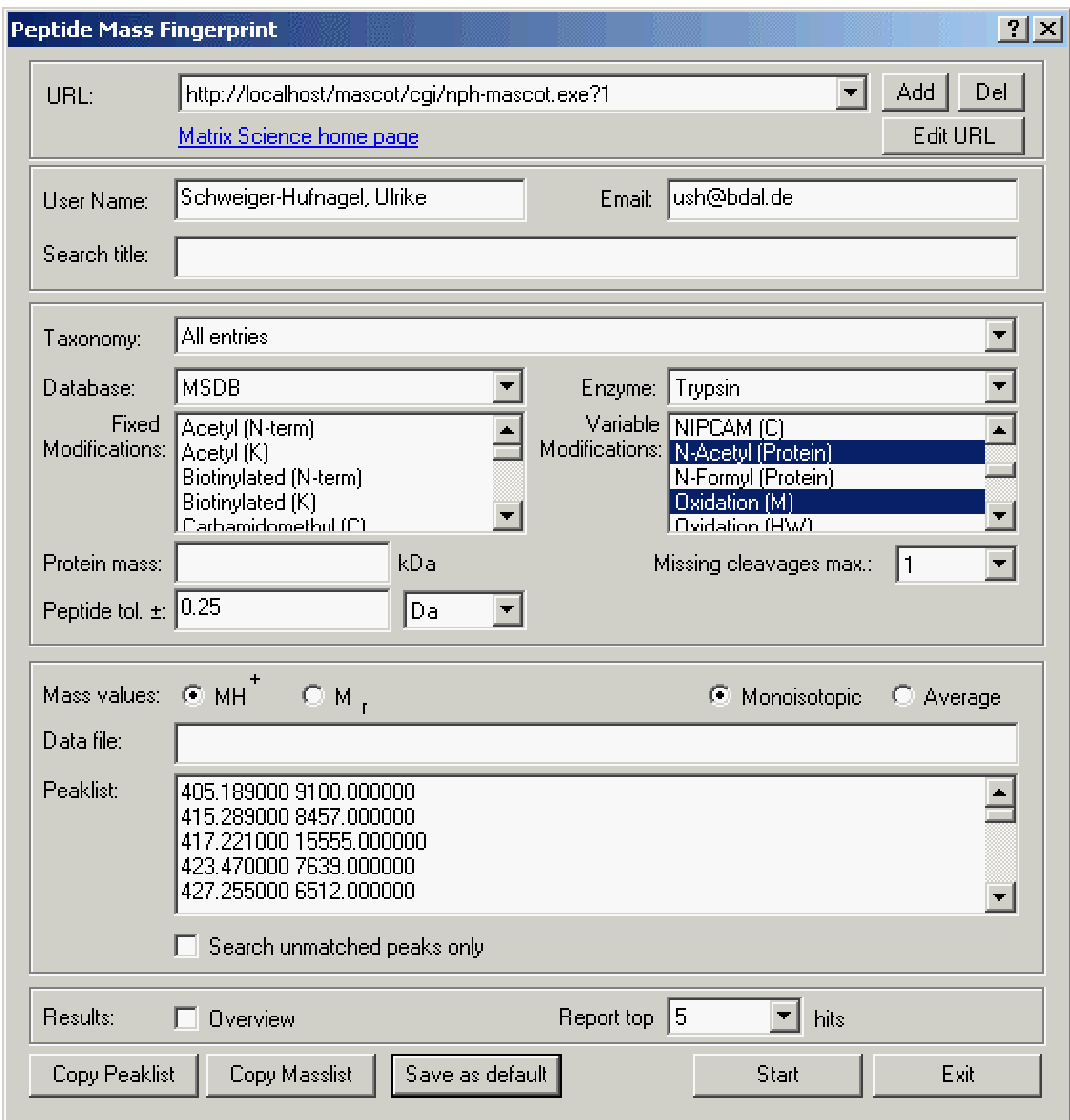


Figure 42

Upon choosing **GetHits** and the desired result number(s), e.g. number 1 or 2 or 1-2, the sequence information is sent to the peaklist in BioTools.

Then comes the search result (Protein Summary Report, Figure 44).

{MATRIX SCIENCE} **Mascot Search**

Licensed to: Bruker Daltonik, (2 processors).
Finished uploading search details...
IMPORTANT: If you get disconnected or choose not to wait for your search results
DO NOT RESUBMIT THE SEARCH. Your results will be sent by email when the search is complete
Searching....
.....5% complete
......10% complete
.......15% complete
...

Figure 43

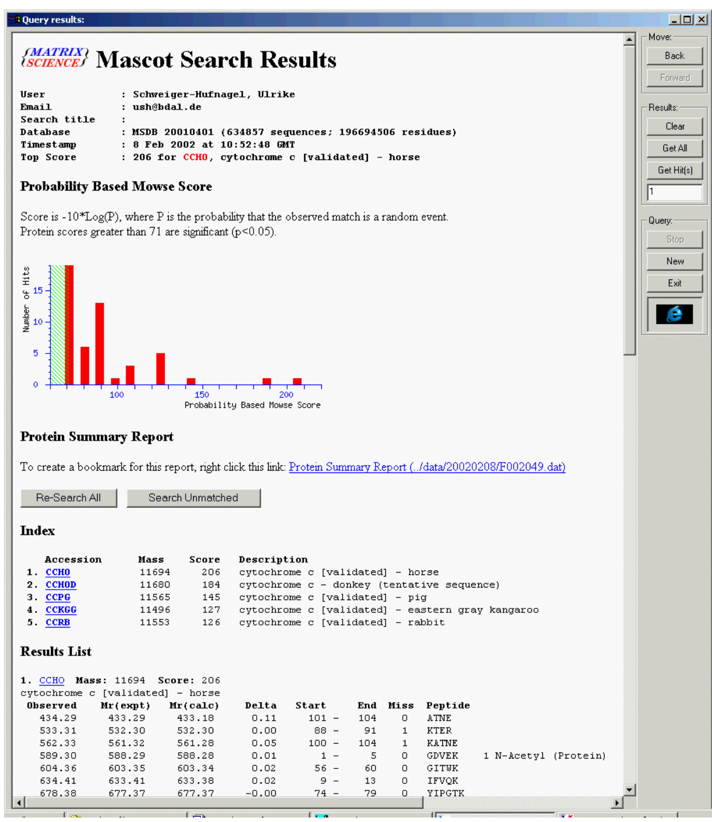


Figure 44

E.4.1.3. The Database Search result in BioTools

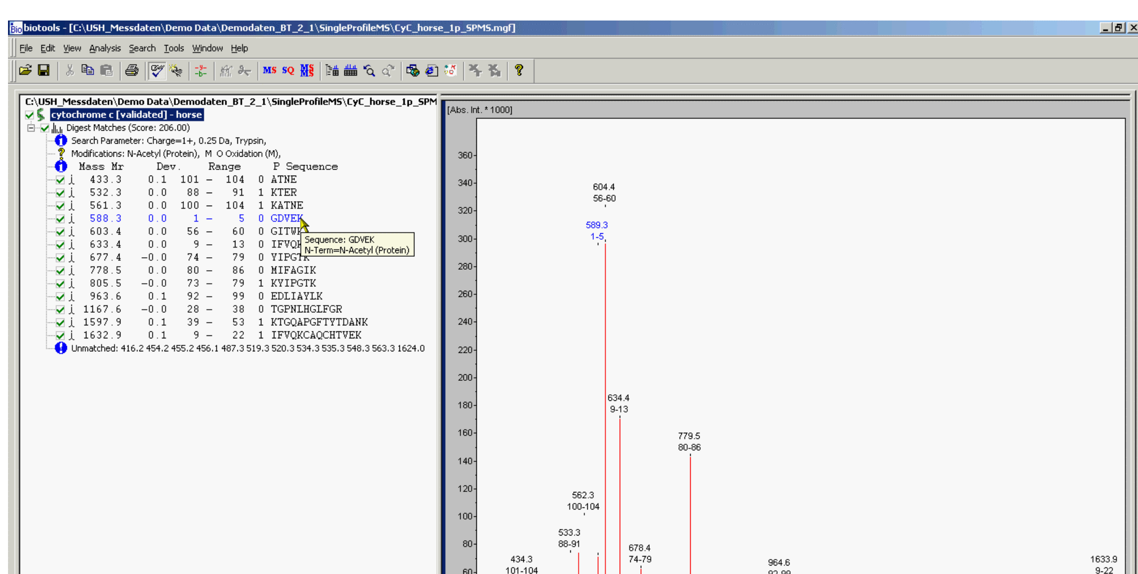


Figure 45

Treeview

The treeview (on the left side of the screen, Figure 45) now shows additional information: the Mascot score, the search parameter, the resulting protein name(s), modifications (blue font for modified peptides) and a result table with the following content:

- measured molecular mass
- mass deviation
- amino acid range
- partials (missing cleavages)
- peptide sequence.

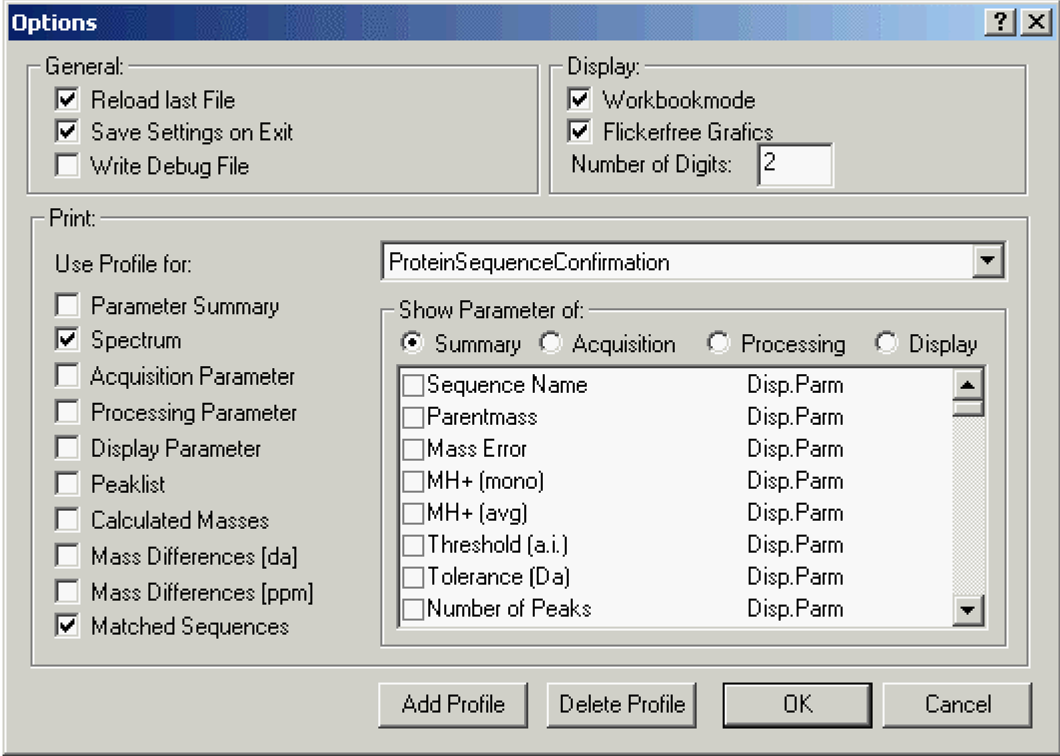


Figure 46

Undesired results can be removed irreversibly by unchecking them, clicking them with the Right mouse button and selecting **DeleteUncheckedSublevel.**

Spectrum window

The Spectrum window (on the upper right side of the screen) shows the matched and unmatched signals when **View | MatchedAndUnmatchedPeaks** and the "Digest Masses" in the treeview are activated.

Sequence Viewer window

The Sequence Viewer window contains three parameters:

- The intensity coverage
- The Sequence coverage for the MS signals
- The isoelectric point of the protein.

Gray bars below the protein sequence show the sequence coverage.

Connection between the result window parts

Upon highlighting one mass in the treeview the corresponding signal is labeled in the mass spectrum (mass and amino acid range) in the Spectrum window and the sequence is also highlighted green in the Sequence Viewer window. Vice versa, when a peptide is activated by clicking on the gray bar in the Sequence Viewer window it is also labeled in the Spectrum window and highlighted in the treeview.

Printing

For printing the result an appropriate Printing profile (Figure 46) is required, and this is created in **Tools | Options**. Choose **AddProfile** and type in a desired name. Select "Spectrum" and "Matched Sequences" as parameters. Then, the spectrum, the mgf-peaklist and the result list can be printed upon choosing **File | Print**.

Saving

Upon activating the disk button from the toolbar the Sequence Confirmation result is saved in an additional file (MDI-format).

E.4.2. Refined Search

E.4.2.1. Starting the Refined Search

In BioTools choosing **Search | SearchForMasses** after activating "digest masses" or double-clicking "digest masses" in the treeview starts the Sequence Confirmation tool. Then, the Search For Masses window (Figure 47) opens. The following parameters can be set:

Parameter	Explanation / Recommendation
Monoisotopic/ Average	Monoisotopic masses should be selected.
Error	Select the unit and the value.
Enzyme	Choose the enzyme that has been used for the digest.
Specific/unspecific cleavages	For both options an unlimited number of missing cleavages (partials) is analyzed.
Modifications	Upon selecting **EditMods** in the Search For Masses window (Figure 47) modifications can be analyzed. Then, the

Modification Editor window (Figure 48) is opened. Choose one residue type or "all" and activate the desired modifications (left mouse button). Select the modification type and **Add** the modification to the list on the left side of the window. **Edit type** enables the user to define modifications. **Ok** sends the selected modifications back to the Search for Masses window.

Mass Selection — Select the masses that should be used for the search.

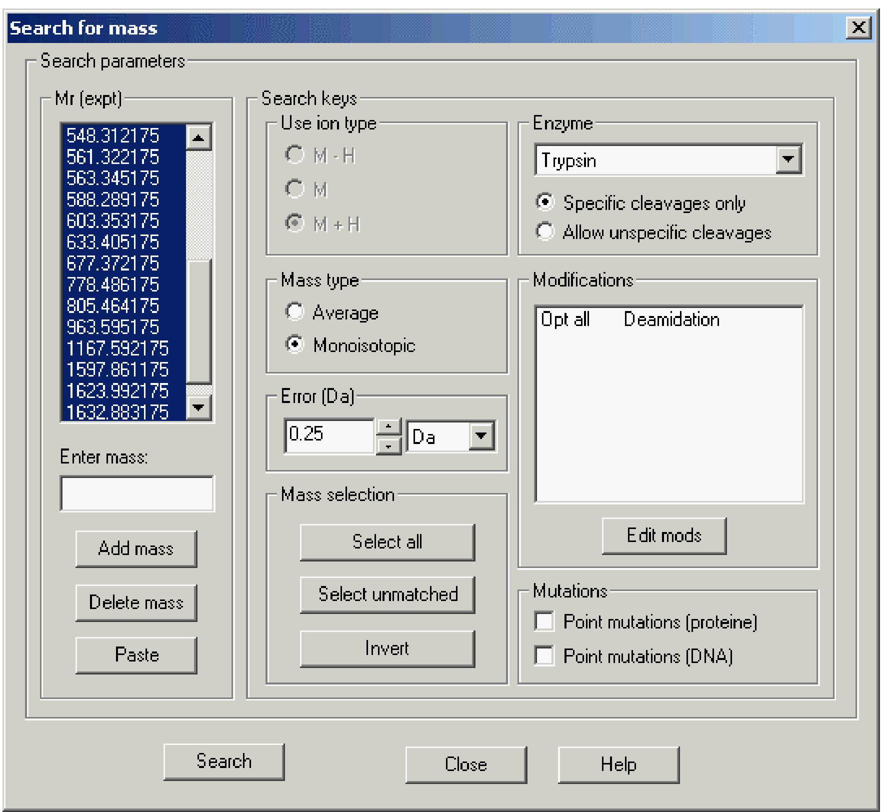


Figure 47

For the example file deamidation, a modification which is not provided for the Mascot search, was chosen as an optional modification as shown in Figure 48.

Choosing **Search** in the search for masses window starts the search, and the calculated proposals are shown in the Mass Search Result window. After **select**ing **all** masses the window will look as shown in Figure 49.

Choose **SendToMS** in the Mass Search Result window to send the result to BioTools.

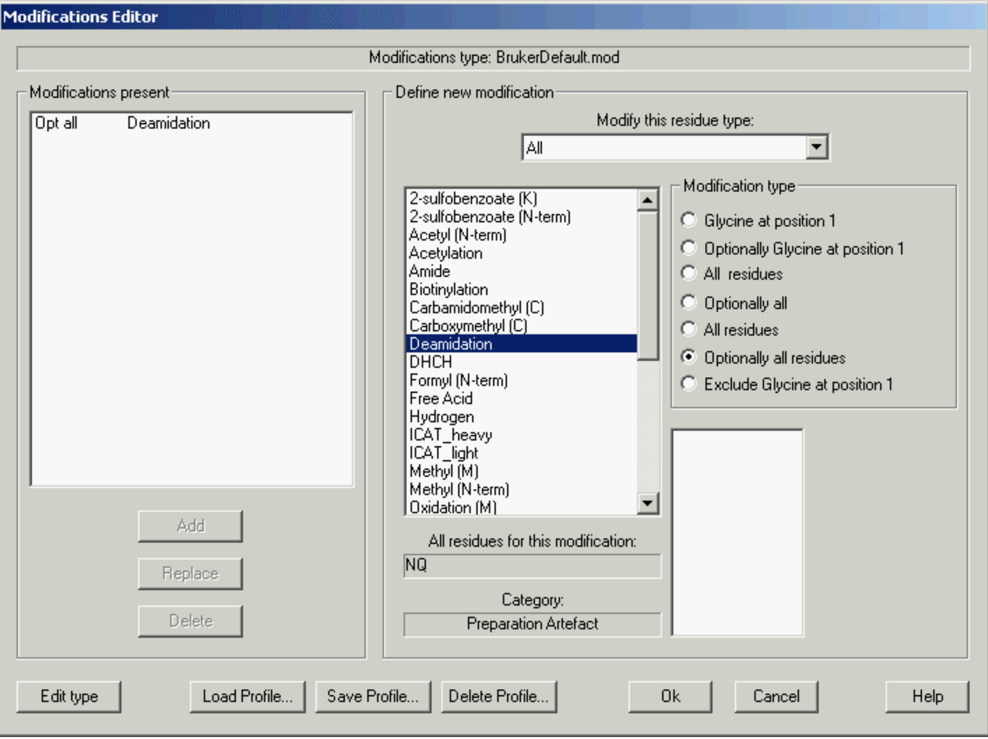


Figure 48

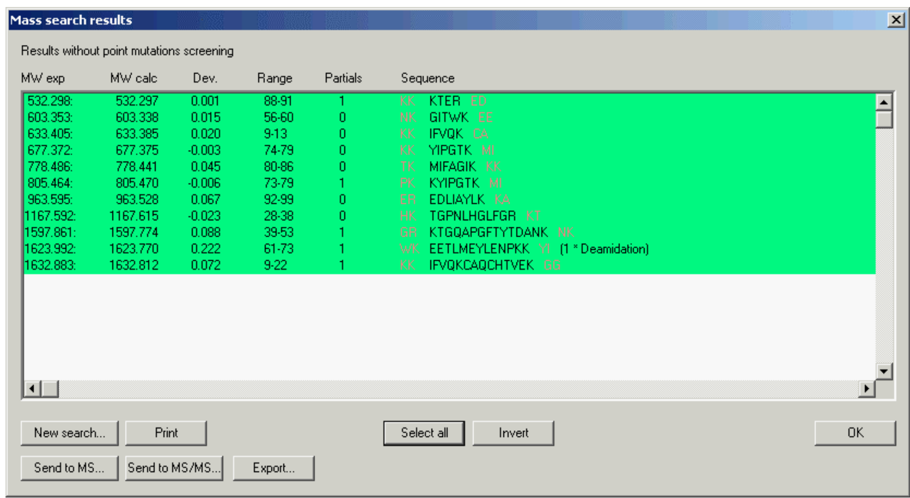


Figure 49

E.4.2.2. The Refined Search result

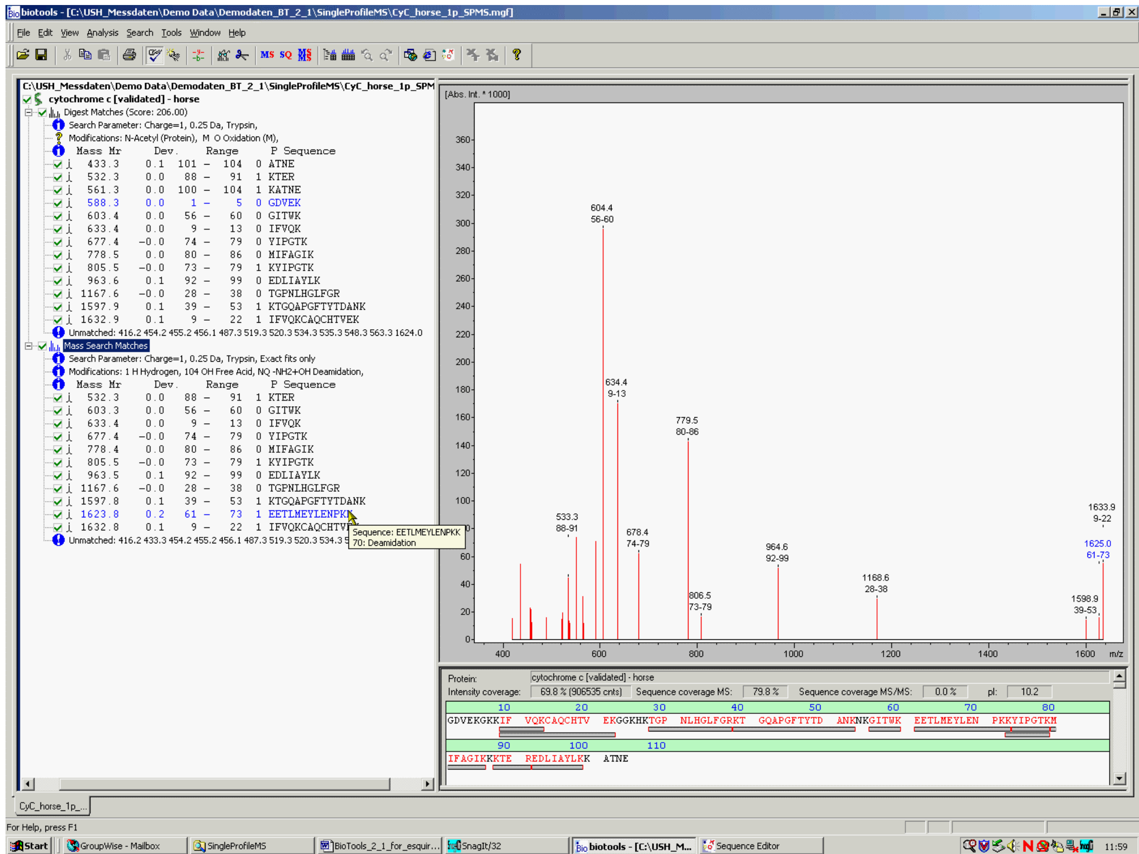


Figure 50

Selecting results to be shown

Either the Database Search result or the Refined Search result (Figure 50) are shown, or both results are combined (Figure 51) upon activating the protein name in the treeview which is "cytochrome" in this example.

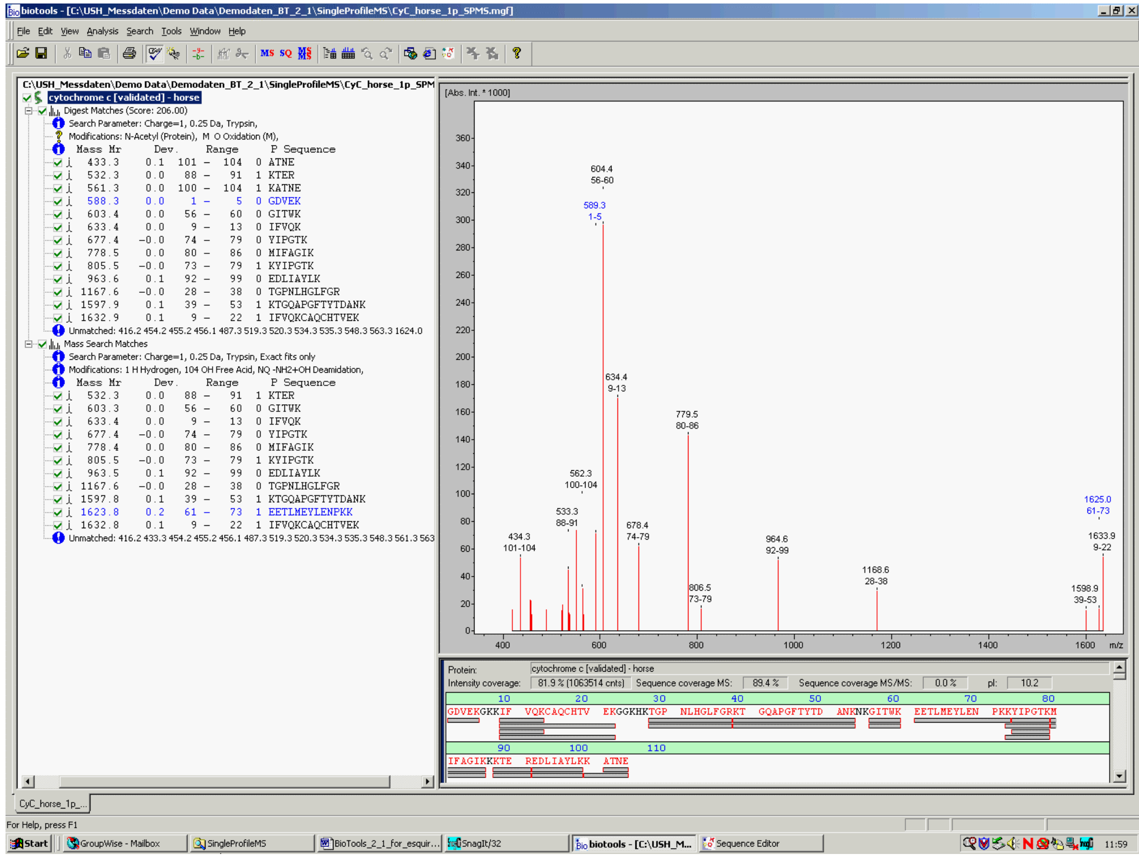


Figure 51

The treeview, the Spectrum window, the Sequence Viewer, printing and saving have been described in E.4.1.3 The Database Search result in BioTools. But one difference exists to the treeview described before: here, in the Refined Search, the treeview shows the calculated neutral molecular masses and not the measured ones!

D DeNovo Sequencing with BioTools

D.1 Introduction D-2
D.1.1 Why DeNovo Sequencing? D-2
D.1.2 Information Content of MS/MS Spectra D-2
D.2 DeNovo Sequencing with esquire3000 + Data D-3
D.2.1 Transfer of Peaklists from DataAnalysis 3.1 to BioTools 2.2 .... D-3
D.2.1.1. Data Preparation in DataAnalysis D-3
D.2.1.2. New Export Parameter in DataAnalysis 3.1: "Normalize MS(n)-Data" ... D-3
D.2.1.3. Activating the Compound in DataAnalysis D-4
D.2.1.4. Direct Transfer of the Peaklist of one LC-MS/MS Compound D-4
D.2.1.5. Export of the Peaklist from one LC-MS/MS Compound (mgf-format) D-4
D.2.1.6. Using peaklists created by DataAnalysis 3.0 and previous versions D-7
D.2.2 DeNovo Sequencing D-8
D.2.2.1. DeNovo Sequencing for the Urgent Reader D-8
D.2.2.2. General Procedures for DeNovo Sequencing D-9
D.2.2.3. Useful Suggestions D-17
D.2.2.4. Excursion to a BLAST Search D-28
D.3 DeNovo sequencing D-29
D.3.1 General Description D-29
D.3.2 Processing in BioTools D-30
D.3.2.1. LIFT-TOF/TOF MS Analysis of a Tryptic Peptide D-30
D.3.2.2. Combining DeNovo Sequencing with Homology Searching D-38

D.1 Introduction

D.1.1 Why DeNovo Sequencing?

In many instances, the protein identification based on MS/MS spectra of a proteolytic peptide fails. Amongst the reasons, why such an analyses may have failed are simple technical problems, such as

- Mass errors larger than those allowed in a database search
- Peptide nature different from expectation due to, e.g., unspecific cleavages
- Variation of search conditions, recalibration or the combination of MS/MS data of several peptides

Beyond these more trivial reasons, there are fundamental aspects causing library searches to fail such as:

- Post-translational modifications of the peptide
- Sequence errors in the database, which is particularly true with EST databases
- Sequence of the protein of interest is not contained in the database. This situation occurs quite frequently while working on organisms with genomes, which have not been sequenced yet.

In all these cases, the extraction of sequence information from MS/MS spectra or fragment ion spectra including higher orders (MS^n) without any prior knowledge about the sequence, i.e., *DeNovo* sequencing, allows to elucidate the structure and sequence of a particular peptide. Such peptide sequences can then be used to identify a protein solely by homology searches such as BLAST.

D.1.2 Information Content of MS/MS Spectra

All MS/MS spectra of peptides contain fragment ions, which result either from single cleavages of peptide backbone bonds and multiple cleavages (see Appendix, Peptide Fragmentation). Electrospray spectra typically contain y and b ions and MALDI spectra additionally contain a and i internal fragments and neutral loss fragments such as b-NH3, etc. High energy CID spectra, as they can be obtained using the CID-LIFT option on the ultraflex-TOF/TOF, can additionally provide side chain fragmentation products for particular amino acid residues. Tryptic peptides typically provide w ions in such spectra, which allow distinguishing between the isobaric amino acid residues L and I due to a different neutral loss from the side chain.

D.2 DeNovo Sequencing with esquire3000 + Data

D.2.1 Transfer of Peaklists from DataAnalysis 3.1 to BioTools 2.2

Enolase_MHplus_1755_96_MS2MS3.mgf is an example mgf-file used to demonstrate the *DeNovo* tool provided by BioTools 2.2.

D.2.1.1. Data Preparation in DataAnalysis

In DataAnalysis, FindCompounds, MassListFind and Deconvolution can be performed as described in esquire application tutorial "Protein Identification of LC-MS/MS Data".

D.2.1.2. New Export Parameter in DataAnalysis 3.1: "Normalize MS(n)-Data"

In **Method | Parameter | MascotExportOptions** (DataAnalysis 3.1) "Normalize MS(n)-Data" can be activated. This option refers to the intensity of signals in the fragment spectra (MS(≥3)) which is lower than the intensity of the original spectrum. When it 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.

For *DeNovo* sequencing it is recommended to use "Normalize MS(n)-Data". Then, the intensity of the MS(3)-signals increases, and those signals do not get "lost". But before exporting the peaklist (see below), all MS(≥3)-spectra containing only noise should be deleted because otherwise the noise peaks are amplified.

D.2.1.3. Activating the Compound in DataAnalysis

Then, the compound of interest should be activated in the treeview as shown in Figure 1.

ENOstenhMS3.d
Chromatograms
Compound Mass Spectra
Cmpd 1, MSn() 2.7 min
+MS, 1.0min (#5)
+MS2(878.91), 1.5min (#6)
+MS2(878.91->*), 2.3-4.4min (#7-#10)
Cmpd 2, MSn() 6.4 min
Cmpd 3, MSn() 10.1 min
Cmpd 4, MSn() 13.8 min
Cmpd 5, MSn() 17.5 min
Cmpd 6, MSn() 21.2 min
Cmpd 7, MSn() 24.9 min
Cmpd 8, MSn() 28.6 min
Cmpd 9, MSn() 32.3 min
Cmpd 10, MSn() 36.0 min
+MS, 34.3min (#59)
+MS2(878.52), 34.9min (#60)
+MS2(878.52->*), 35.6-37.8min (#61-#64)
Cmpd 11, MSn() 39.8 min
Cmpd 12, MSn() 43.5 min
Cmpd 13, MSn() 47.2 min
Cmpd 14, MSn() 50.9 min
Cmpd 15, MSn() 54.6 min

Figure 1

D.2.1.4. Direct Transfer of the Peaklist of one LC-MS/MS Compound

Automatic transfer of the compound of interest from DataAnalysis to BioTools is completed via **Tools | BioTools** or via the toolbar Bio-button.

D.2.1.5. Export of the Peaklist from one LC-MS/MS Compound (mgf-format)

The peaklist can be exported as an mgf-file via **File | Export | Compounds.**

In BioTools, use **File | Open** or **File | Find** to open the mgf-file which contains the peaklist of one compound.

When the location of the mgf-file is known, choose **File | Open Spectrum**, select mgf as data type and the location of the file (Figure 2). Select the file with the left mouse button. To load multiple files, use CONTROL or SHIFT key and highlight the files with left mouse button. Then **Open** the file(s).

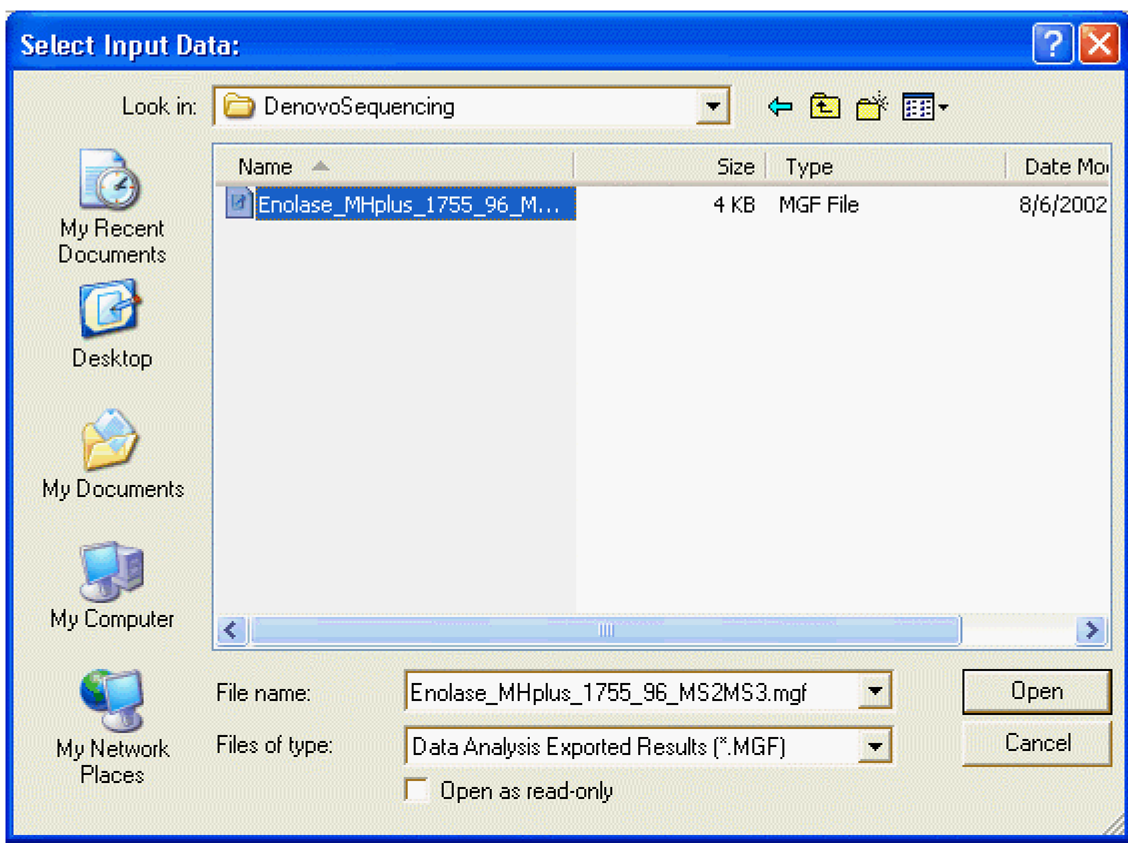


Figure 2

When the location of the mgf-file is not known, choose **File | Find**, select mgf as data type and the folder where the file is located (Figure 3). Upon selecting **FindNow**, this folder is searched for all mgf-files which are presented. Choose the file with the left mouse button. To select multiple files, use CONTROL or SHIFT key and highlight the files with left mouse button. Then, **Open** the file(s).

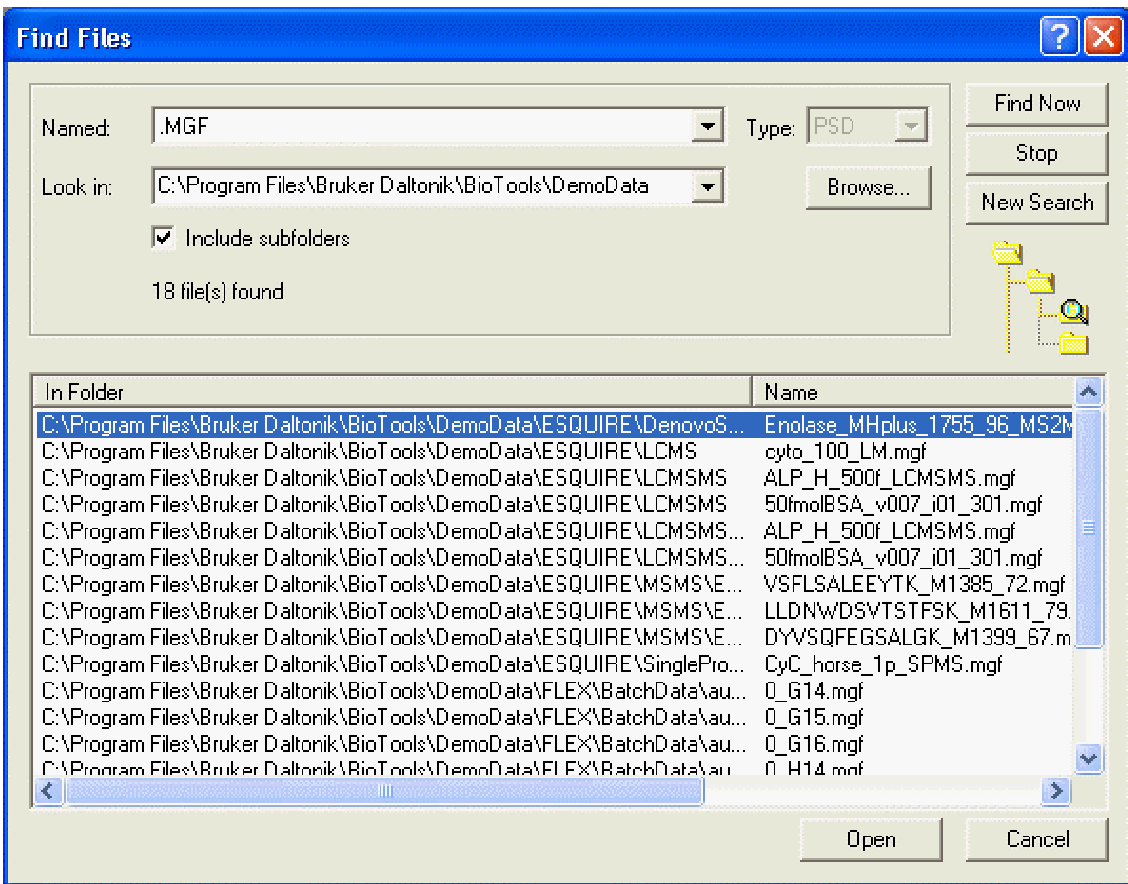


Figure 3

In any case, the Enolase_MHplus_1755_96_MS2MS3.mgf MS(n) histogram spectrum is shown (Figure 4). Biotools recognizes, that the mgf-file contains the peaklist of only one compound.

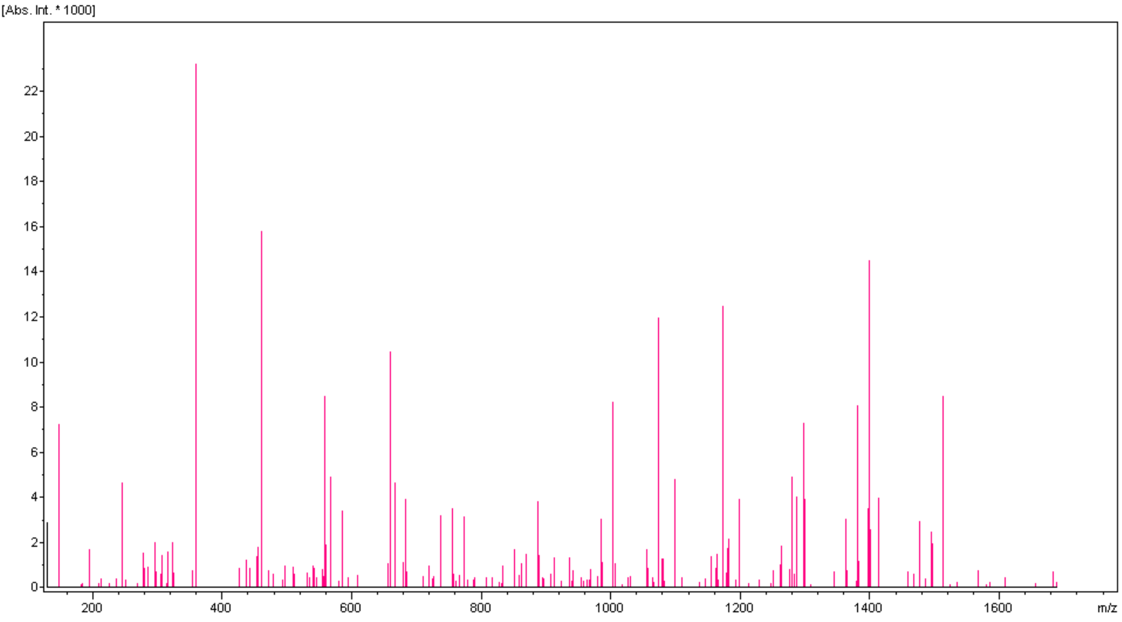


Figure 4

In the treeview on the left side of the screen the file name and its location are shown.

D.2.1.6. Using peaklists created by DataAnalysis 3.0 and previous versions

Peaklists from previous DataAnalysis versions can be used for *DeNovo* sequencing as well. However, only bsc-files can be used which requires that only deconvoluted data (or only non-deconvoluted data) be exported. Furthermore, it is not possible to limit the number of exported peaks. To use *DeNovo* sequencing, DataAnalysis 3.1 is highly recommended.

If DataAnalysis 3.1 is not available please refer to "using BioTools for esquire series data" in the BioTools 2.1 tutorials, which describe the bsc-file handling.

D.2.2 DeNovo Sequencing

D.2.2.1. DeNovo Sequencing for the Urgent Reader

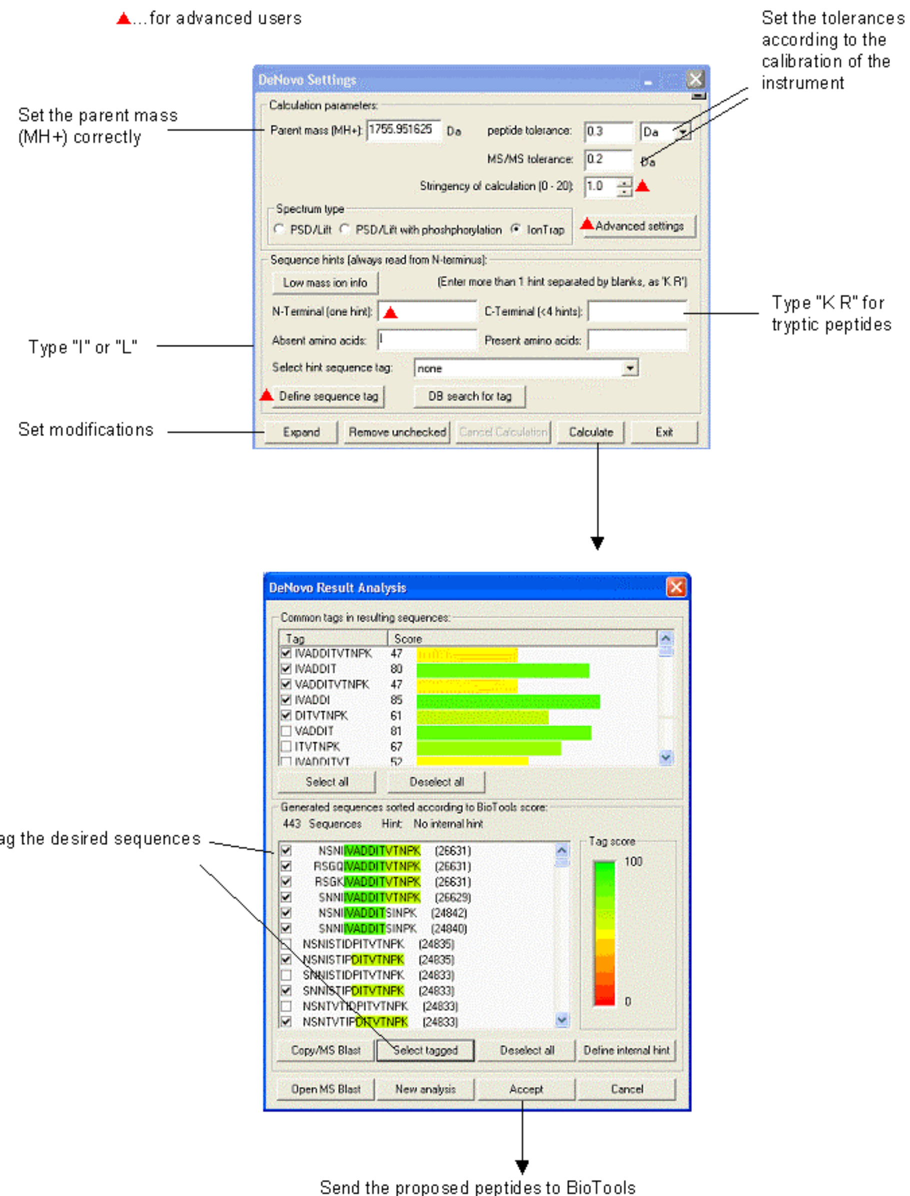


Figure 5

Figure 5 shows how basic parameters are set in the *DeNovo* Settings window (**Analysis | Full DeNovo Sequencing** in BioTools 2.2). The use of terminal or internal hints and other advanced parameters is described in sections F.2.2.2 General procedures for DeNovo Sequencing and F.2.2.3 Useful Suggestions. In *DeNovo* Result Analysis window, peptide sequences can be pre-selected (check-marked) and sent to BioTools subsequently by clicking on **Accept**.

D.2.2.2. General Procedures for DeNovo Sequencing

D.2.2.2.1. DeNovo Settings Window

Choose **Analysis | Full DeNovo Sequencing** in BioTools 2.2 to open *DeNovo* Settings window, and the parameters described below can be set. For the data file Enolase_MHplus_1755_96_MS2MS3.mgf the parameters shown in Figure 6 were used.

D.2.2.2.2. Explanation and Recommendations for Parameters and Operations

Parameters and Operations	Explanation and Recommendations
Parent Mass field	Handles the information about the fragmented precursor ion m/z and the deconvolution (priority 1) or the Global charge limitation (export parameter in Data Analysis ≥ 3.0, priority 2). This means that: When deconvolution was performed successfully the resulting charge and the m/z are used to calculate MH^+. When no deconvolution was performed the charge from the Global charge limitation and the m/z are used to calculate MH^+. When no deconvolution was performed and no Global charge limitation was set the Parent Mass field contains the fragmented m/z . It is important to ensure that the correct MH^+ is entered. This value can be changed manually if necessary.
Peptide tolerance	Mass accuracy obtained from the peptide (precursor) mass measurement. Depends on the calibration of the respective instrument (TOF, Ion Trap, FTMS)
MS/MS tolerance	Mass accuracy obtained for the fragment masses. Depends on the calibration of the respective instrument (TOF, Ion Trap, FTMS)

Parameters and Operations	Explanation and Recommendations
Stringency of calculation	The stringency of calculation is a threshold value to describe the match between spectrum peaks and a putative partial amino acid sequence. If stringency is set low enough, sequences are returned even if very few peaks are matched. A higher value for this parameter may increase the speed of the calculation however the number of generated partial sequences will be reduced as the condition for acceptance becomes more stringent. A **default value of 5** is appropriate in most cases. Sometimes lower values are necessary to move the algorithm successfully past the initial stage, especially when there are few peaks in the spectrum at low masses.
Spectrum type	The instrument type is recognized automatically. For D*eNovo* calculation internal scoring is optimized for each spectrum type. A sequencing calculation for a phosphorylated peptide must be initiated by choosing the respective center option.
Advanced settings	The number of partial sequence candidates can be chosen (Figure 7). This parameter defines is the number of sequence tags used for the calculation of peptide sequences. A high number of sequencing candidates slows down the speed of calculation but returns a greater variety of resulting sequences. Values between 0 and 200 are possible. A default setting 100 is appropriate in most cases.
Low mass info	**Click this button to determine the presence of amino acids** by inspecting the low mass region of the PSD/LID MS/MS data for specific immonium ions. The result of this search is displayed automatically in the boxes "Absent" and "Present". N.B. Useful for MALDI TOF MS/MS data only. Not applicable to esquire ion trap data.
N-terminal	If already known, the N-terminus can be keyed in here, either as a partial sequence (e.g. "HLNI" for the peptide HLNITGR) or just one amino acid.
C-terminal	Supply any information known about the C – terminus. (i.e. **Type in "K R"** (including a blank) for peptides from a tryptic digest. Up to 4 C-termini are possible.)
Absent amino acids	Since Ile and Leu have the same mass, one of them should be entered in this field. Gln and Lys have similar masses (difference 0,034), and one of them can be typed here. But take care for peptides with multiple Lys and Gln, when the MS/MS tolerance is set to e.g. 0.1.
Present amino acids	Irrelevant for esquire data

Parameters and Operations	**Explanation and Recommendations**
Select internal hint sequence tag	When multiple internal hints have been defined in the Define Internal Hint window (Figure 8), one entry can be selected for the calculation.
Define sequence tag	A sequence tag can be typed manually as shown in Figure 8. Choose "Define manually". Type the lower mass of the tag. Type the tag sequence from the left to the right side of the spectrum (see example in F.2.2.3.3, Defining the Internal Hint Manually). When more intense peaks have been chosen in esquire spectra it is recommended to choose "y-ions". Then click **OK**.
DB search for tags	The manually defined sequence tag can be selected in the pull-down window and used for a sequence tag search as described in F.2.2.3.3, Sequence Tag Search.
Expand	The *DeNovo* Setting Parameter window is expanded (Figure 9), and modifications can be selected. This is reversed by **SHRINK**.
Remove Unchecked	Unchecked entries are removed from the BioTools treeview.
Cancel Calculation	The calculation is canceled.
Calculate	Used to start the calculation.
Exit	The dialogue window is closed.

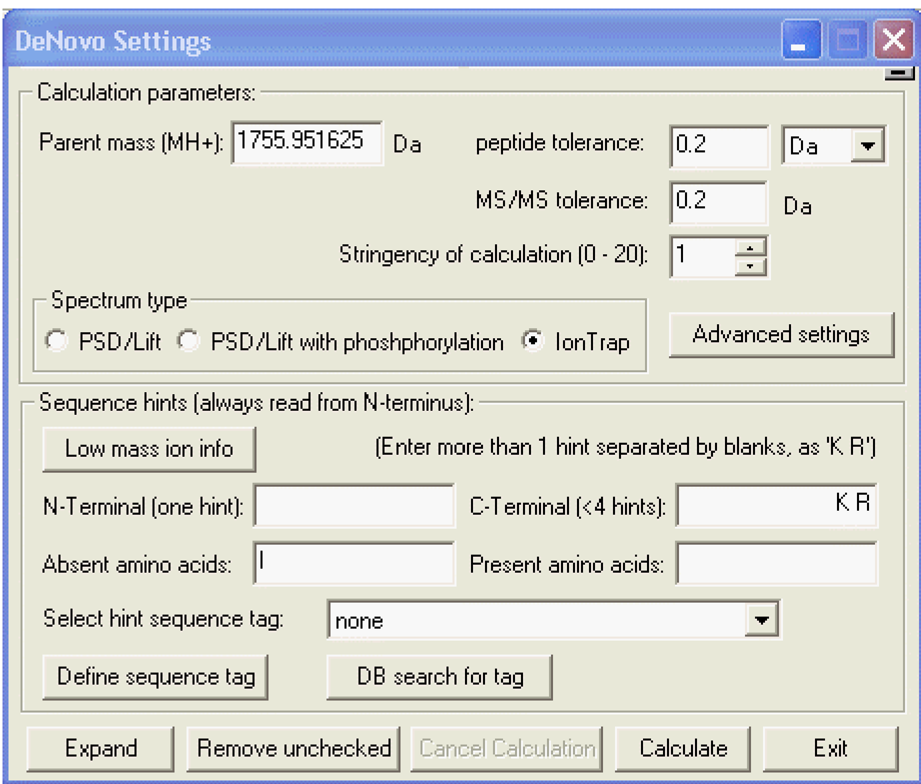


Figure 6

Advanced Settings

Number of partial sequence candidates: 100

OK Cancel

Figure 7

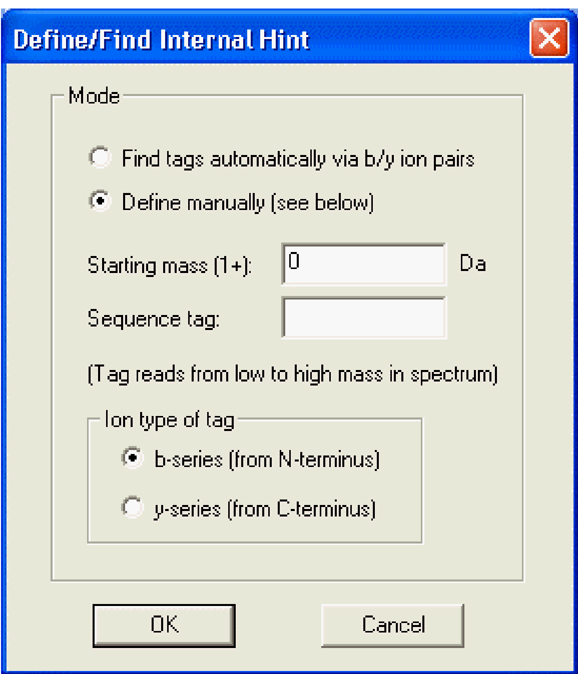


Figure 8

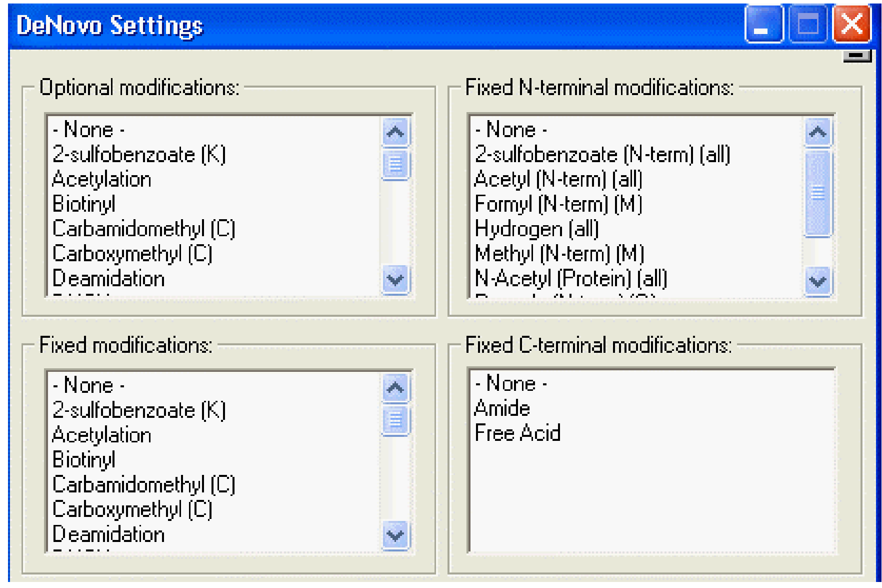


Figure 9

D.2.2.2.3. DeNovo Result Analysis Window

An accomplished DeNovo calulation will open the *DeNovo* Result Analysis window as shown in Figure 10 for Enolase_MHplus_1755_96_MS2MS3.mgf, using the parameter set shown in Figure 6.

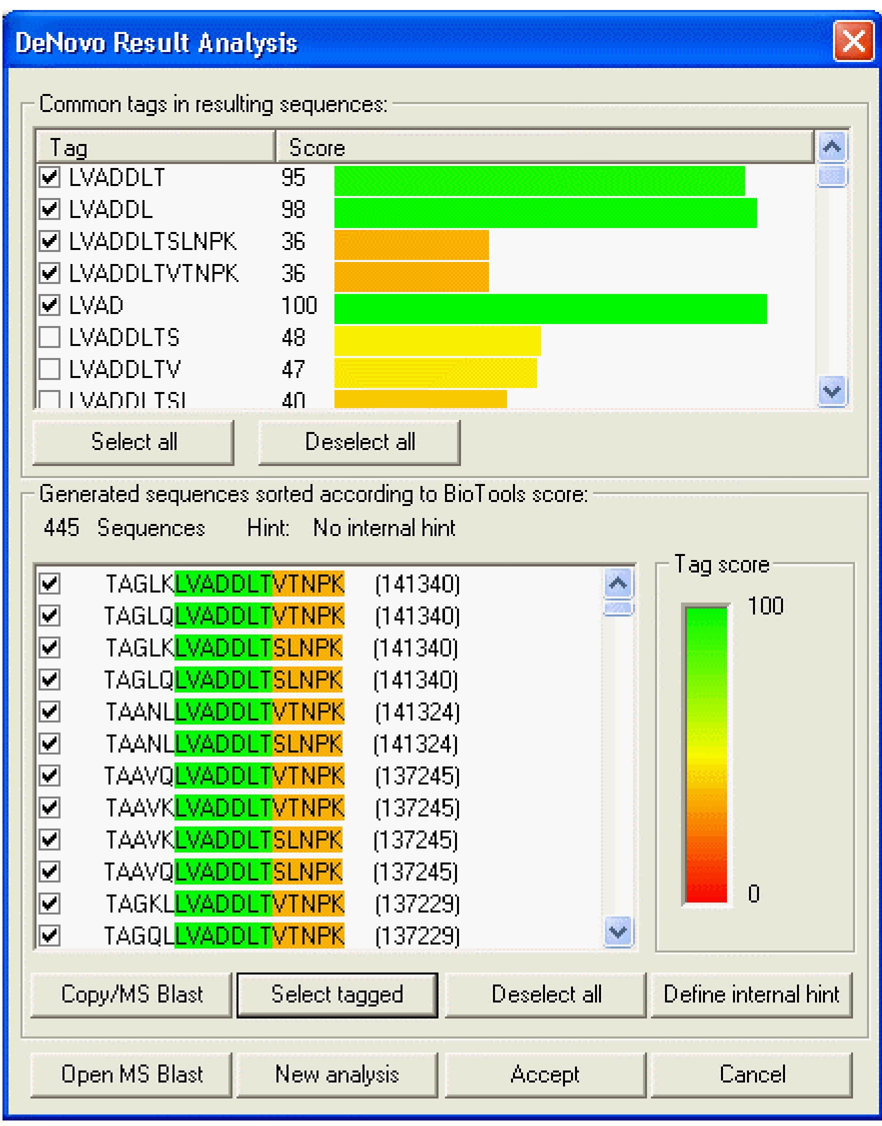


Figure 10

D.2.2.2.4. Explanation of entries in DeNovo Result Analysis window

Parameter/ Operations	**Explanation / Recommendation**
Common tags in resulting sequences	The successful match of MS-MS data to likely sequence motifs is displayed in the upper section of the window. Matching motifs are ranked and given a color coding depending on the score obtained in the DeNovo calculation.
Select all common tags	All common tags can be selected.
Deselect all common tags	All common tags can be deselected.
Generated sequences sorted according to BioTools scores.	The lower section of the window gives a display of the complete sequences which were generated to match the MS/MS data under investigation. The peptide list is sorted according to the respective score and motifs are displayed color coded.
Select tagged	This button selects all full sequences that have been tagged as marked in the Common Tag list.
Deselect all	All peptide sequences can be selected.
Define internal hint	By selecting **one sequence tag and one peptide containing this tag,** an internal hint is created for a subsequent sequencing step (**NewAnalysis**) or for a sequence tag search (described in F.2.2.3.3, Internal Hints for DeNovo Sequencing).
Copy MS Blast	All tagged peptide sequences can be copied to clipboard.
Open MS Blast	This button opens the web browser installed on the system and navigates to the current URL for EMBL MS Blast homology searches: http://dove.embl-heidelberg.de/Blast2/msblast.html. F.2.2.4 Excursion to a BLAST Search and the website for further information on the use of MS Blast for homology searches. This URL can be modified by including an entry in the *DeNovo* section ('URL for MS Blast' in MSTool.ini).
New analysis	This button closes the Result dialog, reactivates the Settings dialog and copies the selected sequences to the Treeview.
Accept	This button closes both the Result and Settings dialogs and copies the selected (check marked) sequences to the Tree View.
Cancel	The *DeNovo* sequencing is canceled.

The Common Tag part of the *DeNovo* Result Analysis window shows the resulting sequence motifs calculated together with a score. The order of listing is determined by the match score multiplied by the length of the tag . The first five tags are tagged automatically.

It is quite simple to inspect and review the quality of a given sequence match. To get a quick overview, it is sufficient to highlight a sequence in the list, and the ms-ms spectrum will immediately be annotated. Scrolling through the list allows differential comparisons.

For future reference these sequences can also be sent to the Biotools treeview. Firstly the peptide sequences are tagged (Figure 11) and sent to BioTools (Figure 12), where all tagged sequences are listed in the treeview and the annotated spectra can be viewed. Again a direct visual match comparison is possible by scrolling through the sequences.

	Sequence	
☑	TAGLKLVADDLTVTNPK	(141317)
☑	TAGLQLVADDLTVTNPK	(141317)
☑	TAANLLVADDLTVTNPK	(141316)
☑	TAANLLVADDLTSLNPK	(139527)
☑	TAAVKLVADDLTVTNPK	(137222)
☑	TAAVQLVADDLTVTNPK	(137222)
☑	TAGQLLVADDLTVTNPK	(137221)
☑	TAGKLLVADDLTVTNPK	(137221)
☑	TAGKLLVADDLTSLNPK	(135432)

Figure 11

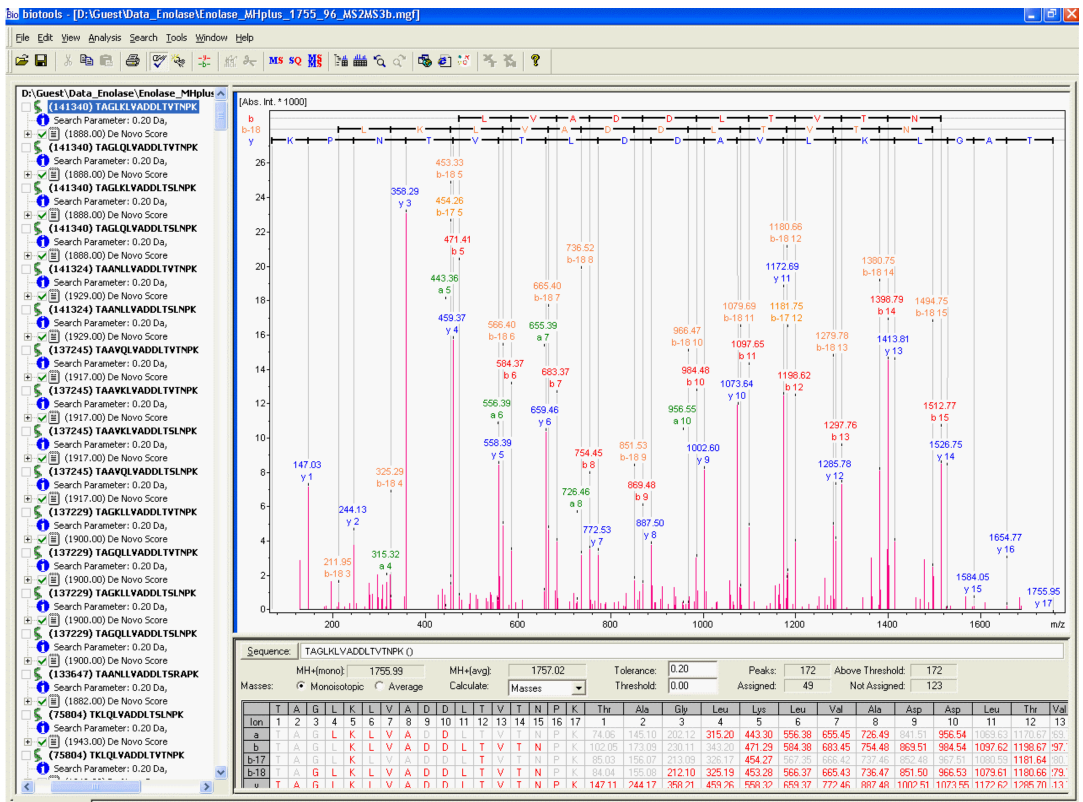


Figure 12

D.2.2.3. Useful Suggestions

Several approaches can be implemented to get good DeNovo sequencing results.

D.2.2.3.1. Using MS(3) data

Using only MS(2) data for *DeNovo* sequencing often results in peptides with degenerated termini – commonly 2 to 3 amino acids remain uncertain. This can be overcome by including MS(3)-peaks in the peaklist.

Figure 13

The upper part of Figure 13 shows one peptide bond cleavage resulting in an b- and an y-ion. Subsequent fragmentation of the y-ion (MS(3)) results in smaller fragments (y-ions and internal fragments) which were not accessible in the first fragmentation step.

The data file Enolase_MHplus_1755_96_MS2MS3.mgf was exported from one enolase compound containing MS(2)- and MS(3)-data (Figure 15 and Figure 14). The MS(3)-data helps to enrich the low mass range much better than the MS(2)-data. It is recommended to use the combined peaklist for *DeNovo* sequencing.

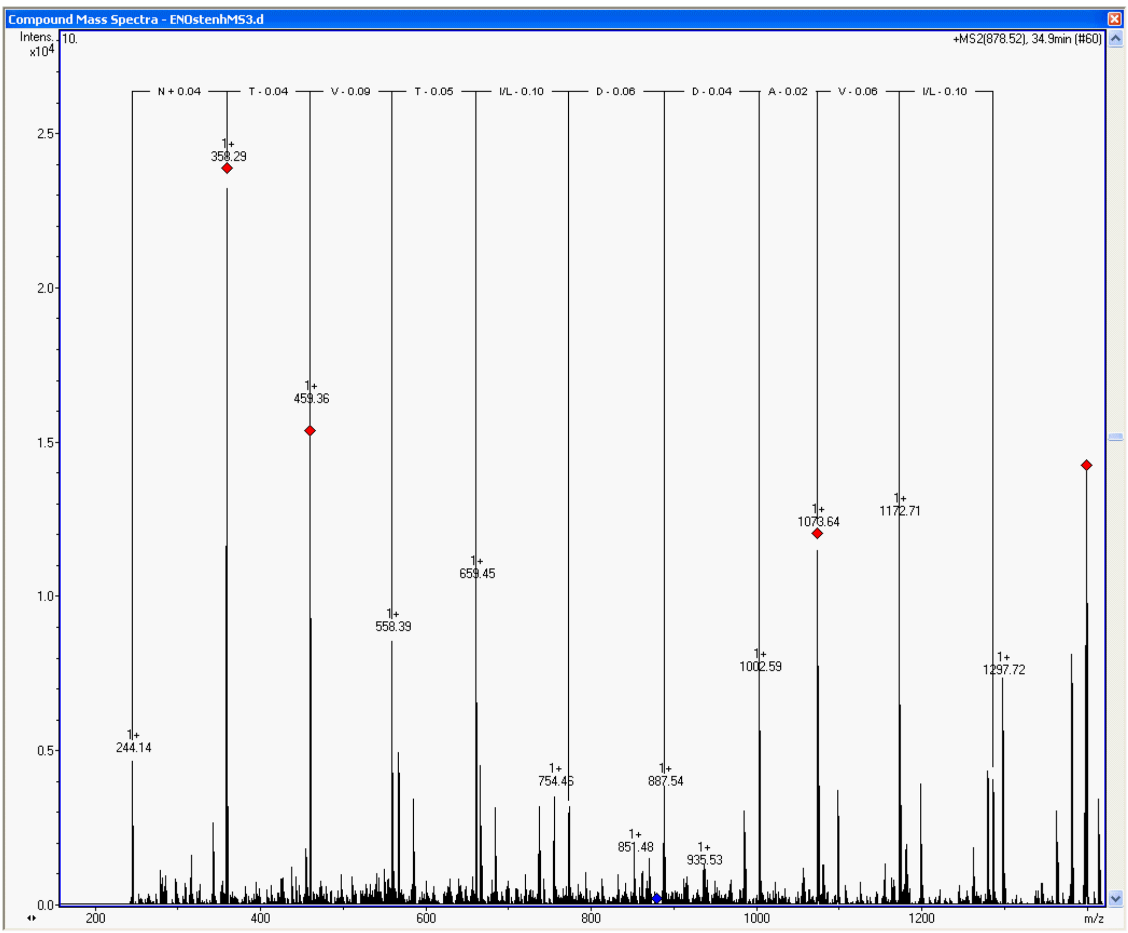


Figure 14

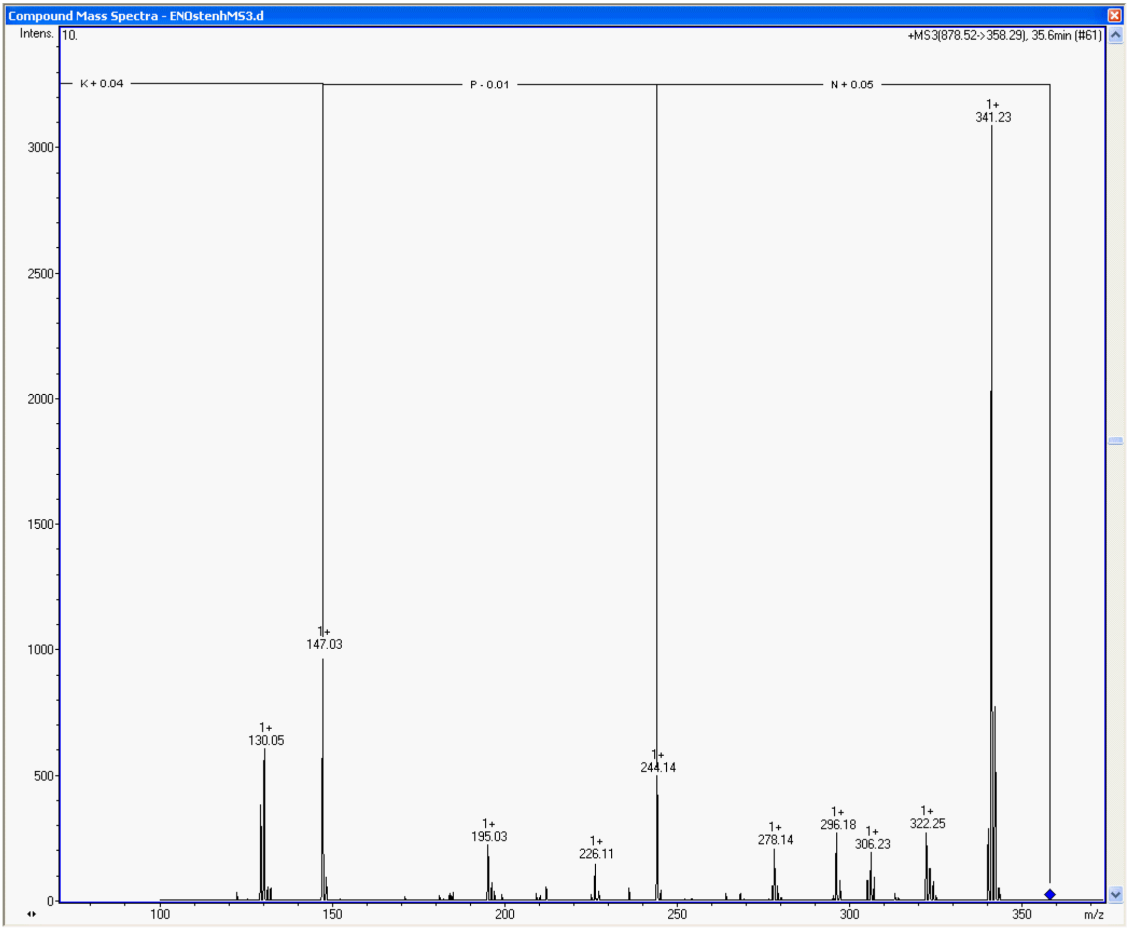


Figure 15

D.2.2.3.2. BioTools Threshold

If the MS(n)-spectrum is very noisy it might be helpful to submit only the more intense peaks for the calculation. An intensity threshold can be set in two ways:

- Choose **Analysis | Set Threshold** and use the mouse to move the line to an appropriate level. Then fix it by clicking the right mouse button.
- Or type a value in the fragmentation table.

The fragmentation table displays the total number of peaks and the number of peaks above the threshold. Usually, values of 50 to 150 work fine.

D.2.2.3.3. Internal Hints for DeNovo Sequencing

Since in ion trap tryptic MS(n)-spectra y-ions tend to be more intense than b-ions the highest peaks in the spectrum most probably are y-ions. Using **Annotation | Annotate** in DataAnalysis the highest peaks can be connected to give an internal hint. The MS(2)-spectrum (Figure 14) shows an adequate example. A clear y-series starts at 244.15 with the sequence NTVTLDDAVL.

Choosing Internal Hints from the Tag Suggestions

In *DeNovo* Result Analysis window one sequence tag with the sequence NTVTLDDAVL or a subset of at least one peptide sequence containing this tag can be selected to **define** an **internal hint** (Figure 16).

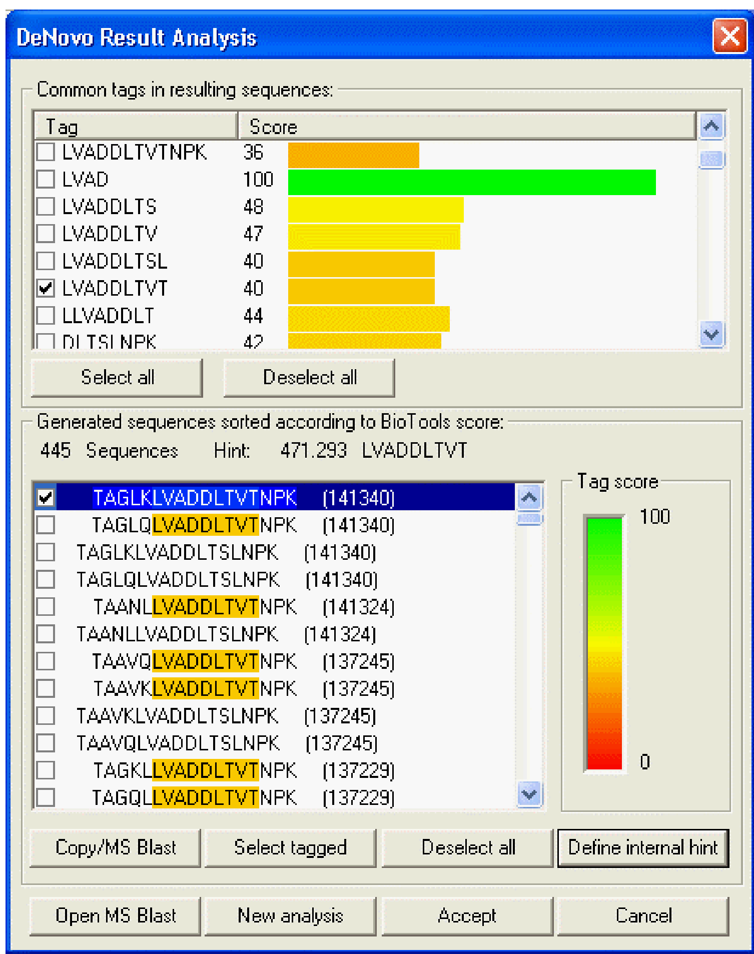


Figure 16

This can be used for a subsequent cycle of *DeNovo* sequencing (**New Analysis**). This tag is shown in *DeNovo* Settings window (Figure 17). Click on **calculate** to generate sequence tags. Then the peptide sequences in *DeNovo* Result Analysis window can be pre-selected (Figure 18), and sent to BioTools (**Accept**).

Figure 17

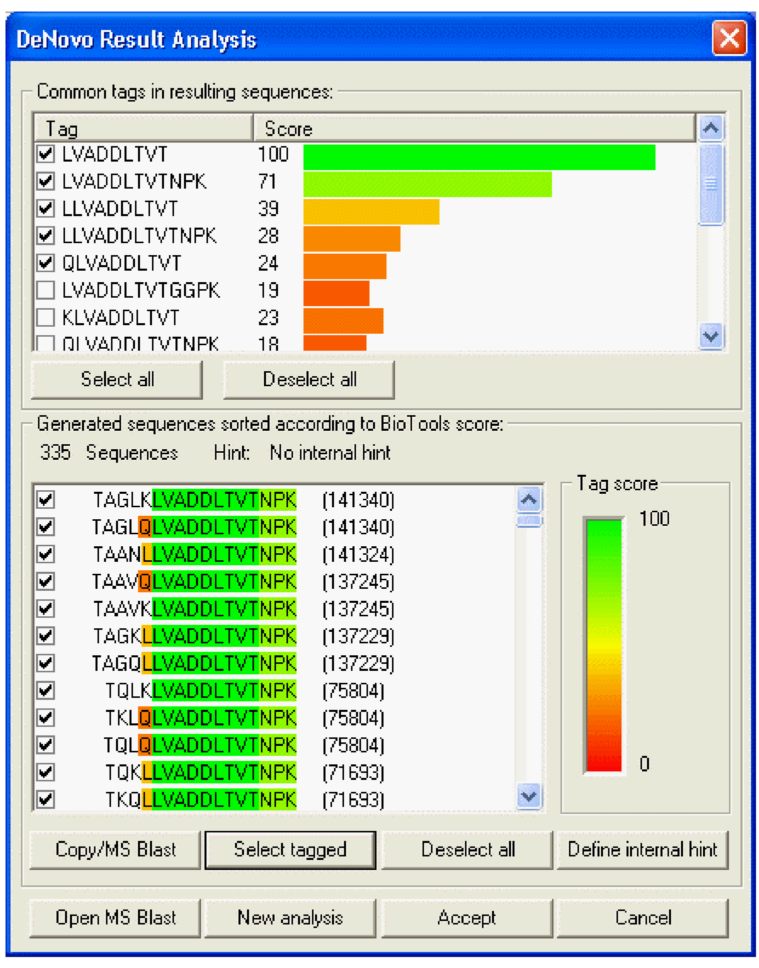


Figure 18

Defining the Internal Hint Manually

By choosing **Define Sequence Tag** in *DeNovo* Settings window, a sequence tag can be defined manually in Define Internal Hint window. The parameters used for Enolase_MHplus_1755_96_MS2MS3.mgf are shown in Figure 19. When "Define manually" is activated, the low mass from which the partial sequence starts should be entered first followed by the sequence tag from low to high mass in the MS(n)-spectrum. Upon clicking **OK,** the tag is sent to *DeNovo* Settings window, where the tag is rearranged from the N- to the C-terminus(Figure 20). Click on **calculate** to generate sequence tags. Then the peptide sequences in *DeNovo* Result Analysis window can be pre-selected (Figure 18), and sent to BioTools (**Accept**).

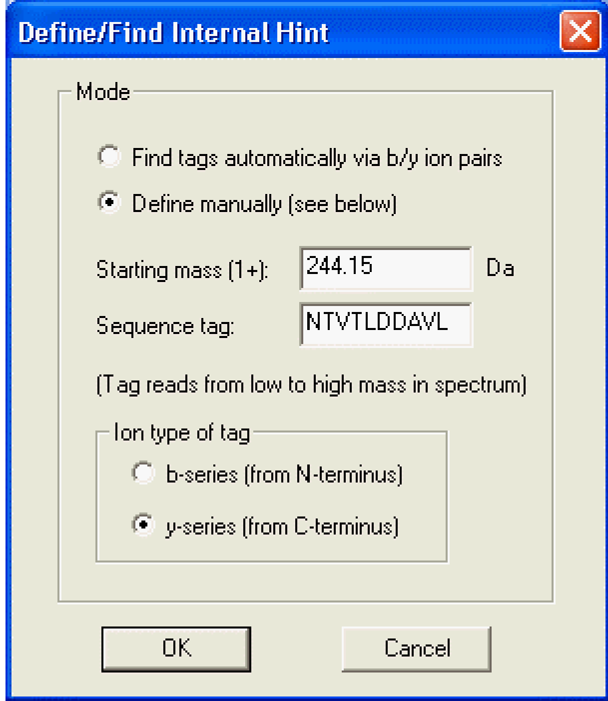


Figure 19

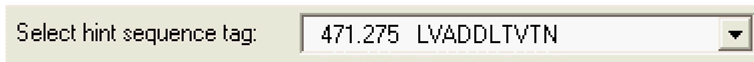


Figure 20

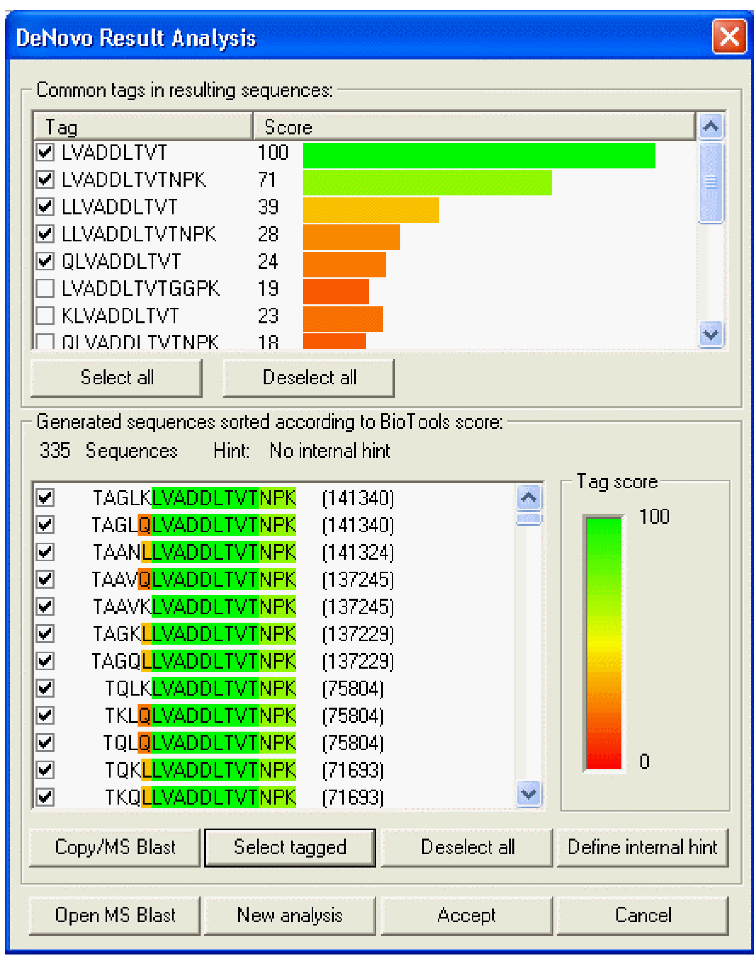


Figure 21

Sequence Tag Search

A sequence tag can be defined as described in chapter Choosing Internal Hints from the Tag Suggestions or Defining the Internal Hint Manually. Then, **Db Search For Tag** in the *DeNovo* Settings window can be clicked, and the Sequence Tag Search window appears.

For Enolase_MHplus_1755_96_MS2MS3.mgf the internal hint was defined as shown in Figure 19, and the parameters from Figure 22 were set in the Sequence Tag Search window.

Only a few important parameters from the Sequence Tag Search window are mentioned here:

- The URL is the URL for sequence tag searches at EMBL.
- The sequence tag typed in the Define Internal Hint window is recalculated for b-ions in the Sequence Tag Search window. Therefore, **b-type sequence ions** should always be selected in the Pattern Matched field.

Upon clicking **Start,** the sequence tag search is started and the results are generated as shown in Figure 23.

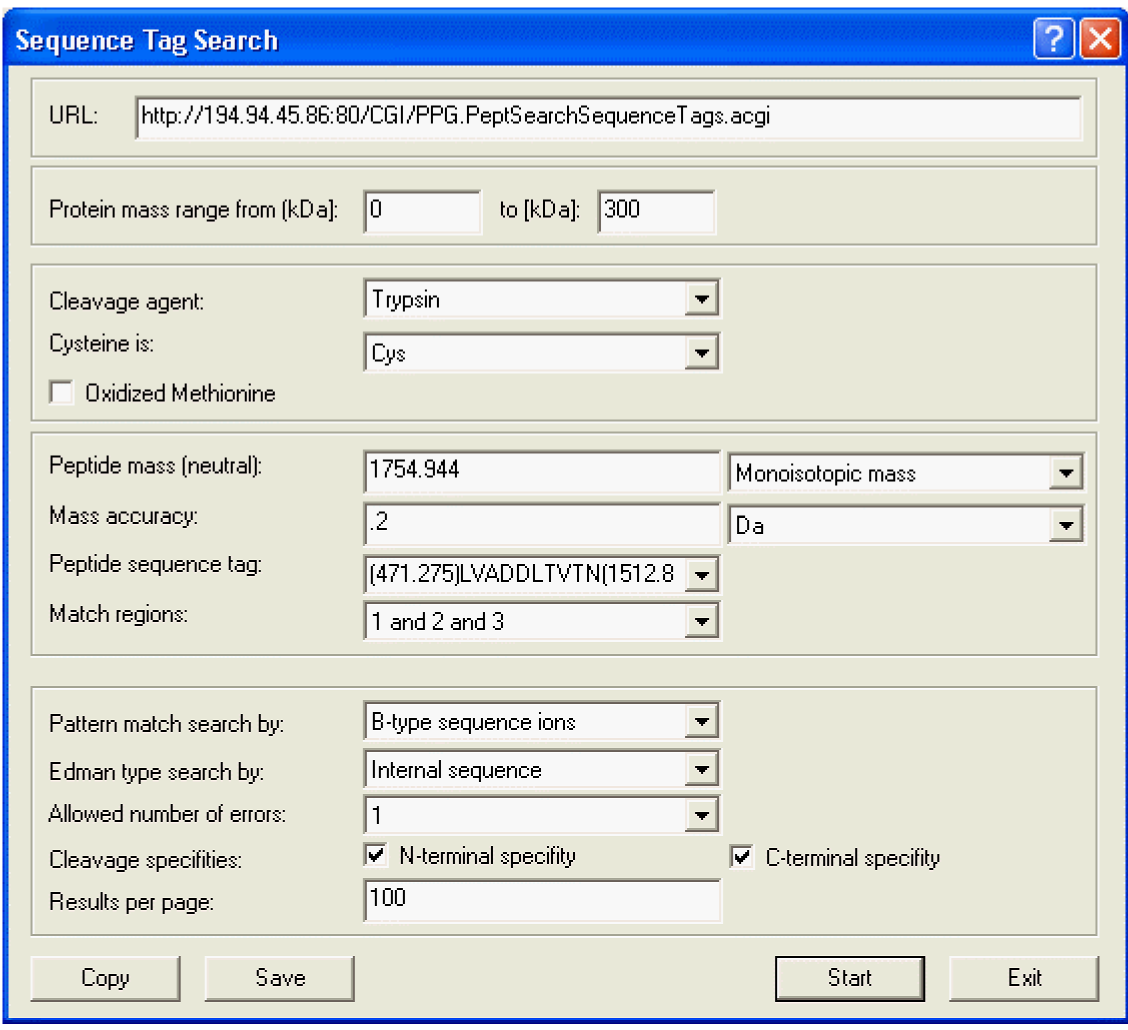


Figure 22

Search result

4 matches were found. Showing matches 1 through 4.

Peptide Sequence matched/ Peptide found sort	Mass [kDa] sort	Database accession sort	Protein Name sort	Digest
TAGIQIVADDLTVTNPK	46.671	swissprot:P00924	Enolase 1 (EC 4.2.1.11) (2-phos	
TAGIQIVADDLTVTNPK	46.63	pdb:1ELS (not known by SRS)	1ELS ENOLASE (E.C.4.2.1.11) (	
TAGIQIVADDLTVTNPK	46.816	trembl:X99228	SCCHRVII_3 gene: "ENO1"; S.cer	
TAGIQIVADDLTVTNPK	46.802	trembl:J01322	SCENOA_1 S.cerevisiae enolase g	

Figure 23

D.2.2.3.4. Defining Termini

The C- and/or the N-terminus can be manually defined in *DeNovo* Settings window. For the MS(2)- and the MS(3)-spectrum (enolase example) the following information was combined to define the C-terminus:

- As described in F.2.2.3.3 Internal Hints for DeNovo Sequencing the most intense peaks in the spectrum were assumed to be y-ions.
- The MS(2)-spectrum gave the tag "VTI/LDDAVI/L" starting from 244.15.
- This tag was extended by "P" starting from 147.03.
- 147.03 is known as the 1st y-fragment of a peptide with a C-terminal "K".
- From this a C-terminus "VADDLTVTNPK" was defined (Figure 24).

Click on **calculate** to generate sequence tags. Then the peptide sequences in *DeNovo* Result Analysis window can be pre-selected (Figure 18), and sent to BioTools (**Accept**).

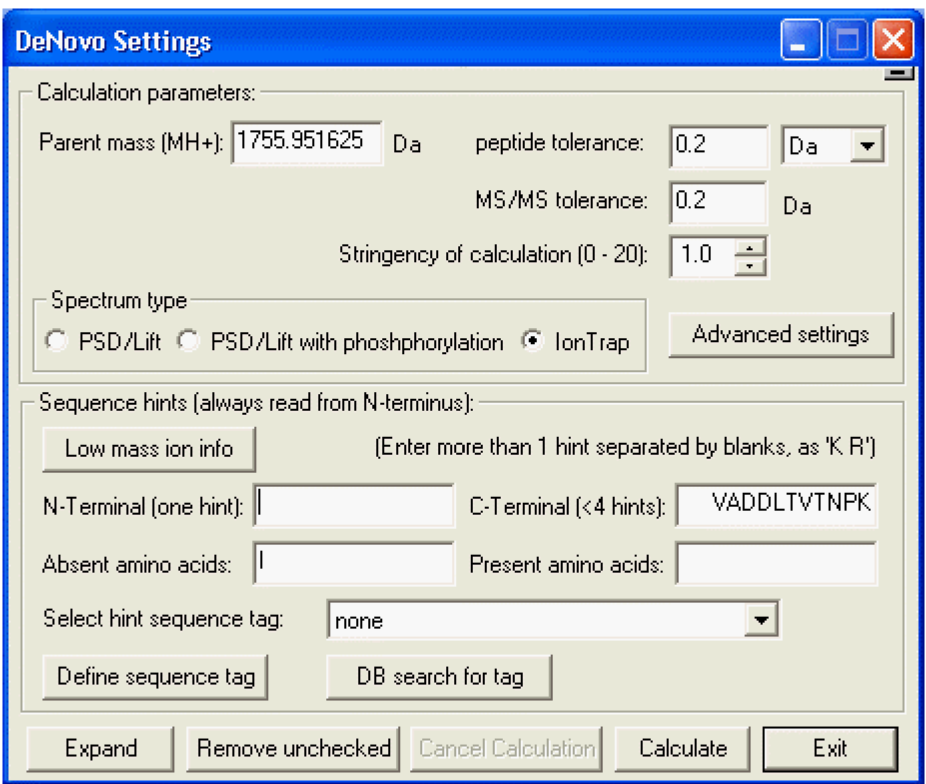


Figure 24

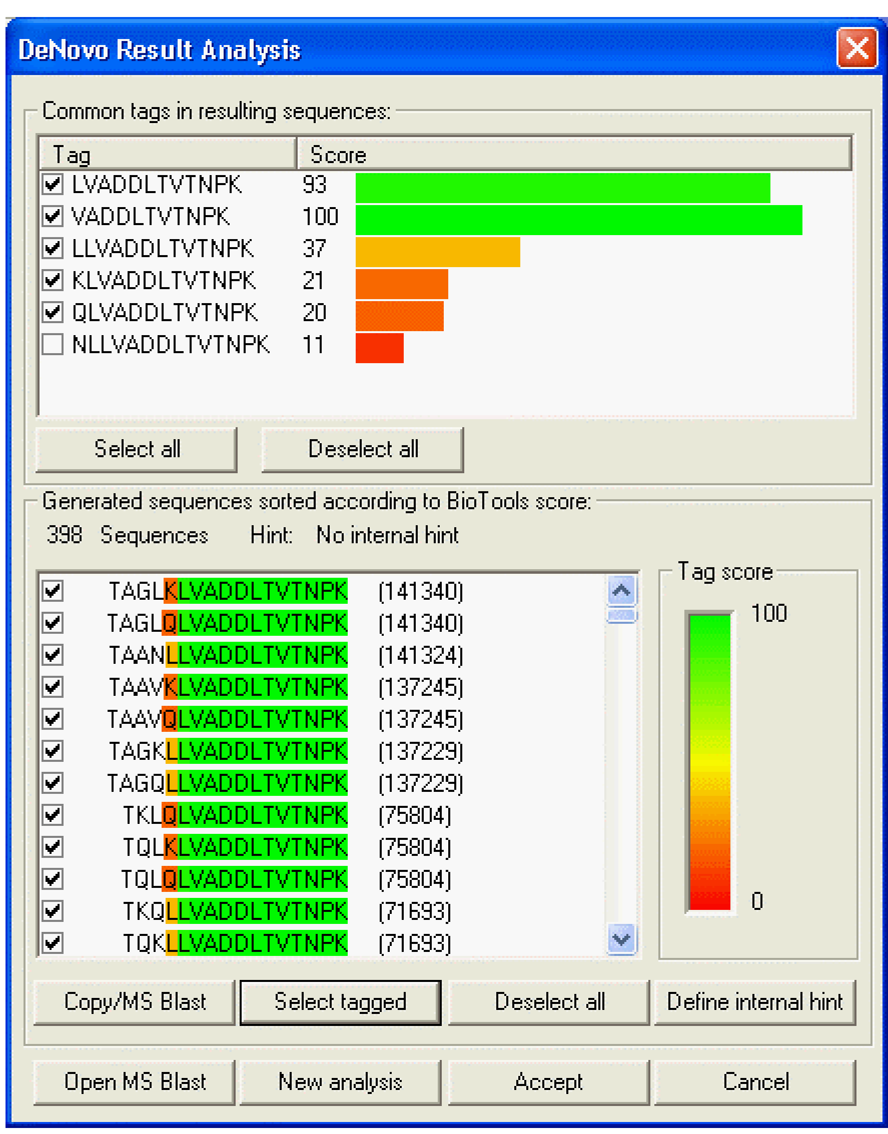


Figure 25

D.2.2.4. Excursion to a BLAST Search

Choose **Copy MS Blast** and **Open MS Blast** from *DeNovo* Result Analysis window (Figure 10). Then paste the clipboard, which contains all calculated and selected peptide sequences into the sequence field (Figure 26). For further information see site http://dove.embl-heidelberg.de/Blast2/msblast.html on the use of MS Blast for homology searches. Start the Blast search to get the results.

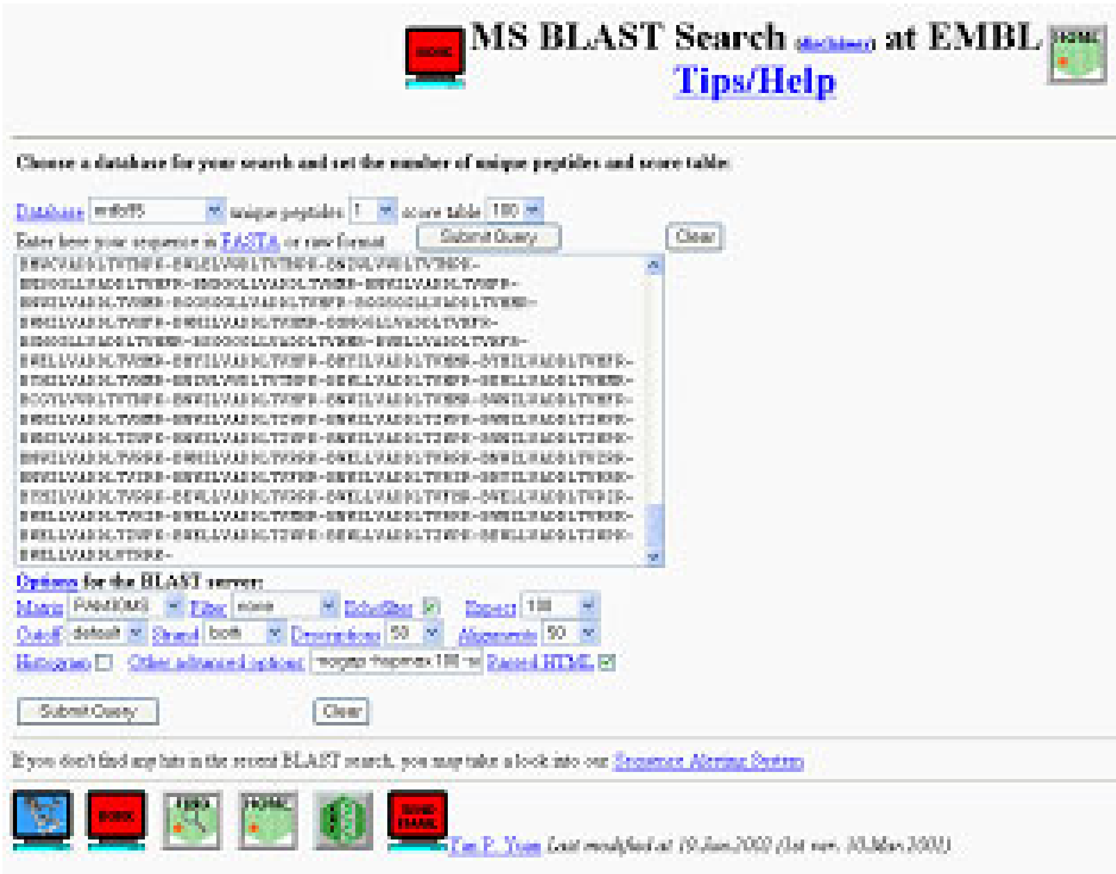


Figure 26

D.3 DeNovo sequencing

D.3.1 General Description

OUTLINE

The DeNovo generation sequence information from a peptide generally involves four steps beginning with suitable mass spectra followed by the translation of the mass spectral fragment data:

Step I Acquire an MS spectrum of the peptide or the digest providing utmost mass accuracy.

Step II Acquire MS/MS Spectra of one or more peptides from a protein digest and specify the best known precursor mass for the peptide of interest.

Step III Screen for known I-type ions, which may allow predicting which residues may be present or absent in the peptide, and for the y1 ion: m/z 175 indicates C-terminal R (Arg) and m/z 147 C-terminal K (Lys).

Step IV Screen for mass differences in the spectrum, which agree with the known masses of amino acid residue masses and assign partial sequences (tags). If one achieves the generation of tags, which make up for the full molecular weight and agree well with the different fragment ion types with a certain redundancy, i.e., b and y ions partially covering the identical sequence, "the peptide is sequenced"

The last step is the most time consuming step if done manually, taking from as little as 30 min to as much as 2 weeks per peptide. The *-novo* functionality in BioTools is designed to largely automate steps III and IV and to enable the biochemist – not only the mass spectrometrist – to successfully *DeNovo* sequence peptides and use the information for protein discovery and characterization.

Several other strategies have been developed over the last years and suggested readings are:

- I.A. Papatyannopoulos "The Interpretation of Collision-Induced Dissociation Tandem Mass Spectra of Peptides" Mass Spectrometry Reviews 1995, 14, 49-73.
- R. S. Johnson http://www.abrf.org/ResearchGroups/MassSpectrometry/EPosters/ms97quiz/SequencingTutorial.html

D.3.2 Processing in BioTools

STEPWISE FAMILIARIZATION WITH FULL DENOVO SEQUENCING - A WORKED EXAMPLE

The following example for peptide sequencing taken from the 2002 Martinsried Workshop "Micromethods in Protein Structure Analysis" serves to illustrate the stepwise explanation of an unknown sequence from raw ms-ms data. The ms-ms data was generated with two different instruments: an esquire*3000* + ION TRAP and the ultraflex MALDI TOF-TOF. Using either one of the two datasets the peptide sequence been successfully elucidated as shown in the following chapters.

D.3.2.1. LIFT-TOF/TOF MS Analysis of a Tryptic Peptide

Load the spectrum DemoData\FLEX*DeNovo* Sequencing\LID.LIFT into BioTools:

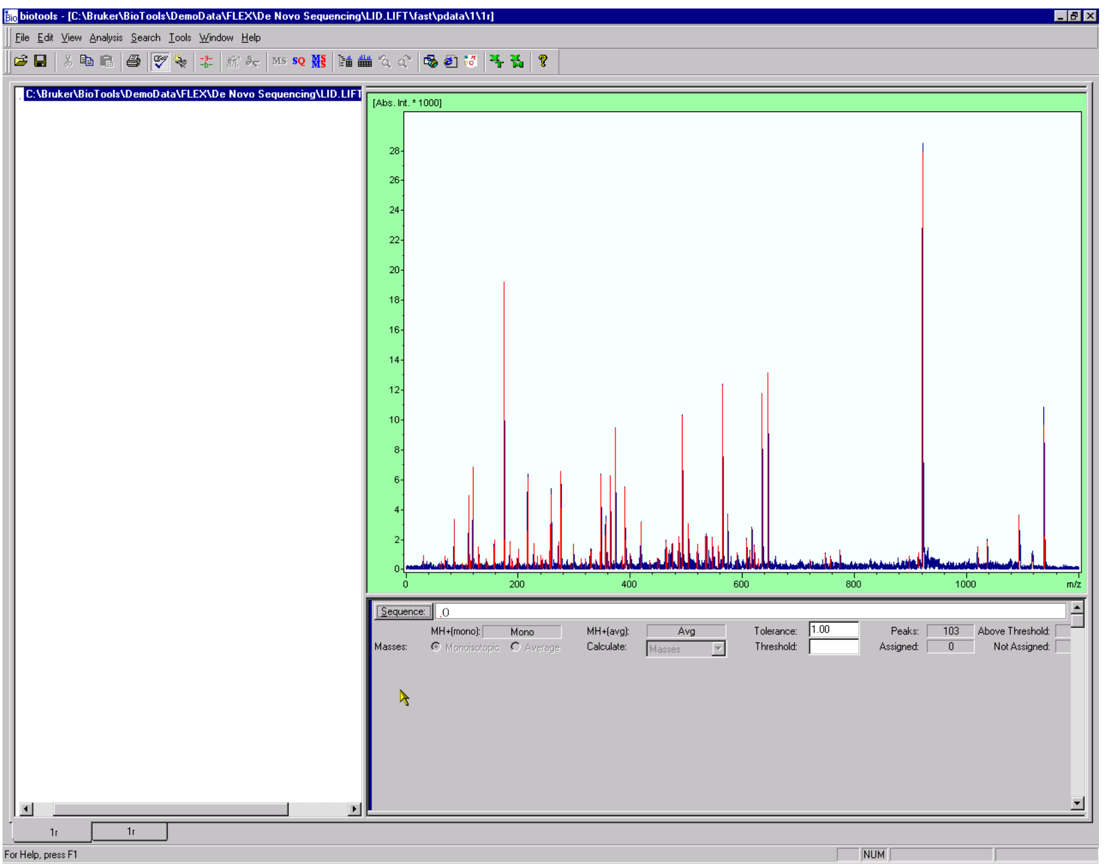


Figure 27

Open Pull Down Menu Option **Analysis | Full *DeNovo* Sequencing** to start the analysis. The precursor mass is automatically displayed in the new window, it is part of the data set and it has been specified during spectra acquisition. The expected **mass tolerances for MS and MS/MS** peaks are manually inserted into the dialog as shown in Figure 28.

The **stringency of calculation default setting of 5** can be used. In case of a calculation where no results were obtained this value may be reduced down to 0, however some less certain sequence candidates may now pop up in the following result list.

The **advanced settings** option of the **number of candidates** can typically be left on the **default value of 100 sequences**. If no results are obtained this number can be raised,. however, the calculation time will increase in doing so.

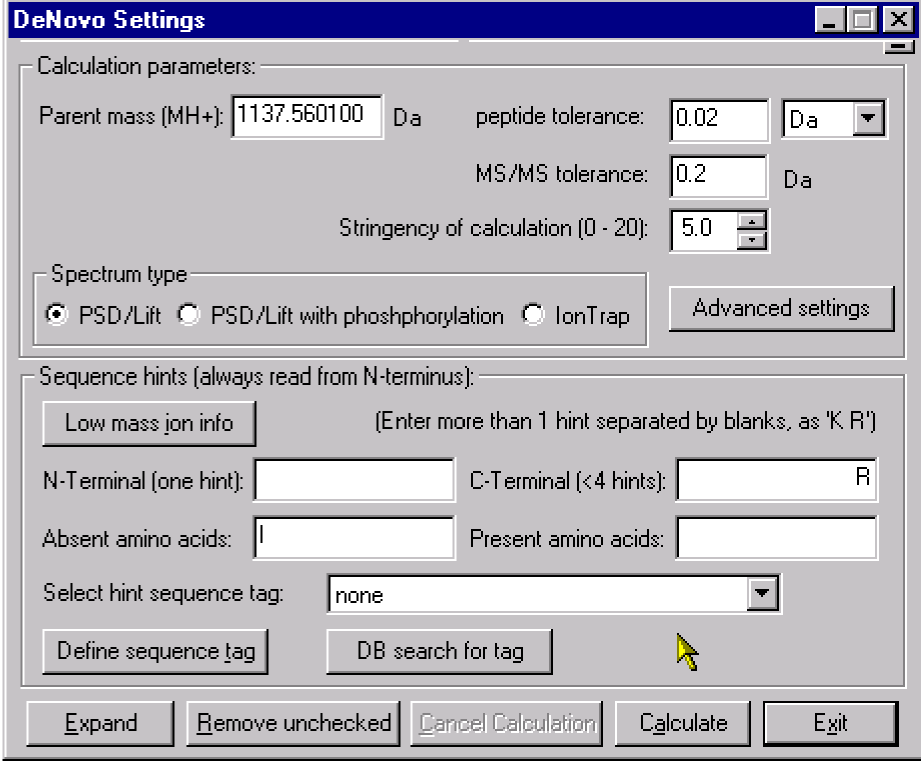


Figure 28

Sequence hints can be given in order to restrict the calculation and to apply any existing knowledge about a peptide. Such hints may be derived from the low mass ions (i-type and y1), by invoking the **Low mass ion info** button. This automatic search for present/absent amino acids has been limited to 6 residues: Y, F, W, H, L, and V,

which are thought to be most reliable. Please keep in mind that this filter routine typically gives good results, but the results should be interactively confirmed by checking the low mass range for correctly labeled signals.

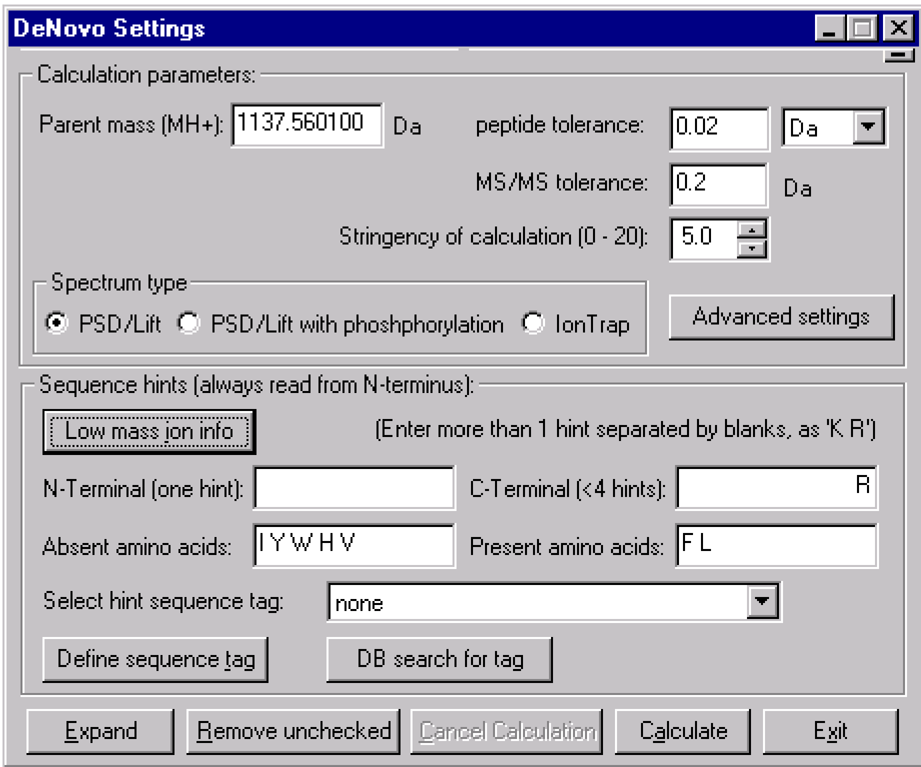


Figure 29

Any ambiguity in signals in this range, for example a high noise level, should be used to manually reduce the number of amino acids listed as hints for the next *DeNovo* calculation.

Manual editing of the sequence hints is possible, and even known sequences from either terminus can be added into the terminal hint fields, e.g., if a partial sequence is available from Edman sequencing, enzymatic sequencing, etc.

The presence or absence of the following low mass ions is related to the respective hints:

72	**V**, easy to miss due to incomplete peak picking
86	**L**
110	**H**, false positives if intensity is not VERY strong
120	**F**
136	**Y**
147	**K**(y1)
159	**W**
175	**R**(y1)

Sequence tags can be generated or manually entered prior to the calculation. On a given ms-ms spectrum we may define a hint sequence tag, based on information available from an abundant series of (a,)b and y ions. The automatic search for such series easily allows to obtain a sequence tag hint.

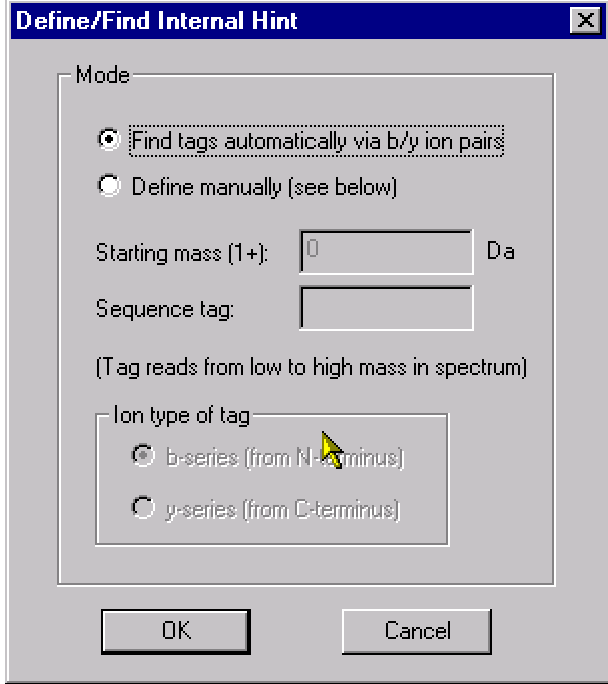


Figure 30

To invoke an automatic search, press the *Define sequence tag* button and select *Find tags automatically*. Press *OK*. In the example data some sequence tags are suggested, select '493.199 AAL'.

In cases like this, the abundance of corresponding sets of a, b and y ions in the MS/MS spectrum provide sufficient redundancy to calculate sequence tags automatically. A properly selected tag in the *DeNovo* Settings dialog may dramatically facilitate *DeNovo* calculation. These tags also may be generated as a result of the *DeNovo* calculation in order to seed a second round calculation keeping major clear sequence motifs fixed.

Alternatively, a sequence tag could be defined manually here by specifying the starting mass and the sequence of a tag.

> **Note:** The sequence tag concept was originally developed by Matthias Mann for MS/MS-based library searches. A tag has the syntax (start mass) SEQUENCE (end mass) – although in BioTools we use only (start mass) SEQUENCE to avoid over determination –. It comprises of 2 mass values and a sequence readout, which links these 2 mass values, assuming both mass peaks belong to the same ion series. Sequence tags are highly specific for a peptide and even for the complete protein sequence database. As a consequence, they can be used to do protein identification as it has been implemented in the Peptide Search program (EMBL).

Press the *Calculate* button for the *DeNovo* sequence calculation. After calculation, the *DeNovo* Result Analysis Dialog appears, which facilitates the interactive result evaluation.

The upper part of the dialog contains a list of sequence tags, which are sorted by tag length and quality of tag assignment in a combined way. The BioTools MS/MS scoring scheme is applied to these tags and the score expressed on a relative scale with a maximum of score = 100. The scale 0-100 is color-coded from red to green for good recognition.

All selected tags in this list are highlighted using the respective tag color in the list of resulting sequences in the lower part of the dialog. The BioTools score of the match between MS/MS peaklist and calculation is shown in parentheses. A click with the left mouse button onto an entry in the peptide sequence list will automatically cause the annotation of the spectrum using this sequence, as if it was transferred into the tree view of the spectrum.

DeNovo Settings
Calculation parameters:
Parent mass (MH+): 1137.560100 Da
peptide tolerance: 0.02 Da
MS/MS tolerance: 0.2 Da
Stringency of calculation (0 - 20): 0
Spectrum type
PSD/Lift PSD/Lift with phosphorylation IonTrap
Advanced settings
Sequence hints (always read from N-terminus):
Low mass ion info
(Enter more than 1 hint separated by blanks, as 'K R')
N-Terminal (one hint):
C-Terminal (<4 hints): R
Absent amino acids: I Y H V
Present amino acids: F L
Select hint sequence tag: 493.199 AAL
Define sequence tag
none
493.199 AA
564.232 AL
493.199 AAL
Expand
Remove unchecked
Exit

Figure 31

Using this feature it is easily possible to identify the best matching sequence amongst the top 3 candidates in the list of peptides, in addition to the automatic correct assignment of the top candidate with the score 1248. After selection of each of the three top candidates inspect the mass range 900-1150 and look for the completeness of the y ion series assignment. The correct sequence TDFEAALNTR reads all the way through to the molecular ion and the m/z 921.5 base peak indicates a strong fragmentation at the C-terminal peptide bond of aspartic acid (D), which is very typical to observe and an additional confirmation of the correctness of the assignment. To finish the analysis please select those entries in the peptide list, which you would like to accept for transfer to the BioTools tree view and press the *Accept* button.

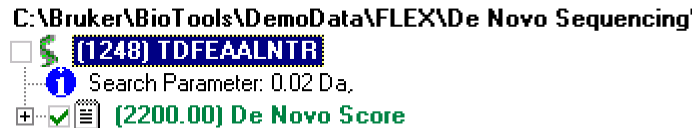


Figure 32

In the BioTools treeview, the BioTools MS/MS score is displayed next to the sequence in parentheses and additionally the internal score of the *DeNovo* calculation algorithm is displayed. This additional information may be helpful in cases, where the standard score is not distinctive.

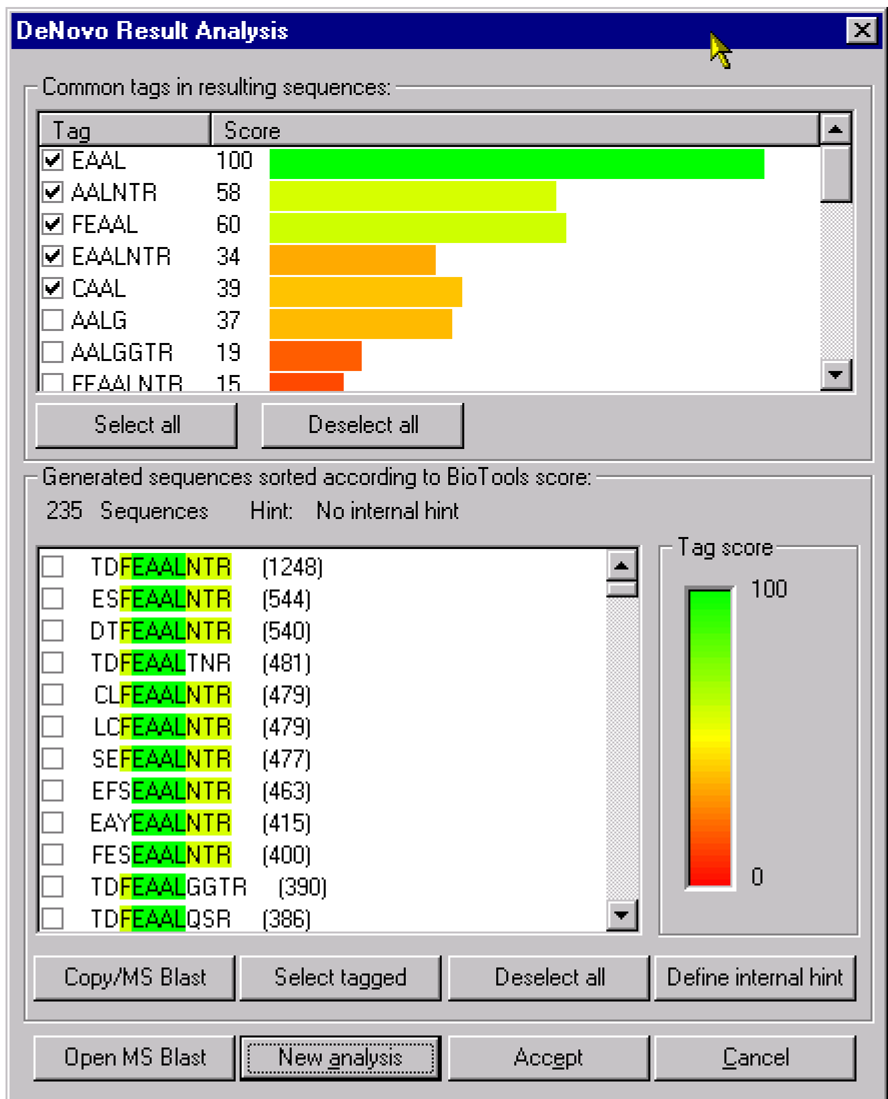


Figure 33

Based on such assignments one could specify now an extended sequence tag hint for a second round of calculation: Select exactly one sequence tag AND exactly one peptide sequence, then push the *Define internal hint* button. *New analysis* brings you

back to the *DeNovo* Settings dialog and the sequence hint is selected automatically for the next calculation.

Using MS-BLAST for homology searches based on the *DeNovo* sequencing is described.

One open question remaining is whether or not the assignment of Leu 7 is correct or whether Ile 7 is the correct result. This question cannot be answered with the currently loaded LID-LIFT file (fragmentation patterns similar to PSD); rather high-energy collisions must be used.

Therefore, transfer the final sequence from the LID-LIFT spectrum to the SequenceEditor by pushing the *Sequence* button underneath the spectrum display.

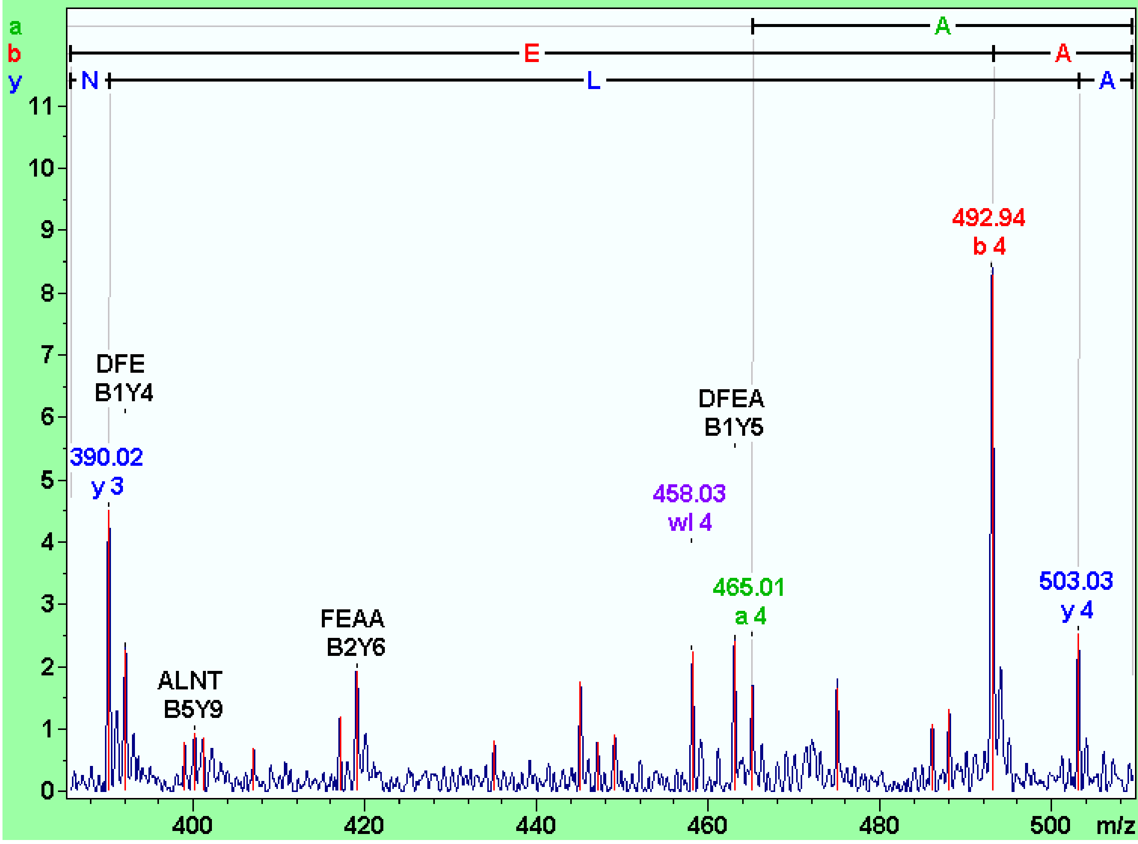


Figure 34

Please load the spectrum DemoData\FLEX*DeNovo* Sequencing\ CID.LIFT, which is the corresponding CID-LIFT spectrum. In the Sequence Editor push the *BioTools* button to transfer the correct Sequence into the active CID-LIFT spectrum. Select the

ion series set "high energy CID" as annotation parameter ( ) with at least the y and the w- ions selected. At a mass tolerance setting of 0.4 Da you should obtain the annotation of the w4 ion at m/z 458, indicative for Ile at this position (wI 4); Leu would give an annotation wL 4.

As a nice exercise you could try to repeat the full *DeNovo* sequencing analysis with this CID spectrum. But be warned: it is much more difficult to do sequencing with CID data compared to LID data!

D.3.2.2. Combining DeNovo Sequencing with Homology Searching

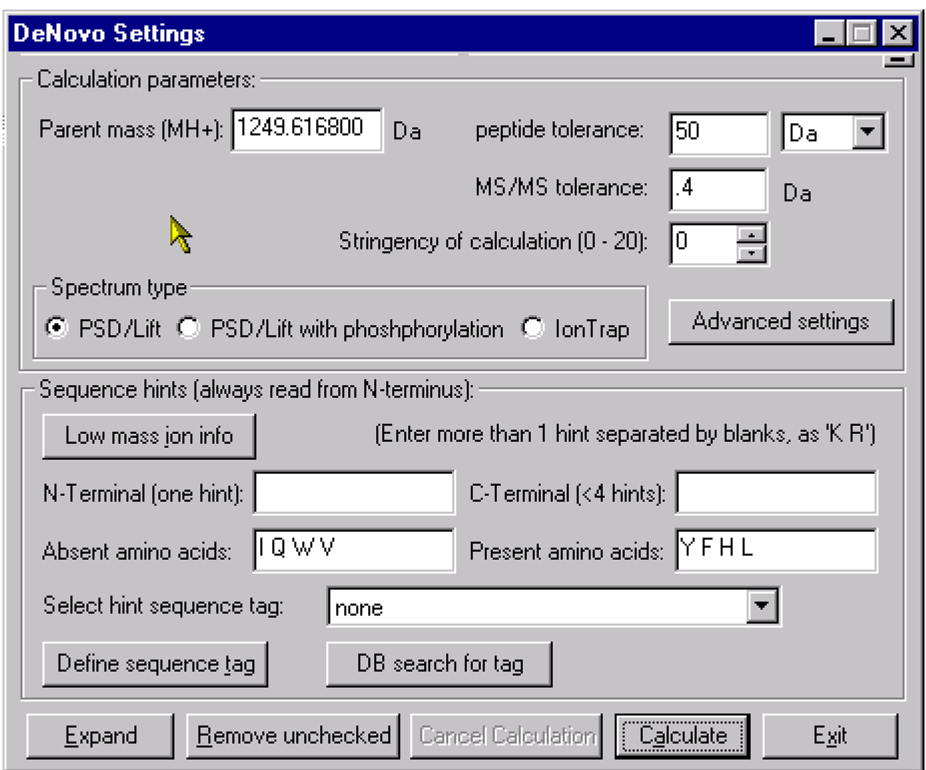


Figure 35

The sequences generated by the *DeNovo* sequencing algorithm may not always provide 100 % correct sequences. However, in most cases a significant part of a candidate sequence is correct. This is a good situation for protein identification based on homology of the determined sequences and not on a perfect match as is typically required for all mass spectra based searches using search engines such as MASCOT.

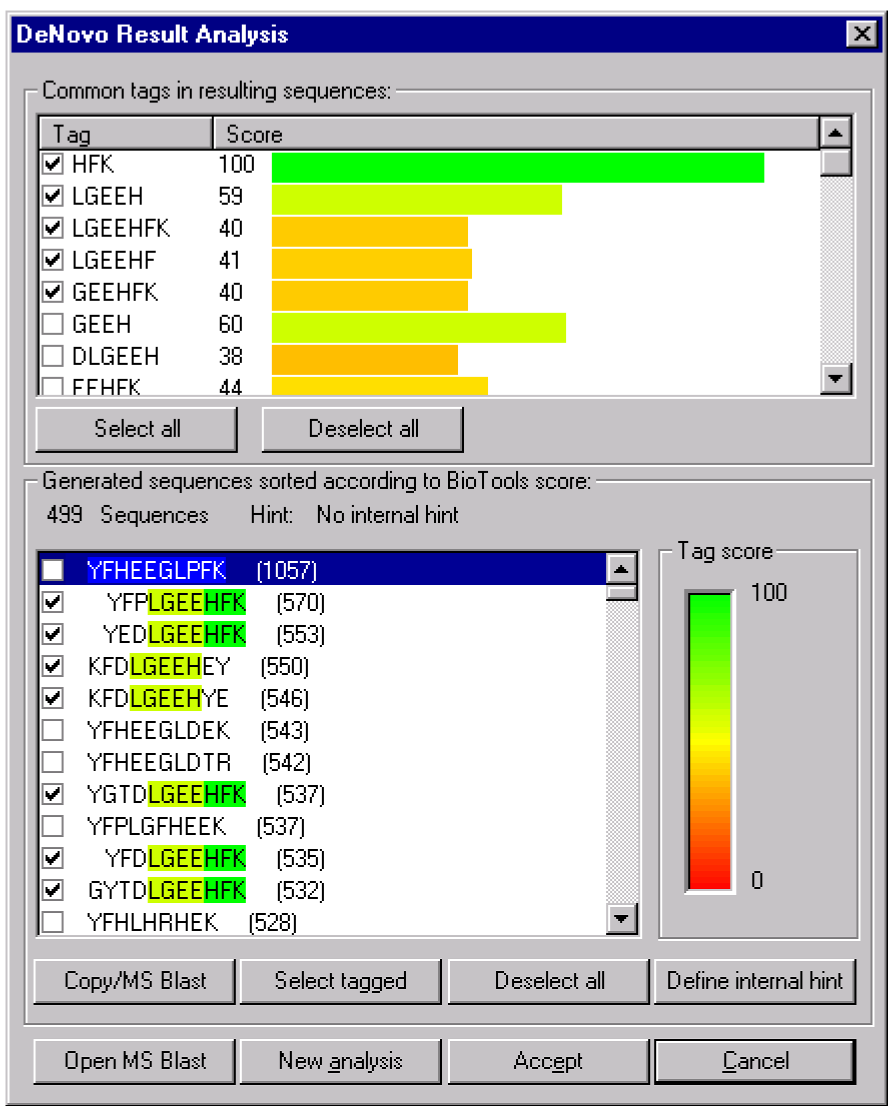


Figure 36

Load the LIFT spectra from a BSA digest under DemoData\Flex*DeNovo*BatchData\0_I14 into BioTools. Select spectrum of precursor mass 1229.6 (i.e., DemoData\Flex\BatchData\0_I14\1\1246.6168.LIFT). Open the DeNovo Settings dialog and specify 50 ppm peptide tolerance and 0.4 Da MS/MS tolerance. After pressing *Low mass ion info* you should get this dialog.

Calculate gives you this DeNovo Result dialog, in which you just press *Select tagged*. This selects all peptide candidate sequences containing either of the selected

sequence tags. You may of course add either sequences or tags to the list, but in this case there is no need for doing so.

Copy/MS Blast copies all selected sequences into the clipboard in a format compatible with MS-BLAST searches at http://dove.embl-heidelberg.de/Blast2/msblast.html. Sequences are separated by hyphens and the following letters are used to code for special purpose:

L stands for Leu and Ile Z stands for Gln and Lys X stands for any amino acid B a putative trypsin cleavage site, stands for Arg or Lys residue preceding the complete sequence

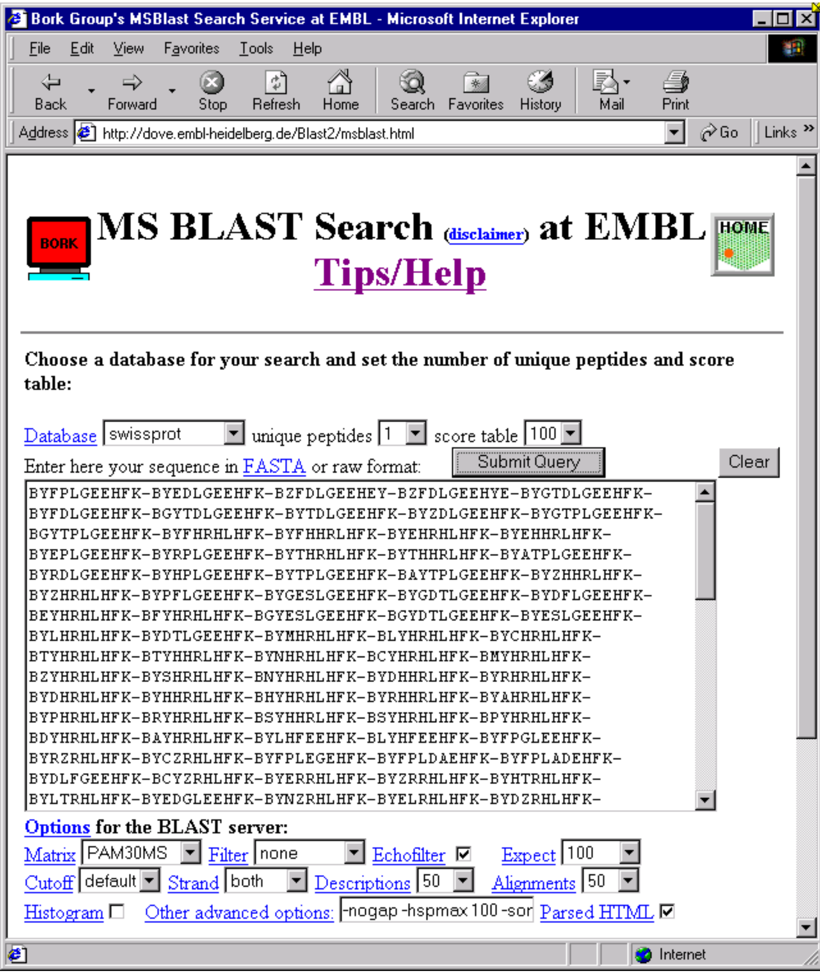


Figure 37

Open MS Blast opens the Browser at the MS BLAST site of the EMBL and paste the sequences contained in the clipboard into the sequence field of MS BLAST. Select the proper database (here SwissProt) and *Submit Query* without changing any other default.

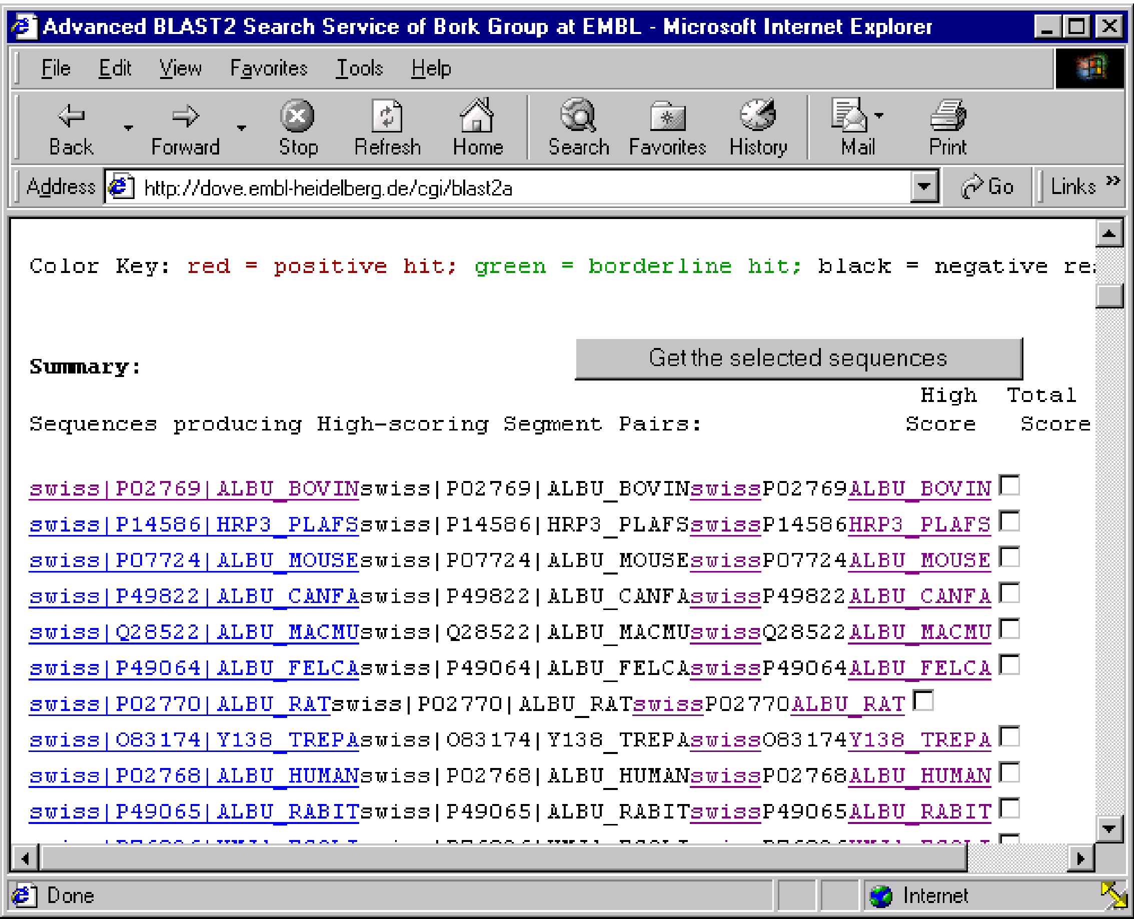


Figure 38

After a while the search result appears containing a whole set of different serum albumin sequences from various species. That is good! But the result is not very specific based on the scores and color coding of results provided by MS BLAST. That is bad. The good news is we can increase the specificity of such homology searches by simultaneously using sequences obtained from several peptides of the same protein.

Please repeat the *DeNovo* sequencing now on some of the other peptides from the same sample: use the LIFT data from m/z 1639.9 or 927.8. Paste the set of

sequences from the clipboard behind the last hyphen in the MS BLAST sequence field each time and repeat the MS BLAST search.

Hints:

The peptide at 1639.9

- V is wrong absent and H wrong present.

The peptide at 927.5

- H is wrong present.

The peptide at 1439.8

- H is wrong present.

The peptide at 1479.8

- V is wrong absent.

For a better understanding of the result interpretation of MS BLAST, the Tips/Help section is highly recommended on that web site.

Sequences may be copied from a result page and fed into BioTools via pasting into the SequenceEditor. Here also the *Search>Sequence* function may facilitate finding the matching sequence motifs.

T Tutorials for SequenceEditor

B Basic Operation

S Modify a Sequence

P Protein Digest

C Fine Structure Characterization

B Basic Operation

This chapter describes the basic appearance and ways to manipulate the SequenceEditor application window.

B.1 Starting SequenceEditor B-2
B.2 Data Format B-2
B.3 Basic Sequence Handling B-2
B.3.1. Create a New Sequence B-3
B.3.2. Open a Sequence B-5
B.3.3. Save a Sequence B-6
B.3.4. Delete a Sequence B-6
B.4 The Molecular Weight of a Sequence B-7
B.4.1. Gross Sequence Molecular Weight B-7
B.4.2. Calculate Multiple Charge-State m/z Values of Molecular Ions B-7
B.4.3. SeqInfo - Composition B-9
B.5 Standard Modifications of the Sequence B-10
B.5.1. Cysteine state B-10
B.5.2. Termini B-10
B.6 Visualize particular Residues and Sequence Motifs B-11
B.6.1. Search for Sequence Motifs B-11
B.6.2. Highlight Aminoacid Residues and N-Glycosylation Consensus B-11
B.7 Export Sequence to MS/MS Spectrum Display B-11
B.8 Printing Data B-11

B.1 Starting SequenceEditor

On completion of the installation of BioTools a Bruker Daltonics program group is created; this contains a SequenceEditor program icon.

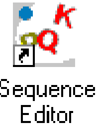


Figure B-1, SequenceEditor application

SequenceEditor can be started by double clicking on this icon either on the desktop or within the spectrum display of BioTools. The SequenceEditor application window appears on the screen.

B.2 Data Format

SequenceEditor can read data both in Bruker (*.sqs) and the GPMAW (*.seq, Lighthouse Data) formats. The data can be saved in the Bruker format only.

Note ***All sequences are stored and displayed only as single-letter code***

B.3 Basic Sequence Handling

The figure below shows the SequenceEditor application window as it will appear after program start. Multiple sequences in multiple windows can be opened in the sequence view and all calculations apply to the active sequence window.

The Sequence Editor contains two main elements:

The tree view (TV) on the left contains the sequence database files (*.sqs) which are comprising of up to several sequence entries.

The sequence view (SV) displays the currently opened sequence text, including modifications and crosslinks.

When the SequenceEditor is first started, the last loaded data set is displayed.

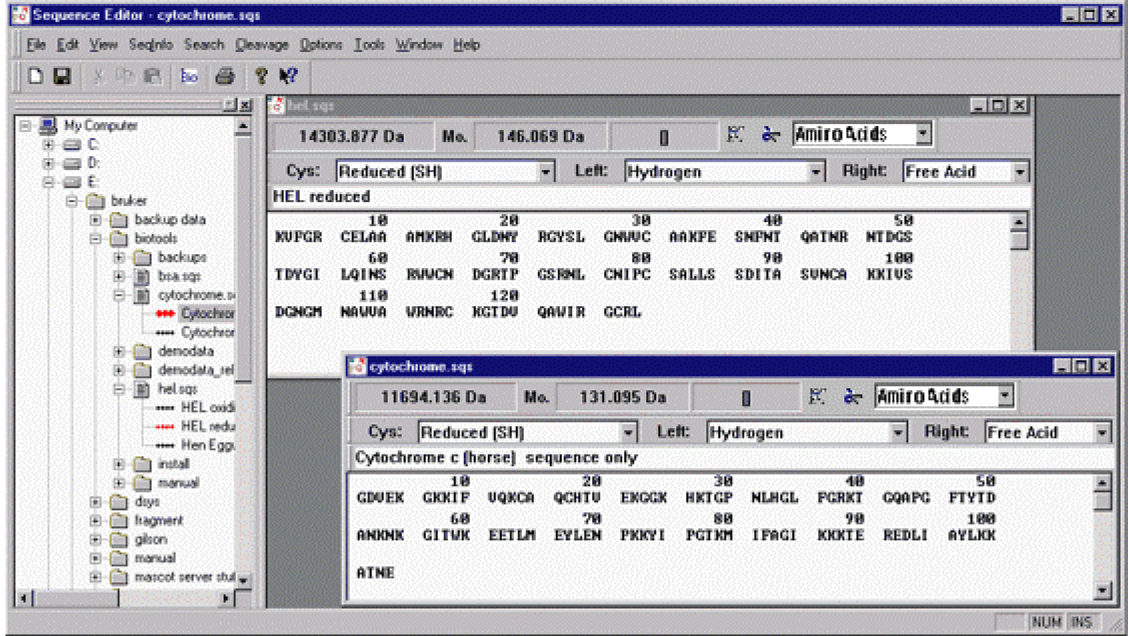


Figure B-2, SequenceEditor application

B.3.1. Create a New Sequence

The button opens a new blank sequence sheet and modifications are set to default values: the cysteine state to "Reduced (SH)" and the N-and C-terminal to "Hydrogen" and "Free Acid" respectively. Default building block list is **Amino Acids,** however, the default setting can be otherwise selected under Tools/Options. Any other type can principally be defined, currently also a DNA building block list is provided with the system.

A new sequence can be now entered manually **(single letter code only!)** or pasted from the clipboard. Non-alphabetical characters are eliminated from the text in the clipboard. A sequence name can be typed or pasted into the sequence name field.

Note ***Blanks and numbers do not need to be removed manually in the SE after copying a sequence from a internet database in a browser window.***

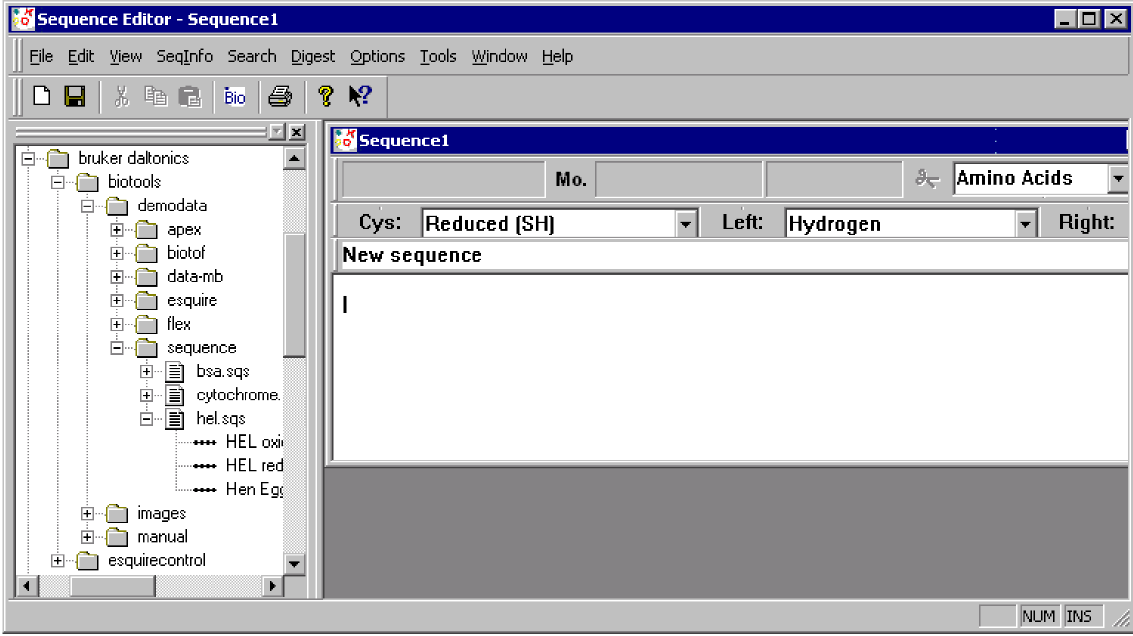


Figure B-3, Empty sequence window

Typing results in inserting/overwriting the new entered residues into the sequence. Any part of the sequence can be selected and the standard, cut, copy and paste functionalities apply.
Most of the changes made to the sequence can be undone (*Edit - Undo*).

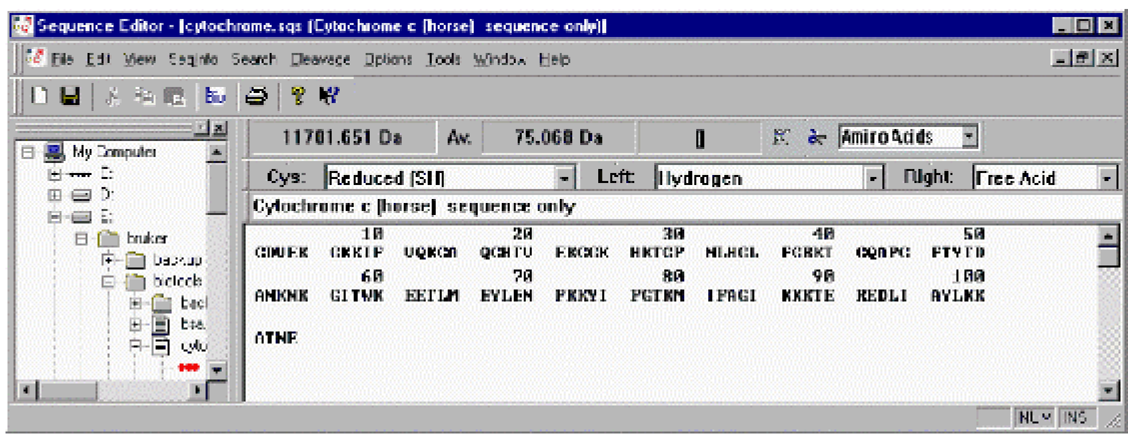


Figure B-4, New sequence

B.3.2. Open a Sequence

Sequences are opened using the TV situated on the left side of the screen. Each sequence file (*.sqs) can contain several protein sequences. A click with the LMB on the respective filename in the TV unfolds the contained sequences and displays the respective titles and the sequence length. In case of opened sequences, the symbol •••• is marked red. The currently active sequence (its window is on the top and has the input focus) gets the symbol: ••• assigned.

Double-click LMB on the selected sequence to open it. The simultaneous read operation of several sequences residing in the same file is possible.

Note ***Loading of GPMAW-format files (extension ".SEQ") is supported. If saving of modified information into such *.seq file is attempted, the new file will be stored as Bruker *.SQS file.***

The details of the tree structure are explained in following figure.

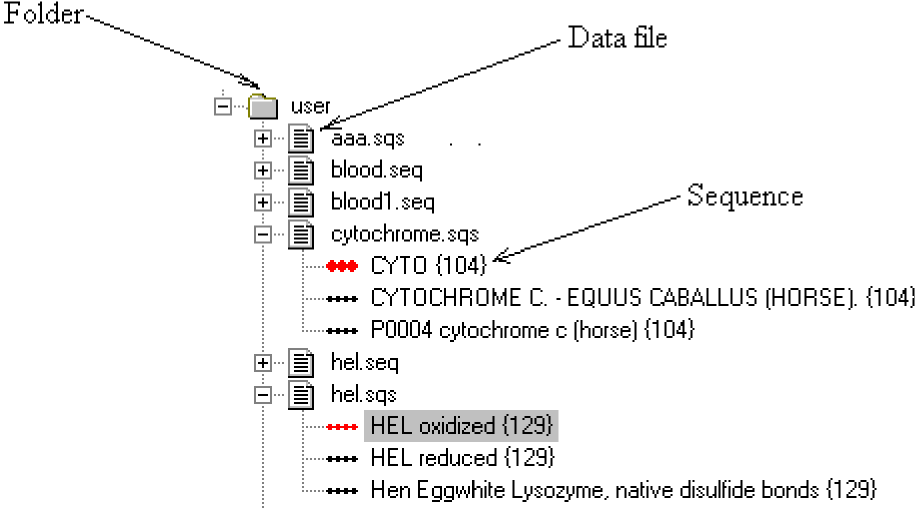


Figure B-5, Entries in the tree view window

B.3.3. Save a Sequence

The [save icon] button saves the active sequence with its current name within the previous *.sqs file.

If you worked on a new sequence, your SequenceEditor displays the "Save As" dialog box:

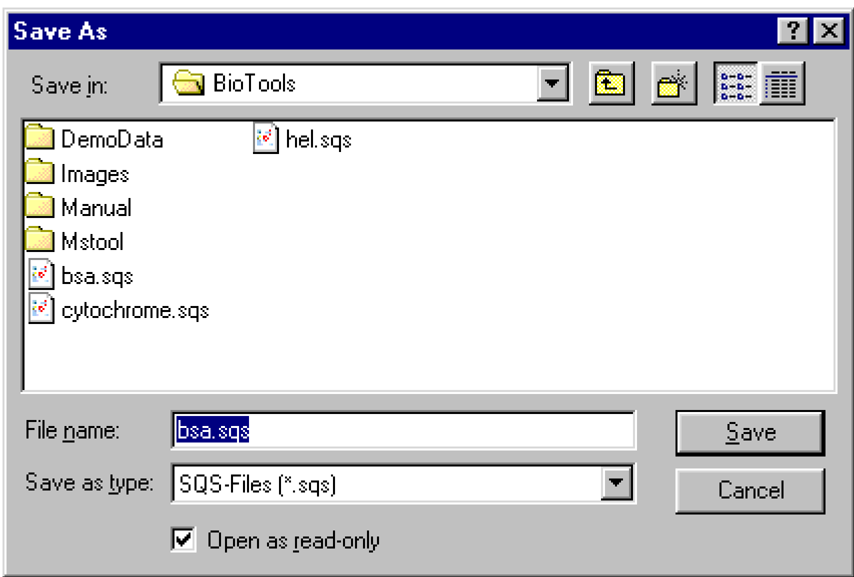


Figure B-6, Save sequence dialog

If an existing *.sqs file is selected, the sequence name will become a new entry within the existing *.sqs or a new one can be specified for sequence storage. The entry name will be as specified in the sequence name field.

B.3.4. Delete a Sequence

To delete individual sequence entries within a *.sqs file, click with the RMB onto the particular entry in the sequence tree view and a *Delete sequence* button appears. Press the button and the sequence is deleted from the *.sqs file.

Note ***The sequence database files (*.sqs) can be deleted like any other file type using, e.g., the Windows Explorer.***

B.4 The Molecular Weight of a Sequence

The toolbar situated on top of each sequence window shows the total molecular weight of the sequence, the residue mass under the cursor or of the selected range of residues, and the current number of the residue or the selected range. These mass values are switch-selectable between monoisotopic and average modes (button "Mo"). **Always neutral masses are calculated here!**

Figure B-7, Sequence window toolbar

The current cysteine state ("Cys") and the N-("Left") and C-terminal modifications ("Right") are displayed and can be changed by means of the combo boxes placed in the toolbar. The available options are defined in the BrukerDefault.mod file. The information supported in the lowest line is the title of the current sequence.

B.4.1. Gross Sequence Molecular Weight

In case of any relevant change to the amino sequence, the total mass is recalculated. The calculation considers the full covalent structure of a protein: sequence, cysteine states, termini and all the modifications and crosslinks defined within the sequence. The mass calculation is performed both for monoisotopic ("**Mo.** ") and average ("**Av.** ") molecular weight and either of these **neutral molecular weights** can be displayed. The updated total mass is immediately shown in the application toolbar. Use the *SeqInfo – Charge States* command to get the corresponding information about m/z values.

B.4.2. Calculate Multiple Charge-State m/z Values of Molecular Ions

Neutral molecule masses, as they are calculated based on the elemental composition, are not typically observed in a mass spectrometer. Rather, multiply protonated charge states are observed can be calculated and used for calibration or gross structure verification.

Use the *SeqInfo – Charge States* command to get this information. The *SeqInfo – Charge States* dialog box appears, which has the following entries.

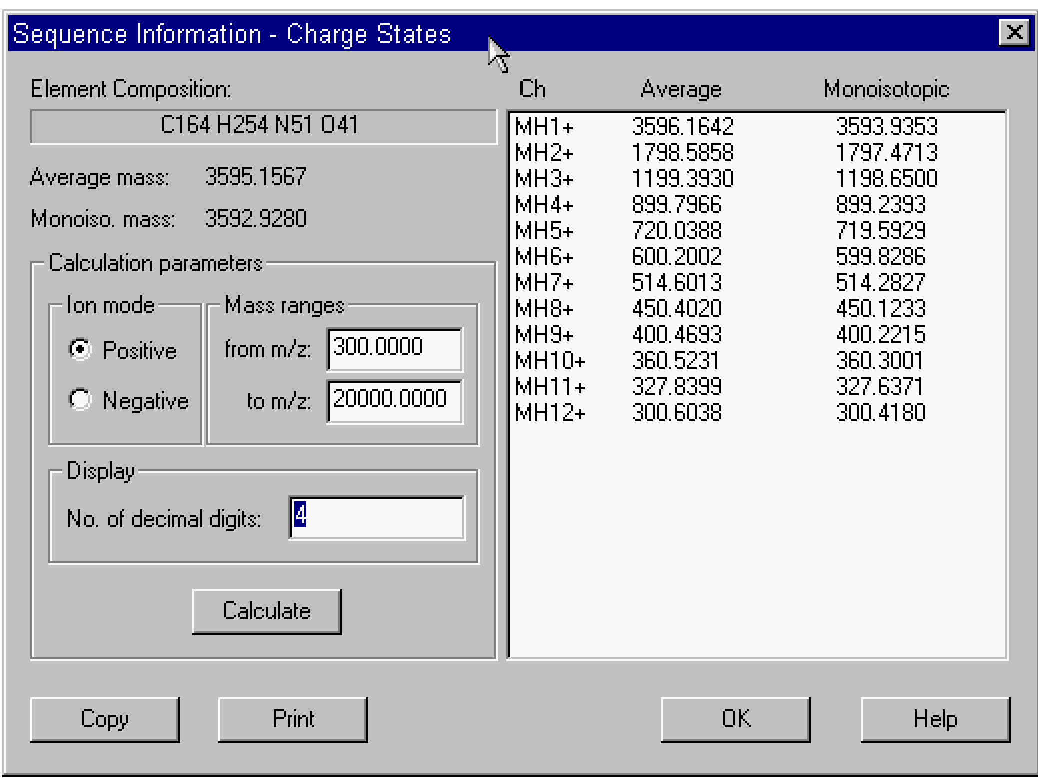


Figure B-8, Sequence Information – Charge States

Element Composition
Shows the elemental composition calculated from the whole sequence including all modifications and crosslinks.
Average mass & Monoisotopic mass
Total sequence masses shown as monoisotopic and average neutral molecular weights.
Digits
Number of post decimal positions used for mass printout.
Ion mode
Switches between the MH^+ and MH^- mode.
From m/z and **to m/z**
Calculations will be done in the mass range between *from m/z* and *to m/z*.
Calculate
Calculate the m/z values of the different charge states of the active protein structure within the specified m/z range for the specified ion mode.
Print
Use this button to print the content of the dialog box.

Copy
Use this button to copy the m/z values into the clipboard. From here they can be pasted into the Bruker NT acquisition and processing programs to allow for simple definition of calibrant masses.

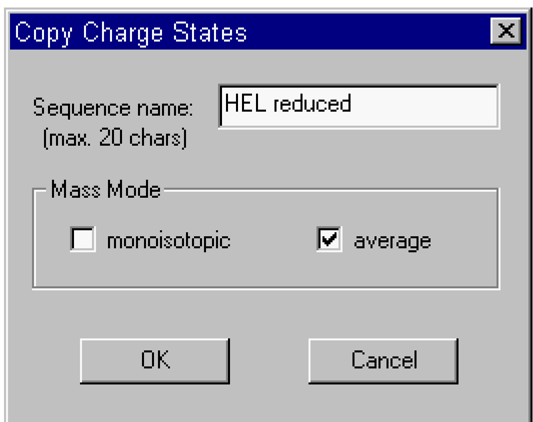


Figure B-9, Sequence Information – Copy Charge States

Here, the desired calibrant name is specified (default is the actual sequence name truncated after 20 characters) and average or monoisotopic masses are selected. Pressing the *OK* button copies the respective masses together with a descriptive header into the clipboard for further use. The entry as used in a calibration files will appear like: HEL reduced [MH]1+ avg, 14314.29070.

B.4.3. SeqInfo - Composition

Use this command to get the amino acid composition of the selected sequence.

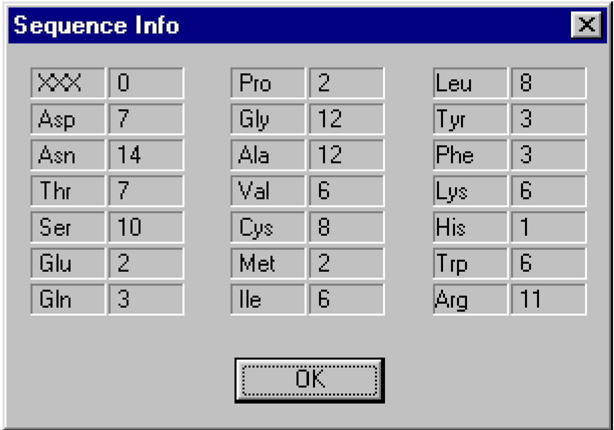


Figure B-10, Sequence Information – Composition

B.5 Standard Modifications of the Sequence

Cysteine is frequently involved in intramolecular crosslinks (disulfide bridges formed by oxidation of the reduced SH side chain functionality) of proteins and is typically subject to chemical procedures like reductive cleavage of the crosslink followed by alkylation. Such standard cysteine states and terminal standard modifications like N-terminal formylation or C-terminal amidation can be defined just by selecting the options in the combo boxes in the header of the sequence display:

Figure B-11, Sequence window toolbar

Note ***Default settings are reduced cysteine, i.e., the free thiol form, N-terminal free α-amino group (hydrogen) and C-terminal free acid. All residues in non-default states are color coded in the sequence.***

If these combo box fields are empty, please load a modifications file (*.mod) using the *modification types editor* via the menu *Edit - Modif.-Types*.

B.5.1. Cysteine state

The Cys-state can be changed by selection of a different modification in the *Cys* combo-box. Each change of the cysteine-state results in a recalculation of the sequence gross molecular weight.

Selecting the "*Oxidized*" cysteine-state automatically causes the *Crosslink Editor* dialog box to pop up and you are expected to define the disulfide crosslinks (see menu Edit - Crosslink-Types). For calculation of the gross mass of the full protein, however, this is not needed and you can just leave the dialog with *OK*. No crosslinks are defined then but the individual cysteine residues are used for mass calculations in the respective oxidized form.

Once disulfide crosslinks are defined in the *crosslink editor*, they are displayed as connecting lines. Changing the cysteine-state from *Oxidized* to another one does not result in loss of this information. Instead, crosslinks will be displayed in gray and become inactive (they are not used for any calculation). Crosslinks are restored if the state is reset again to *Oxidized*.

B.5.2. Termini

The left (N-terminal in the case of proteins, 5´ for DNA) and right (C-terminus and 3´, respectively) terminal groups of the sequence can be changed from the default states. Each change of the modification results in a gross mass calculation. If terminal residues are set to a non-default state, they become color coded in the sequence.

Note ***The choice of modifications in the combo boxes is dynamically dependent on the sequence and can be customized. E.g., the N-terminal pyroglutamylation is only available if the N-terminal residue is Gln! (see menu Modification Type Editor)***

B.6 Visualize particular Residues and Sequence Motifs

B.6.1. Search for Sequence Motifs

If you need to search a known sequence motif within the protein sequence, under *Search – Sequence* you type it in and all sequence stretches containing the motif are highlighted in the sequence.

B.6.2. Highlight Aminoacid Residues and N-Glycosylation Consensus

If you need to screen for particular aminoacid residues, e.g., to estimate the distribution of cleavage sites for trypsin, you can color code them in the sequence display by *View – Show residues* and select from the residues listed in the menu. Also included is the N-glycosylation consensus sequence motif **N X S/T**.

B.7 Export Sequence to MS/MS Spectrum Display

With the Bio button sequence data is sent to the spectrum display for the calculation of MS/MS fragment ions and the visualization of the match between a sequence and an MS/MS spectrum from the APEX, ESQUIRE or FLEX. If only a part of the sequence is selected, such as a proteolytic peptide, only this part will be exported to the spectrum tree view.

B.8 Printing Data

The button allows printing of the active sequence.

S Modify a Sequence

The primary structure of proteins is defined to a great extent by the genetic code, which results in a specific amino acid sequence of the respective expressed protein. However, most proteins express their proper function only after certain modifications were introduced to the amino acid side chains (keyword: post-translational modifications, PTMs) and after the highly specific formation of disulfide bonds, which stabilize a protein thermodynamically and ensure a proper conformation. The definitions of primary structure changes, such as protein modifications and crosslinks are described in this chapter.

S.1 Modifications S-2
S.1.1. Modify Residues in the Sequence S-2
S.1.2. Define New Residue Modifications S-5
S.2 Crosslinks S-7
S.2.1. Edit Crosslinks S-7
S.2.2. Defining New Crosslink Types S-9

S.1 Modifications

There are three **locations** for protein modifications:

N-terminus

C-terminus

Side chains of all aminoacid residues in the peptide chain.

Four **types** of modifications are distinguished in the program:

Local e.g.: C^{118} is palmitylated, or: the N-terminus is formylated

Global e.g.: all C residues are carbamidomethylated

Excluded e.g.: all C residues are carbamidomethylated, except C^{17}

Optional e.g.: either of all the S, T or Y residues are possibly phosphorylated

S.1.1. Modify Residues in the Sequence

1. Double-click the residue to be modified within the sequence with the left mouse button. The *modification editor* (Figure S-1) is opened with a selection list of all modifications available for the residue you have clicked on. (Alternatively select *Edit: modifications* to open the dialog.) SequenceEditor loads the specified modification file and monitors its name in the field situated on the top of the dialog.

All currently applied modifications are listed in the *Modifications present* field. The listbox *Modifications available for* shows all modification types available for the selected residue (here: Arginine, R). If other residue types are also to be modified, select it in the combobox *Select residue type*.

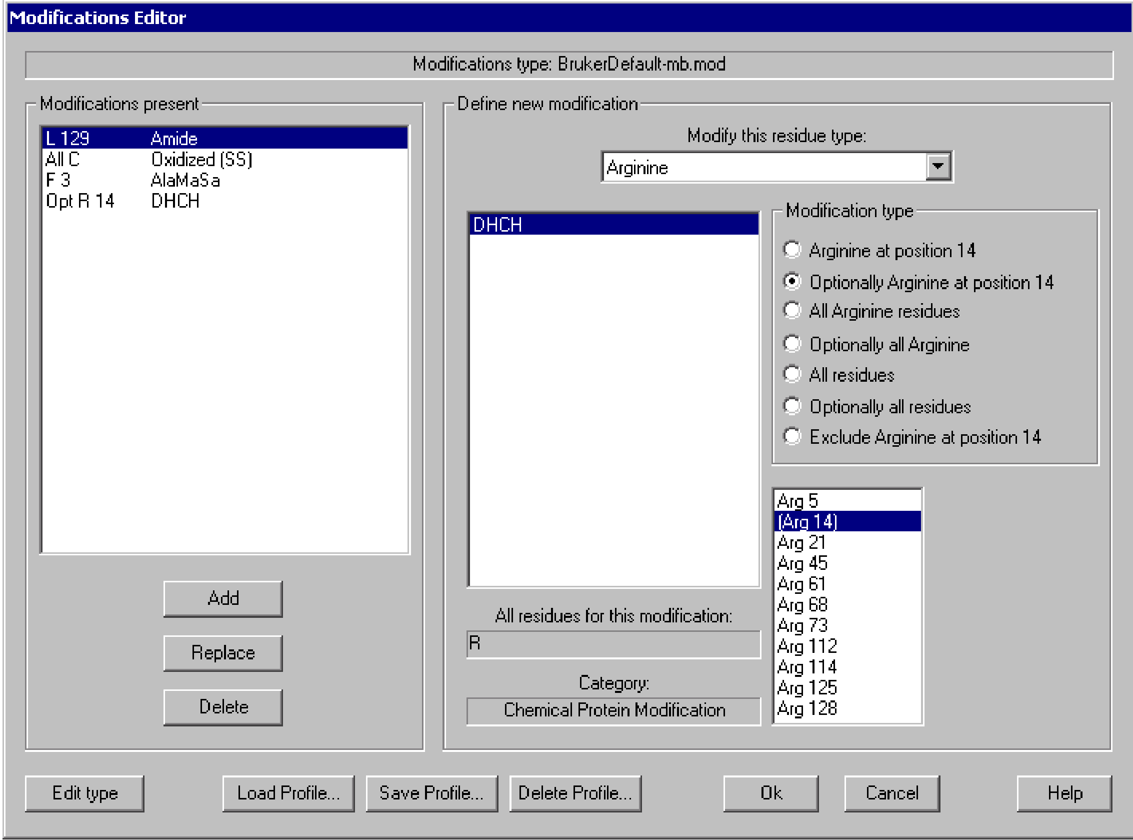


Figure S-1, Modifications Editor

2. In the *Mods types* box specify the type of the modification to be applied:

Arginine at position 14 specifies a **local** modification to be selected. This is the residue the user clicked on to open the *Modifications Editor*. You can select any arginine residue from the sequence clicking on the respective entry in the listbox situated below of the *Mods type* list.

Note ***If a particular residue is already modified or excluded it is shown in parentheses and the selection is not possible.***

All Arginine residues and **All residues** are used to specify a **global** modification. All arginine residues in the sequence or all residues for which the selected modification can be applied will be modified, respectively. All residues is only relevant if the

modification can be applied to several residue types, e.g., S, T and Y residues can be phosphorylated.

Optionally All Arginine and **Optionally All Residues** specify an **optional** modification either on all arginine residues or on all residues for which the selected modification is defined. Optional modifications are defined like global. They are used if you want to screen a sequence for a putative modification.

Optionally Arginine at position 14 specifies an optional modification at R14 only.

Exclude residue position excludes any particular residue from being modified.

3. Select a modification in the *Modifications available for* box. When clicked on, the valid acid types appear in the *All residues for this modification* field. In the field beneath the category of the specified modification is monitored.

4. Press the *Add* button to add the modification to the modifications list.

You can remove any applied modification by means of selecting it in the *Modifications present* list and pressing the *Delete* button. If you want to replace an existing modification with an another one select the present modification in the list and the new one in the *Modifications available for* box and select the *Replace* button.

5. Exit the dialog with the *OK* button to apply all changes or with the *Cancel* button to ignore all changes made in the *Modification Editor*.

The modification types can be defined, updated or removed using the *Modification Types Editor*. Select the *Edit Modification Types* button to call this dialog.

S.1.2. Define New Residue Modifications

By means of the *Modification Type Editor* (Figure S-2) any modification can be defined. To enter the editor, select the menu element *Edit:Modif.-Types* or click on the *Edit Modification Types* button in the *Modifications Editor*. The currently selected modification file (*.mod) is shown at the top of the dialog.

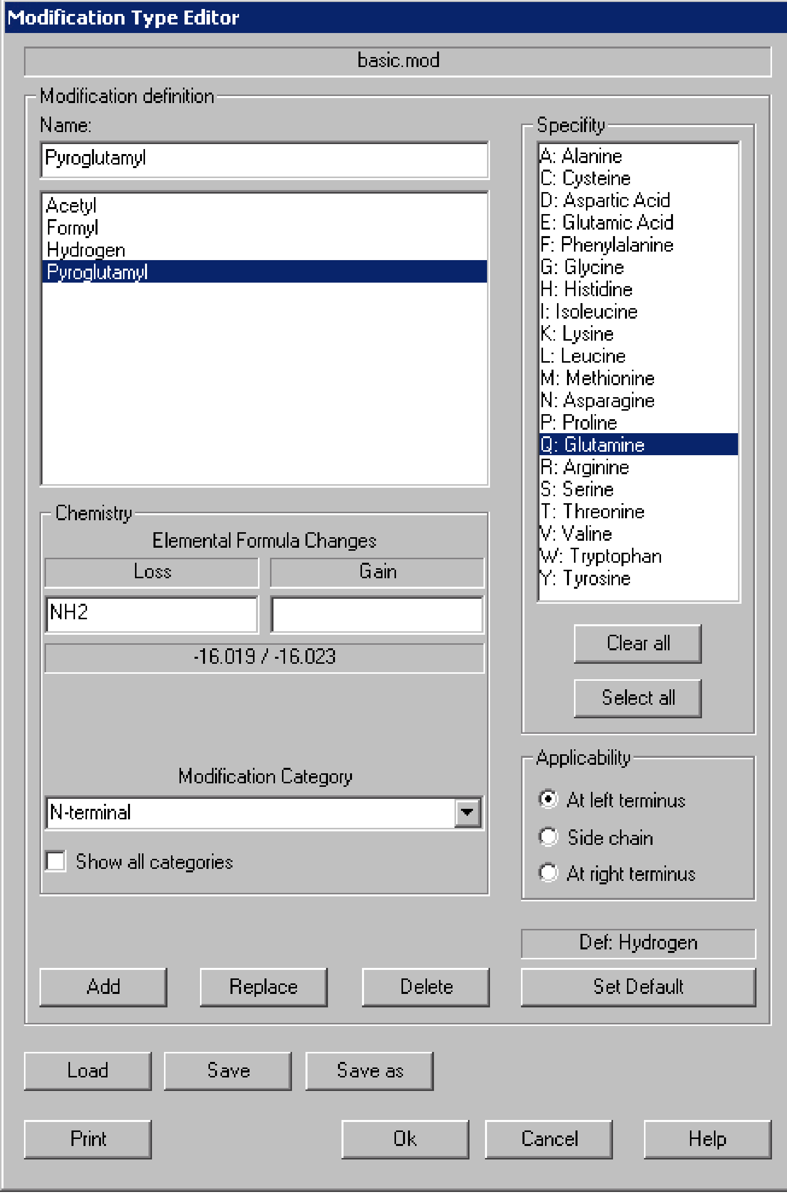


Figure S-2, Modification Type Editor

Note: The default modification file, BrukerDefault.mod, is supplied with the program. It can not be altered by the user. To permanently define and store new modifications you need to save it under another name first (see: *Load*, *Save* and *Save as* buttons).

Follow these steps to define a new modification:

1. **Check whether the modification is already defined**. Select *Show all categories* if the category is unclear to you at this point and scroll through the list of defined modifications. If a name sounds right, select it and verify the full definition of that modification. If you did not define the modification previously, deselect *Show all categories* and continue with step 2.
2. **Type the modification *Name***.
3. **Define *Elemental Formula Changes*** in terms of lost and gained atoms
4. **Define the residue *specificity***, i.e., the aminoacid residues, which are susceptible to this modification.
5. **Define the *Applicability***, i.e., the modification site (N-, C-terminal or side chain)
6. **Select a defined modification category (like Chemical modification) or define a new one.** This is only to give larger sets of modifications a logical structure, which allows simpler modification retrieval. For each category, one modification can be ***Set Default***, like reduced (SH) for cysteine modification category (such modification appears on the top of a alphabetically sorted list). You may enter the charge on the side chain, but in BioTools this information is not used yet.
7. Press the ***Add*** button to add the modification to the mod. list of the current session and ***Save*** to store it permanently into the selected *.mod file.

In the list on the left side of the dialog all currently defined modifications for the selected category (see *Modification Category* field) are shown. If the *Show all categories* box is checked all modifications specified in the modification file are listed. The *specificity* list on the right side monitors the amino acid types for which the selected modification is available.

If you select a modification the *Loss* and *Gain* fields show the changes in the chemical formula caused by it and underneath the gross monoisotopic/average mass change due to the modification.

S.2 Crosslinks

S.2.1. Edit Crosslinks

Crosslinks connect different residues within a single chain through the side-groups of the building blocks. The standard and most prominent way to link different parts of a protein with each other is the disulfide bond, which links cysteine (C) residues. Crosslinks can be defined in *Crosslink Editor* and are used for intact protein molecular weight and enzyme digest calculations. They are shown on the screen as colored lines connecting the respective residues. The color of the crosslinks can be defined under *Options - Colors*.

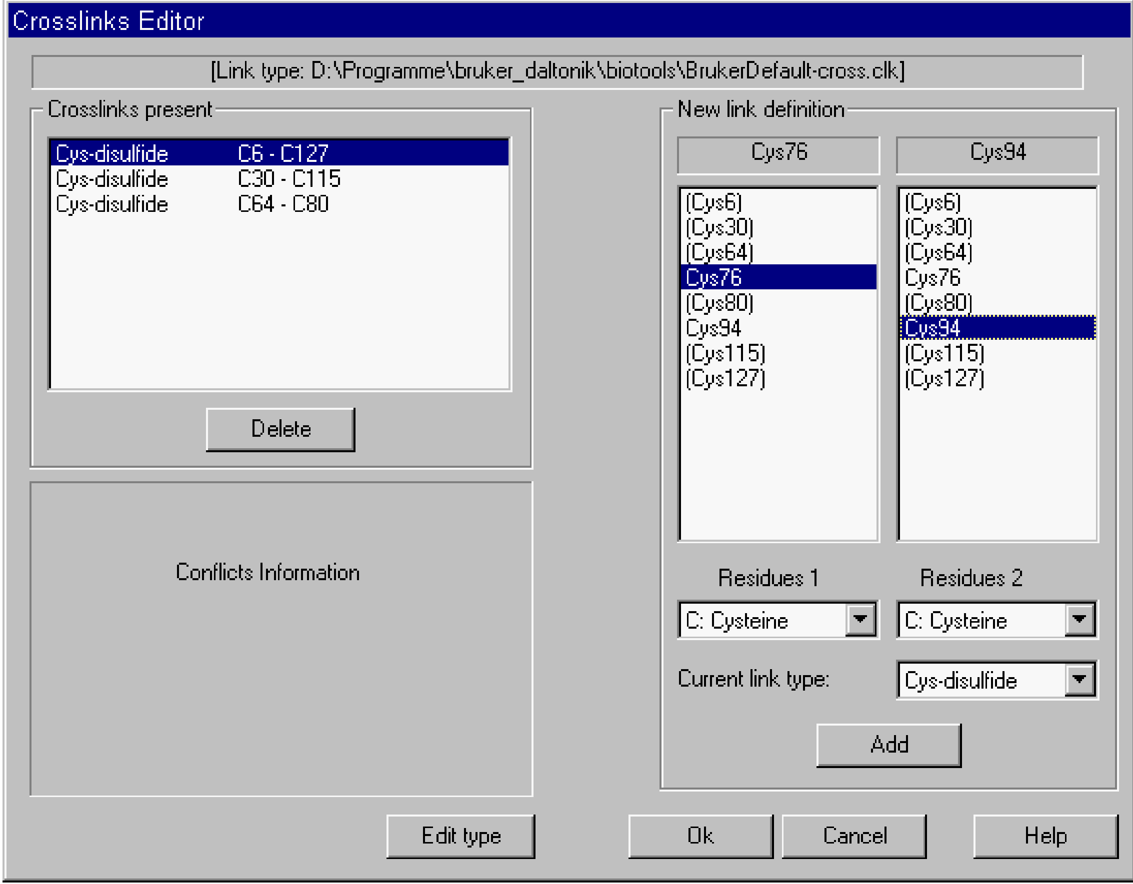


Figure S-3, Crosslinks Editor

Example crosslinks are contained in the default crosslink file *BrukerDefault.clk* and new links can be defined in the *Crosslink Type Editor* (Figure S-3). The default crosslink file can not be changed within the *Crosslink Editor*. Only a crosslink file created by the user can be edited and saved. ***Link type*** shows the currently selected crosslink file.

Crosslinks present shows the list of all specified crosslinks in the active sequence.

Conflicts information: Any detected conflicts in the editor, e. g., selecting of an already modified or crosslinked residue, are monitored in this field.

Follow these steps to define a crosslink:

1. **Select Residues 1 and Residues 2,** i.e., the residue types, which are to be crosslinked, ***Cys*** and ***Cys***.
2. **Select Current link type to specify** the crosslink to be added, ***Cys-disulfide***. This list is dynamically dependent on the defined residues in step 1, therefore **do not exchange steps 1 and 2**.
3. **Select the two specific residues in the residue** lists, like ***Cys14*** in the left selection list and ***Cys17*** in the right list. Residues, which already linked or modified cannot be additionally linked and are shown in parenthesis.
4. Press the ***Add*** button to add this link.
5. Repeat steps 3-4 for every crosslink of that specific type.
6. Press ***OK*** to accept the present links.

Edit link type: Any link can be dynamically defined, replaced or deleted. Press this button to edit the crosslink types in the *Crosslink Types Editor*.

S.2.2. Defining New Crosslink Types

Crosslink definitions are stored in *.clk files, which can be loaded in the *Crosslink Types Editor*. BioTools is shipped with the *BrukerDefault.clk* file, which can not be modified (*Add* button is grey). If you store it under a different name (*save as*), the *Crosslink Types Editor* allows you to specify new crosslinks and to store them in the crosslink file.

The mass change caused by a defined crosslink is considered for the calculation of the total mass of the sequence or in computer digests of the sequence.

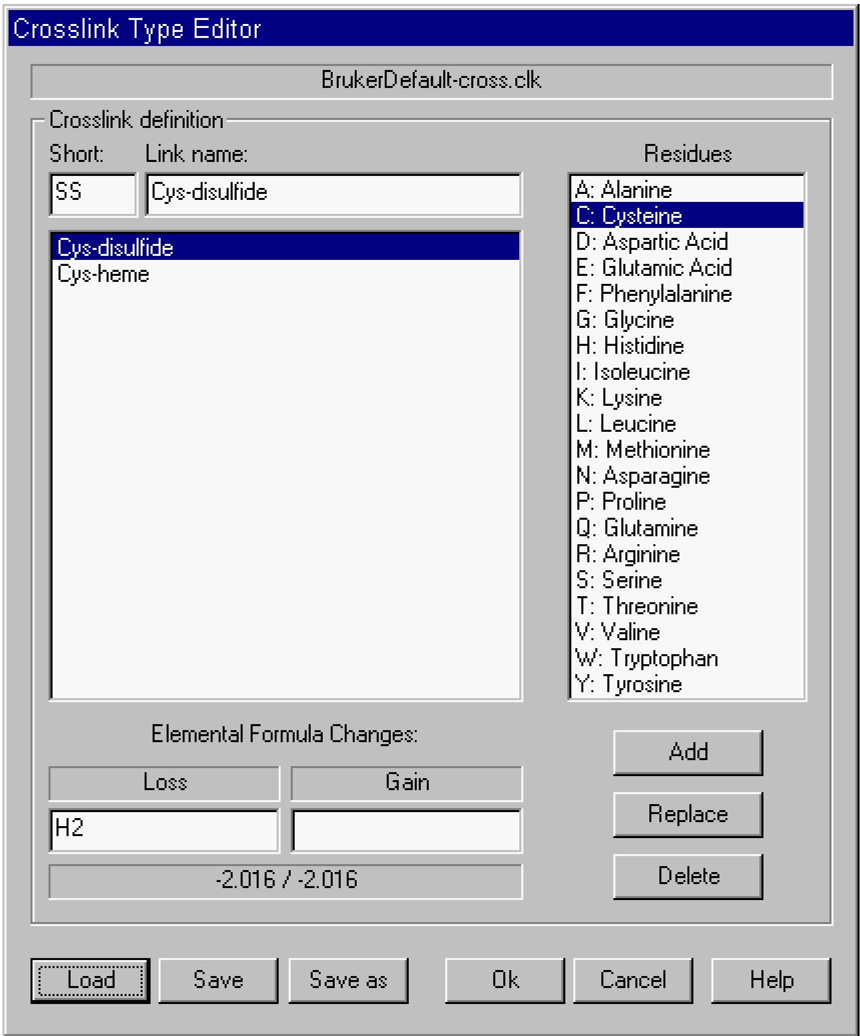


Figure S-4, Crosslink Type Editor

Follow these steps to define a new crosslink type:

1. ***Load*** and select the *.clk file in the BioTools folder, if the right file is not loaded automatically (initially, save the default BrukerDefault.clk file under a new name **– Save as**).
2. Type the ***Link name*** and the 3-letter acronym (***Short***)
3. Type the ***Elemental Formula Changes*** - Losses and Gains - of the total of both involved residue side-chains
4. Select both ***Residues***, which are to be crosslinked.
   Note ***If it is a symmetric link, like a Cys-Cys crosslink, only one residue type is selected.***
5. Push the ***Add*** button and the new crosslink type appears in the crosslink list and the resulting mass of the crosslink appears underneath the elemental formula fields as monoisotopic/average mass.
6. Press the ***Save*** button to add this crosslink permanently to the *.clk file and ***OK*** to finish editing crosslink types.

P Protein Digests

P.1 Introduction P-2
P.2 Perform Enzymatic Digest P-3
P.3 Format the Digest Results P-5
P.4 Export Digest Results to Spectrum...P-6
P.5 Search for Masses in a Sequence...P-8
P.5.1. Search for Unexplained Masses after MASCOT search P-8
P.5.2. Search for Unexplained Masses in a Local Sequence.......................P-12
P.5.3. Screening for Mutations and Modifications using MS/MS Spectrum..P-13
P.6 Edit New Enzyme Specificities P-17

P.1 Introduction

Beyond the calculation of intact sequence molecular weights, the calculation of molecular weights of peptides resulting from digestion with proteolytic enzymes is important for protein structure elucidation and protein identification (words: peptide mapping, mass fingerprint).

Digest agents fragment a particular protein (or DNA or RNA) into smaller pieces after recognition of a particular amino acid residue or a sequence motif comprising of multiple residues. The task in these analyses is to match the mass spectrum with all its peaks to the protein sequence + modifications to account for as much structural information as possible. Artifacts make this task sometimes difficult: unspecific or irregular digests, unknown modifications, contaminations, autoproteolysis of the cutting enzyme, etc.

Calculation results can be used locally, like for the prediction of enzymatic peptide fragment masses or in conjunction with particular datasets: MALDI or LC-MS fingerprints, etc.

Mass fingerprints can be used analytically in two directions in conjunction with experimental data:

Generate a proteolytic digest pattern *in silico* and screen for the respective peaks in the spectrum. This approach allows the suggestion of signals, which escaped automatic or incomplete manual peak annotation procedures up to that point. In other words: it helps to extract more information from a spectrum.

Take the spectral information and search in the sequence, to what extent it accounts for the experimental data. This approach leads to the identification of unexpected digest products, like incompletely or unspecifically digested peptides, which can not be calculated in a simple way otherwise.

BioTools supports both ways to analyze peptide maps for an in-depth correlation between a protein structure and a mass fingerprint.

P.2 Perform Enzymatic Digest

Generate or load the protein sequence into the SequenceEditor, which you would like to digest into the program. Use the button to perform an *in silico* enzymatic digest (Figure P-1).

Note ***This button is also available in the*** **BioTools toolbar*****, but only after selection of the appropriate*** **Digest Matches** ***node in the treeview. Therefore, a theoretical digest can be done on the imported sequence directly without the need to open the SequenceEditor manually.***

Along with the enzyme name, the mass range and the number of missed digest sites (*Partials*). can be defined to control the result list of digested peptides, amongst the other options described in more detail below.

Both monoisotopic and average masses can be toggled in the result list. The results can be sorted by numbers (indices) or by masses. Fragments that are pseudo-independent but connected by crosslinks are listed in the table separately.

Enzyme Select the enzyme of choice.

Note: If the desired enzyme/specificity is missing, define it in the *Enzyme Types Editor*, which you enter from this dialog pushing the *Edit Enzyme* button.

Define the Digest Conditions

Partials The number defines the maximum allowed number of internal digest sites in the peptide output according to enzyme specificity. This accounts for the degree of completeness of a digest.

Limit mass range If selected, the output will contain only peptides within the specified mass range (here from m/z 1000 to m/z 3000).

Deuterium exchange If selected all calculations are done with deuterium instead of hydrogen for all exchangeable hydrogen atoms (i.e. the ones linked to heteroatoms).

Do not cleave if modified If selected there will be no digest at the chemically modified enzyme substrate residue (e.g. Lysine acetylation in case of trypsin).

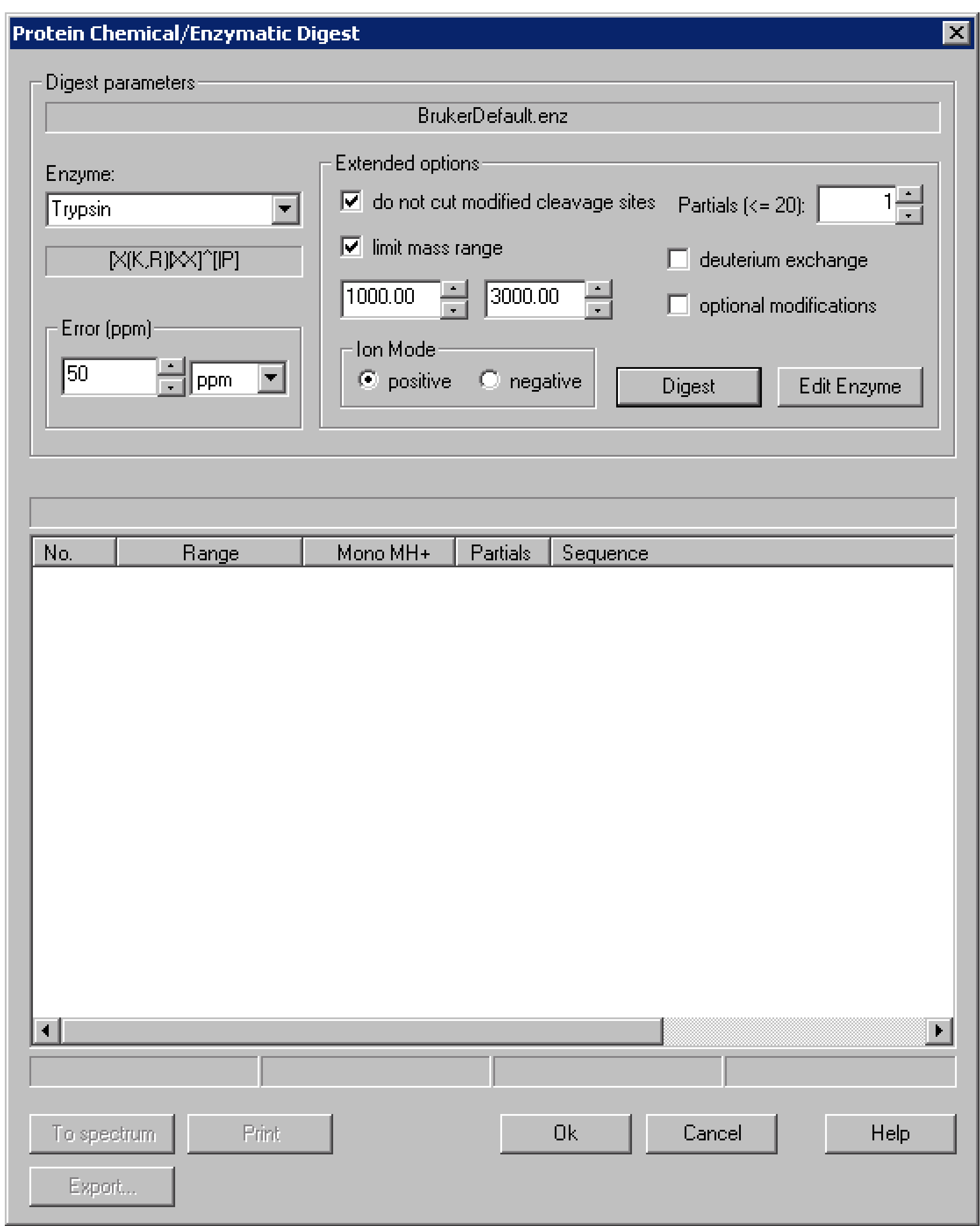


Figure P-1, Enzymatic digest

Optional modifications If there are optional modifications defined for the protein sequence this selection results in the calculation of all permutations of modifications within the conditions of the other options. Proteolytic peptides, which contain a modification are marked in the result list with an asterisk (*) and the number of modifications contained within the peptide. This is helpful to identify the actual state of modification of residues based on peptide molecular weight, which are in question.

Ion mode Allows to select between the calculation of MH^+ or $M\text{-}H^-$ of the peptides.

Error Specify the mass error for matching an experimental peaklist. This is only needed if you want to send the digest results ***To Spectrum.***

To spectrum The theoretical masses of the digest are sent to and visualized in the spectrum display. Matching peaks will be labeled and the others will be highlighted. It allows to screen the spectrum for peaks, which escaped peak annotation.

Digest results The cleaving results are displayed in the report listbox (Figure P-2).

Enzyme file Shows the currently selected enzyme file. The enzyme information available for the system are defined in the *BrukerDefault.enz* file. The default enzyme file can be modified only in the *Enzyme Types Editor* after saving it under a different name.

P.3 Format the Digest Results

Sort the List Click on the ***No.*** field of the listbox header to sort the list by *Indices*, *descending* or *ascending Masses*.

Define Ion Type Click the ***Mono*** field to select the display of the ***Mono****isotopic neutral*, ***Mono****isotopic* ***MH***$^+$ *(or* ***MH***$^-$*)*, ***Ave****rage neutral* or ***Ave****rage* ***MH***$^+$ *(or* ***MH*** $^-$*)* mass. The MH^+/MH^- display depends on the *Ion Mode* (positive or negative).

Note: In order to obtain the mass values of all 4 mass types for a single peptide simultaneously (see the bottom of Figure P-2), click on the entry No., in this case „1".

The peptides connected with crosslinks (#) are shown at the bottom of the digest results list. The corresponding marked (#) single chain peptides are contained in the main list.

Sorted by indices

No.	Range	Mono	Sequence
1	[1- 5]	606.367	KVFGR
2#	[6- 13]	835.387	CELAAAMK
4	[15- 21]	874.411	HGLDNYR
5#	[22- 33]	1267.596	GYSLGNWVCAAK
6	[34- 45]	1428.645	FESNFNTQATNR
7	[46- 61]	1753.830	NTDGSTDYGILQINSR
8#	[62- 68]	935.365	WWCNDGR
9	[69- 73]	517.268	TPGSR
10#	[74- 96]	2334.096	NLCNIPCSALLSSDITASVNCAK
12	[98-112]	1675.796	IVSDGNGMNAWVAWR
13	[113-114]	289.157	NR
14#	[115-116]	249.109	CK
15	[117-125]	1045.537	GTDVQAWIR
16#	[126-129]	446.219	GCRL

Linked peptides:

#/#	range	range	residues	mass
2/16	6-13	126-129	Cys6-Cys127	1281.606

Mono = 606.367 | MH+ = 607.374 | Ave = 605.743 | MH+ = 606.750

Figure P-2, Digest results

P.4 Export Digest Results to Spectrum

The *digest results list* can be exported to the digest spectrum with the ***To Spectrum*** button. All peptides, which match the annotated peaks in the spectrum within the specified mass error are annotated in black. The annotation contains the experimental peak mass and the sequence range. If there is no experimental peak matching a particular theoretical mass, a red label of the theoretical mass together with "Missing" is written to that mass in the spectrum.

Note *The mass type selected in the digest result list (Mono MH+ for MALDI, etc.) MUST match the data type of the spectrum for proper display! However, crosslinked peptides cannot be transferred to the spectrum!*

This is a helpful guide to the eye to go "peak hunting". But be careful that you don´t trick yourself by assigning a meaning to noise!!

Hint: In BioTools you can add missing peaks found by this approach to the peaklist. You switch to **Add peaks mode** in context sensitive menu accessible via right mouse button while the cursor is over the spectrum. You expand the overview window so that you can use it and zoom in to comfortably select peaks (drag the mouse across the mass range of which you want to assign the maximal data point). You add peaks by the drag operation in the Spectrum Window and change the zoom range in the Overview Window either by dragging the zoom range box or using the cursor keys after a mouse click to get the Overview Window active.

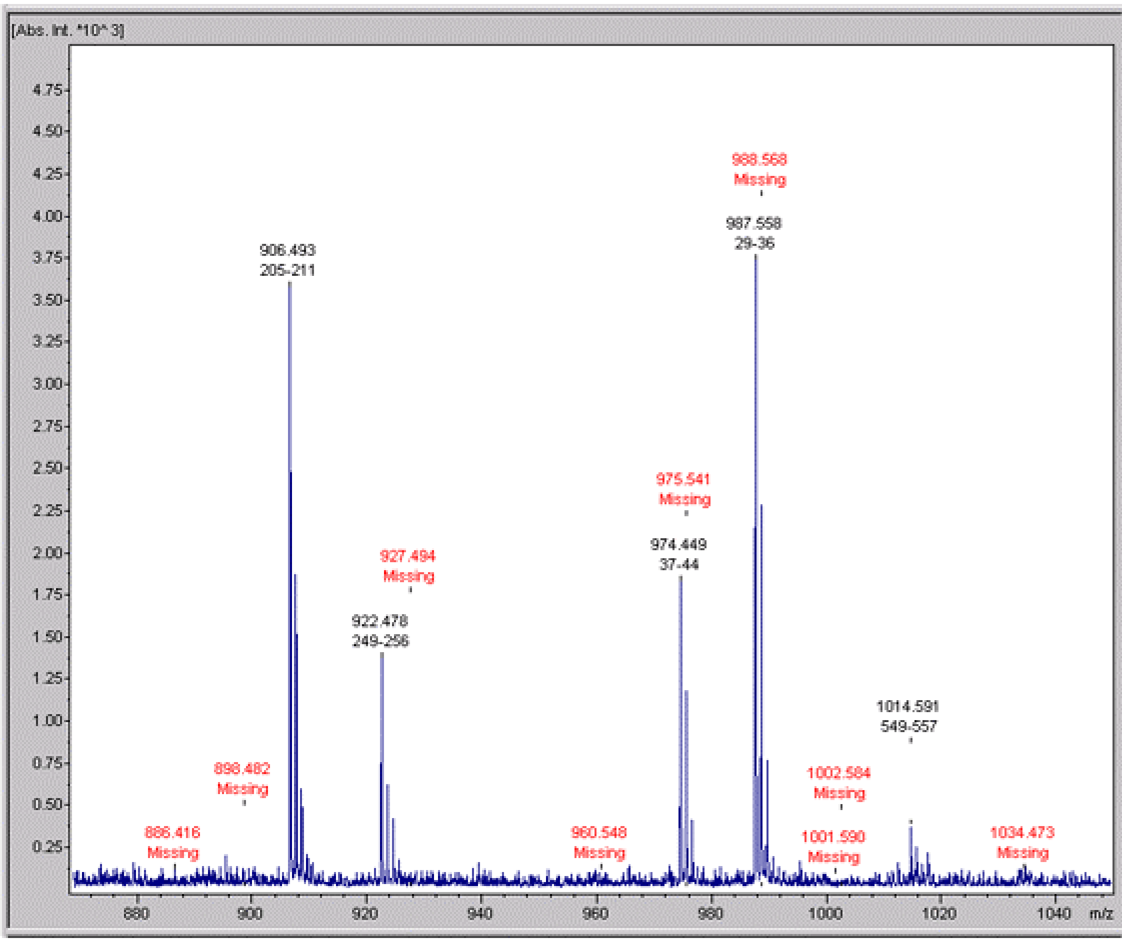


Figure P-3, Digest results in BioTools

P.5 Search for Masses in a Sequence

Take the spectral information and search in the sequence, to what extent it accounts for the experimental data. This approach leads to the identification of unexpected digest products, like **incompletely or unspecifically digested peptides**, or allows screening for **unknown modifications** in the sequence and even **point mutations**.

E.g., trypsin digests typically suffer unspecific digests His or the Chymotrypsin digest sites Tyr, Phe, Trp, Ile and Leu. On the other hand, missed digests are typically observed if the recognized residues Arg and Lys are adjacent to Asp or Glu residues involved in ionic bonds.

Depending of the source of sequence information, two different approaches are possible: 1. After a Mascot Search, the relevant spectra are imported into BioTools and sequence is available for this search, or 2. A sequence is available from any source, dumped into the sequence display and used.

P.5.1. Search for Unexplained Masses after MASCOT search

Initially, a MASCOT search of digest data is done. From the result browser page, either all (***Get All***) or selected entries (***Get Hit(s)***, e.g., "***1-3,7***") can be imported into the tree view of BioTools containing matching sequences and peaks.

Double-click on the **Digest Matches** entry in the tree view of the sequence you want to further work on, or click the **Search for Masses** button (Search for Masses). The whole sequence, together with all modifications, mass errors and enzyme information as specified for the MASCOT search are transferred to the SequenceEditor, which automatically opens the ***Search for Mass*** dialog.

Note ***Modification or enzyme information are passed to the SequenceEditor only, if the definition file of MASCOT on the server and under BioTools (enzymes and mod_file,*.enz and *.mod, respectively) contain identical names for the same entry!***

Click the ***Select unmatched*** button, to search only those peaks, which were not yet assigned. Prior to the search other options, like the mass error or modifications can be changed. Only masses, which are contained AND selected in the mass list will be searched.

Note ***In this search the partials setting is completely ignored and cannot be defined to allow for the display of unexpected results.***

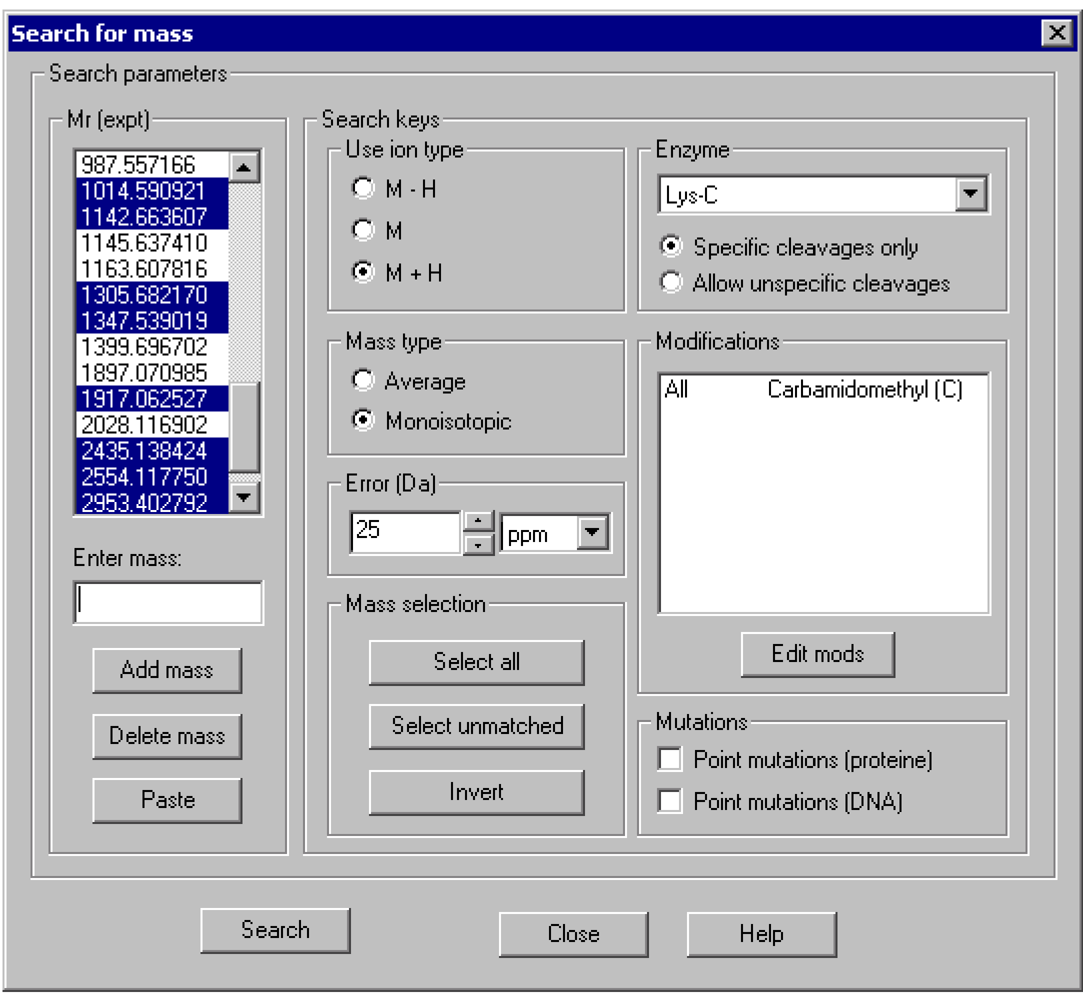


Figure P-4, Search for Mass dialog

In the resulting ***Mass search results*** list, the individual masses can be selected as complete set (***Select all***) or individually.

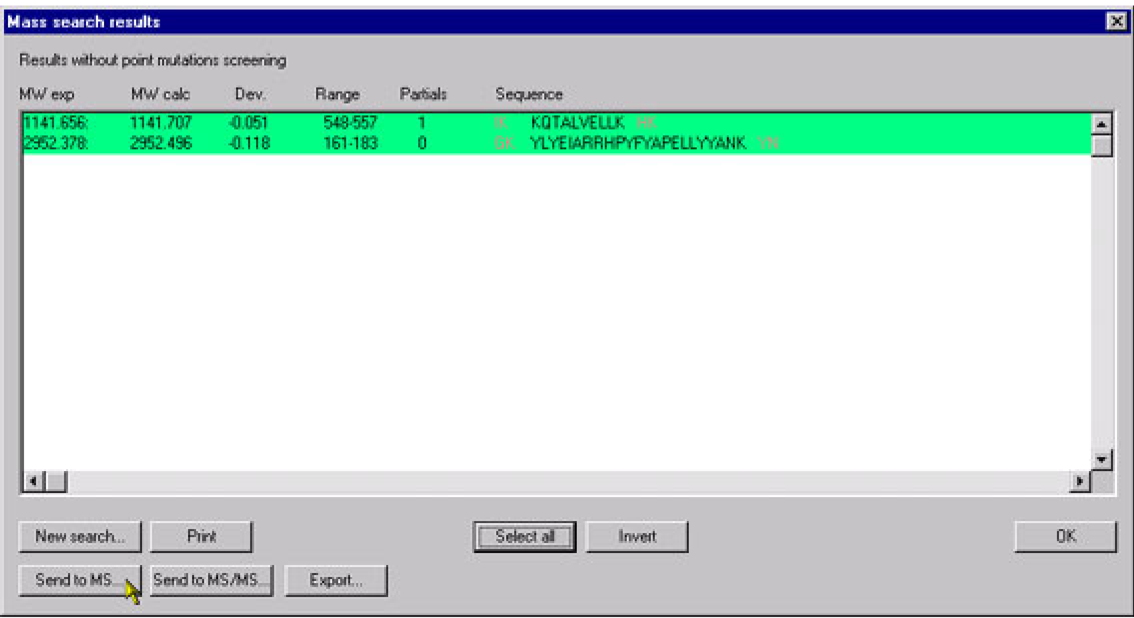


Figure P-5, Mass search results

The selected masses are transferred to the spectrum tree view with the ***To spectrum*** button. They appear as a new ***Mass Search Matches*** entry under the initially selected sequence name. These masses can be visualized in the spectrum either by clicking on ***Mass Search Matches*** or a click on the sequence name, which the displays all peaks in the spectrum, which are accounted for by all subentries of the sequence. If additional search conditions should be evaluated, the ***New search*** button brings you back to the search page to modify the conditions, e.g., select ***Check fits*** in the ***Search for mass*** dialog to allow the output of unspecific digests. Manual investigation of such a list is mandatory and chemical intuition/knowledge is required to make meaningful selections.

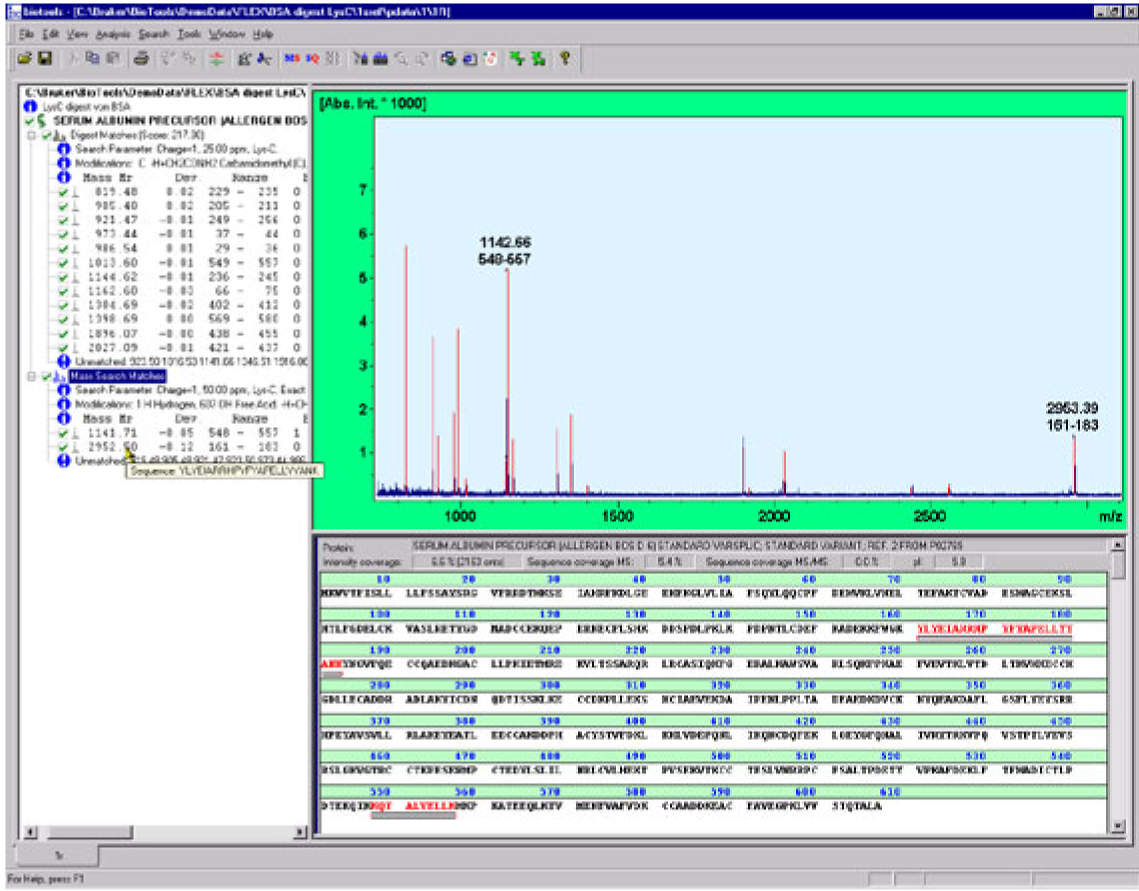


Figure P-6, Results for mass search

Finally you end up with a spectrum, which has been accounted for by a theoretical sequence.

P.5.2. Search for Unexplained Masses in a Local Sequence

If a sequence is locally available in the SequenceEditor for further work, this can be used directly for evaluation of the spectrum, without the need to do library searches. This may be the case for novel proteins being sequenced in your lab or quality control applications for recombinant proteins with a special set of modifications, etc..

A fingerprint spectrum is active and the mass list is copied into the clipboard: Open the **MASCOT Fingerprint** search dialog in BioTools and push the **Copy Masslist** button.

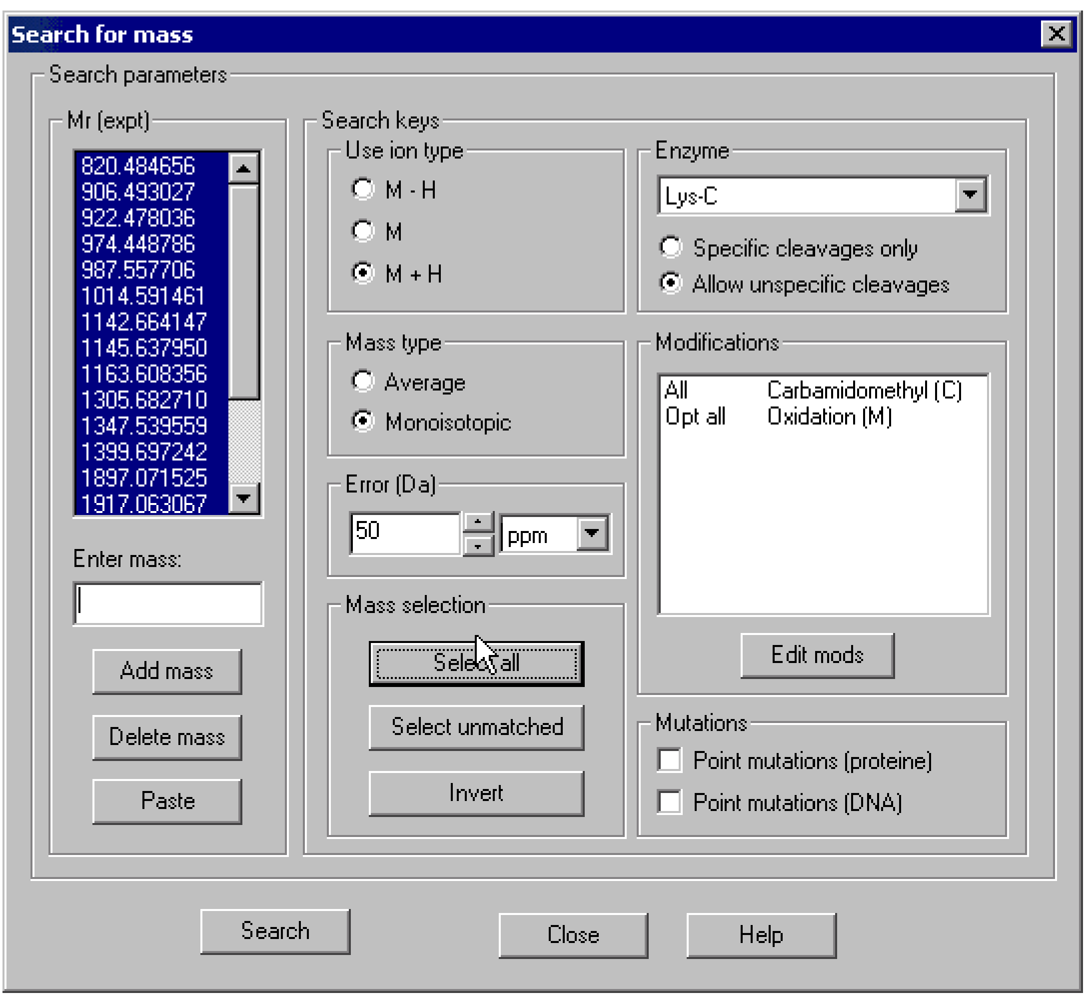


Figure P-7, Search for Mass dialog

Load the particular sequence into the SequenceEditor and define all modifications.

Open the ***Search for mass*** dialog, discard of the previous mass list with ***Select all*** and ***Delete mass*** and push the ***Paste*** button to get the actual peak list from the clipboard into the search dialog. Press ***Select all*** to select the imported peaks for a search and select **M+H** as **ion type** of a peaklist copied from a MALDI fingerprint, M if the peaklist is of neutral mass type, etc..

Define all relevant options, like enzyme and mass error and press ***Search*** to calculate the matching peptides.

Further actions in the ***Mass search results*** list are as described in the previous chapter.

P.5.3. Screening for Mutations and Modifications using MS/MS Spectrum

If simple assumptions did not allow finding reasonable suggestions for the identity of a peak in a digest, an MS/MS spectrum must be acquired and analyzed. You can use the described **Search for mass** operation to extensively screen for sequence errors, point mutations and modifications in the protein sequence, which was identified based on the other peaks in the digest.

The option **Mutations** allows – on a single selected peptide only – to screen the sequence and allow **point mutations** or **sequence errors** in addition to any combination of optional modifications. Of course, you can not identify the valid sequence in the huge resulting list of peptides. Therefore the whole result list is selected in the Mass search window and transferred to the respective MS/MS spectrum using the **Send to MS/MS** button.

Point mutations (proteins) allows to screen for any single amino acid exchange.

Point mutations (DNA) allows to screen only for those single amino acid exchanges, which are permitted by a point mutation (SNP, single nucleotide polymorphism) on the DNA level following the genetic code (Appendix). Such a list may be shorter and helpful in very difficult cases.

Procedure: As a test example, use the Flex demo data sets *BSA digest LysC* and the PSD data set *Bsa1305_LysC.FAST*. Manually load the sequence *BSA Mutant D405Q - example for mutation screening with PSD 1305* into the SequenceEditor. This is an artificial BSA sequence containing a single amino acid residue exchange: Asp-405 was „mutated" into Gln-405, the Lys-C peptide m/z 1305 maps onto this position in BSA. Therefore, 1305 cannot be mapped onto the „Mutant" sequence, if you do not allow for mutations.

1. Select the one peptide of interest from the digest mass list in the **Search for mass** window (see Figure P-8). You can also manually enter the mass and select it, if the mass list is not available right now.

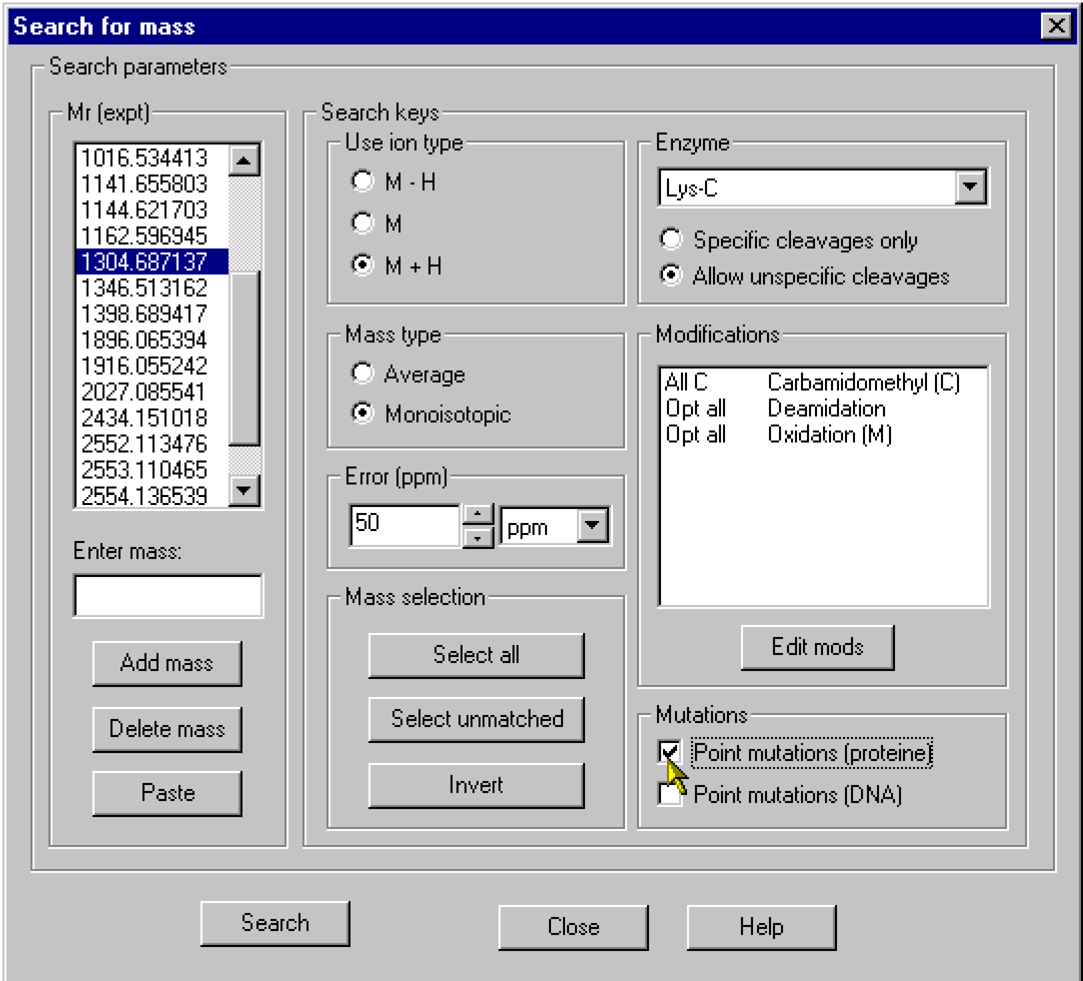


Figure P-8, Search for Mass dialog

2. Specify all suitable modifications (as optional, typically) and a mutation option.
3. Select the respective MS/MS spectrum in BioTools that it becomes the active spectrum.
4. Do the search and press the **Select all** button (see Figure P-9)

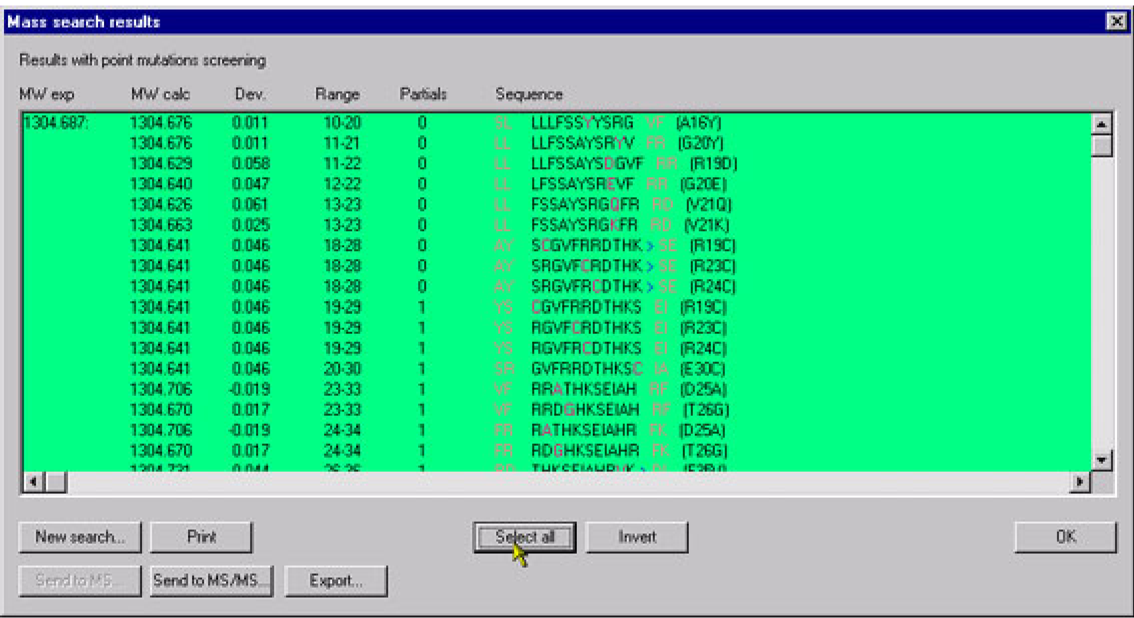


Figure P-9, Mass search results

5. Send to MS/MS transfers the result list of peptides into BioTools, where it is scored for best fit with the spectrum, the top scores are on top of the list and ready for visual inspection (see Figure P-10).

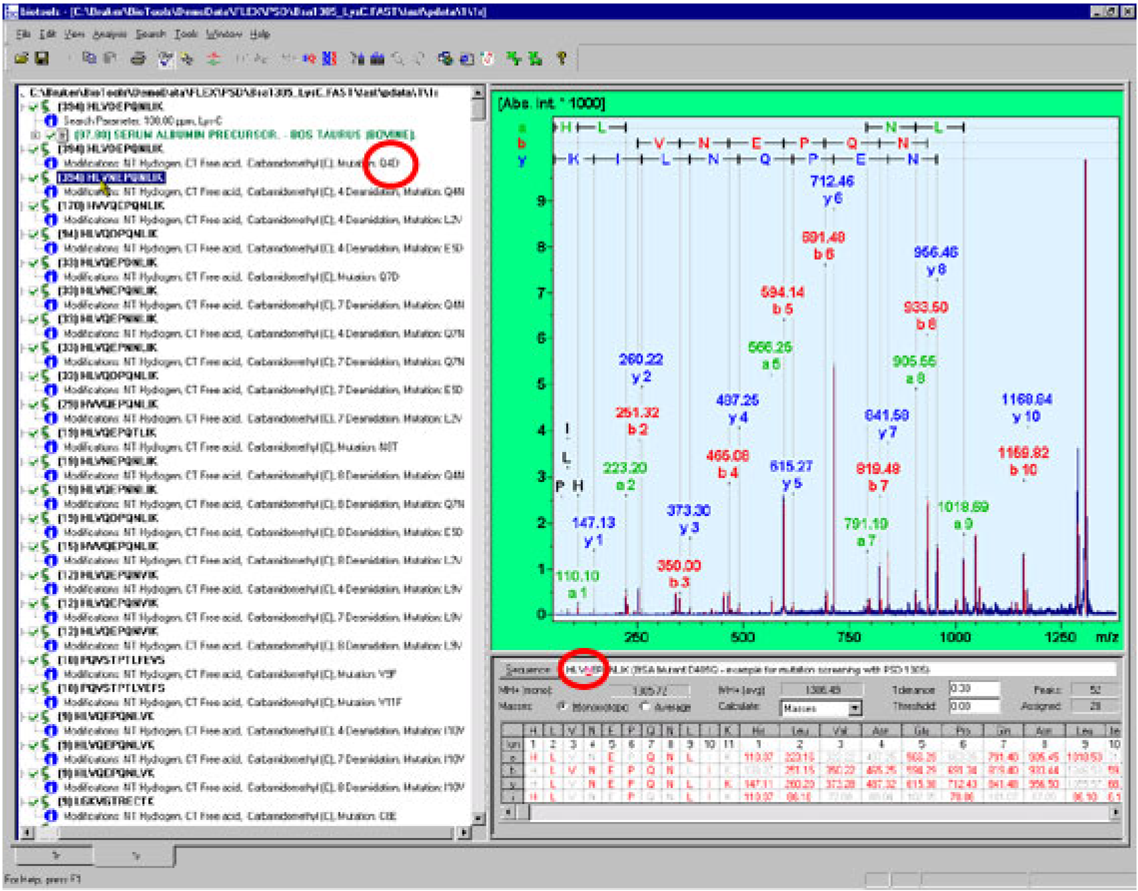


Figure P-10, Results in BioTools

In the tree view of BioTools, mutations in the listed peptides are specified, e.g., as Q4D, which indicates it was necessary to change Gln-4 of the peptide into Asp-4 for a match between spectrum and sequence. Mutations are also highlighted in the sequence text underneath the spectrum as underscore.

P.6 Edit New Enzyme Specificities

The Enzyme Specificity Editor (Figure P-11) is used to define proteolytic enzymes or chemical digest agents. The enzyme name, recognition sequence, disabling sequence properties and the elemental formula change due to the digest can be specified for a new entry. Recognition patterns for exceeding 8 residues can be defined. The digest chemistry can also be combined with terminal modification chemistry by definition of a new name and the proper chemical formula change of the respective terminus (e.g., Trypsin SBAA). The digest rules are coded in crypted text that can be edited by the user within this editor.

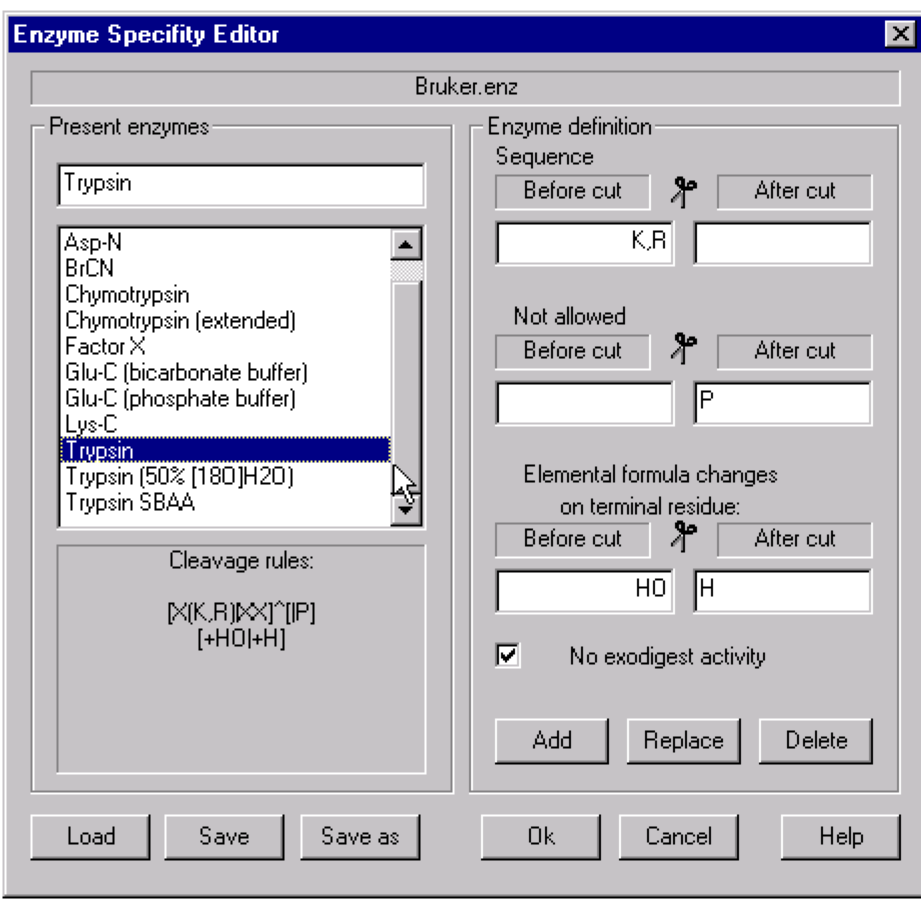


Figure P-11, Enzyme Specificity Editor

Enzyme file Shows the currently selected enzyme file, default file is *BrukerDefault.enz*. The default enzyme file can not be changed in *EnzymeTypesEditor*.

Save this file under a new name to be able to modify it. All changes to the enzyme list will be stored in the current *.enz file if you press ***Save***.

Enzyme The name of the currently selected enzyme or of a new enzyme that should be appended to the enzyme list or that replace the currently selected enzyme.

Modify the current list of enzymes:

Enzyme list The list of all enzymes available in the currently selected enzyme file.

Digest rules:

The encrypted text describing the activity of the selected enzyme appears in this field.

Recognition sequence An amino acid residue or a more complex motif before or after the digest site (symbolized by the scissors). The following definition means that the recognition of the residues K or R or of the sequence DEG will result in a digest: K,R,DEG.

Disabling conditions The digest is not allowed if the specified residue(s) exist(s) direct to the left or to the right of the cutting site.

Formula changes Elemental composition changes of the peptides at both sides of the digest site due to enzyme action. The definition *"-SCH3+O"* for CN-Br means a loss of a SCH3 group at the new C-terminus and a O uptake at the new N-terminus of the proteolytic peptides.

No exodigest activity Select it for enzymes like trypsin, which can´t usually digest a terminal peptide bond.

Add the specified enzyme to the list or **Replace** the specificity of the selected enzyme with the current data.

Delete the selected enzyme.

Modify or use the stored *.enz file.

Load an enzyme file for using or modifying it.

Save all changes to the currently selected enzyme file.

Save as Select this button to save the currently available enzymes under another name.

C Fine Structure Characterization

C.1 Load Data C-2
C.2 Exchange to SequenceEditor C-3
C.3 Change the Modifications C-4
C.4 Modified Amino Acids C-5
C.5 Transfer to BioTools C-6
C.6 Results in BioTools C-7

C.1 Load Data

Use the "Find" option from the "File" menu and load entry "…\DemoData\FLEX\PSD\PhosphoPeptide.FAST\fast\pdata\1\1r.

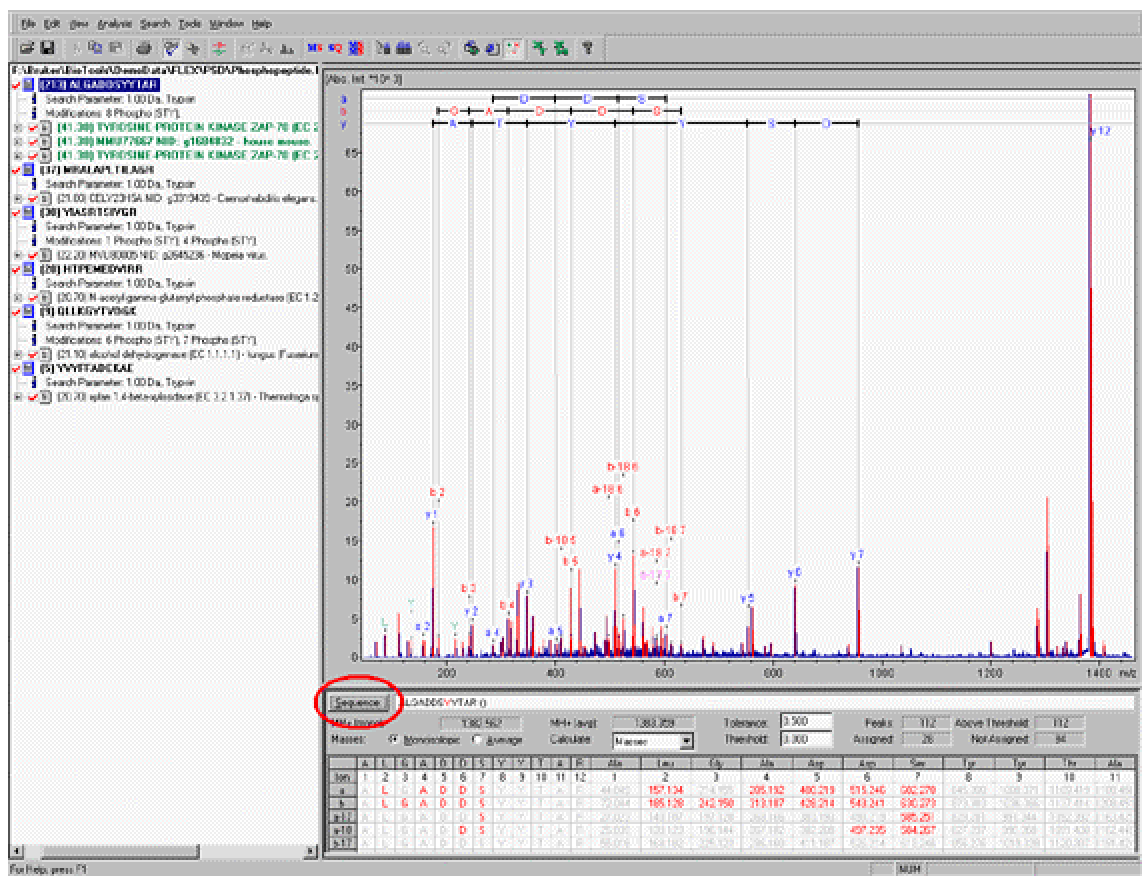


Figure C-1, Phosphopeptide analysis

C.2 Exchange to SequenceEditor

Next click the button "Sequence" to start the SequenceEditor. The SequenceEditor will come up with the sequence already shown in BioTools.

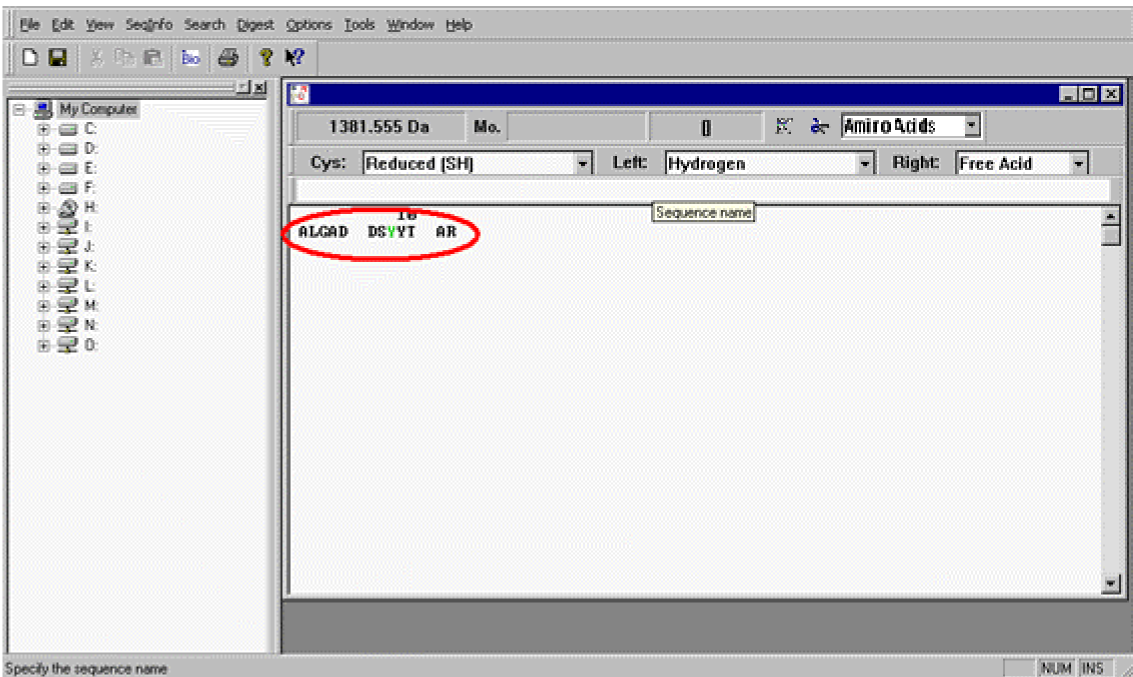


Figure C-2, SequenceEditor with one modified amino acid

C.3 Change the Modifications

A double click within the sequence string will open a dialogue that allows you to change the modifications. Select the marked entries as shown below and replace the existing modification with the new one.

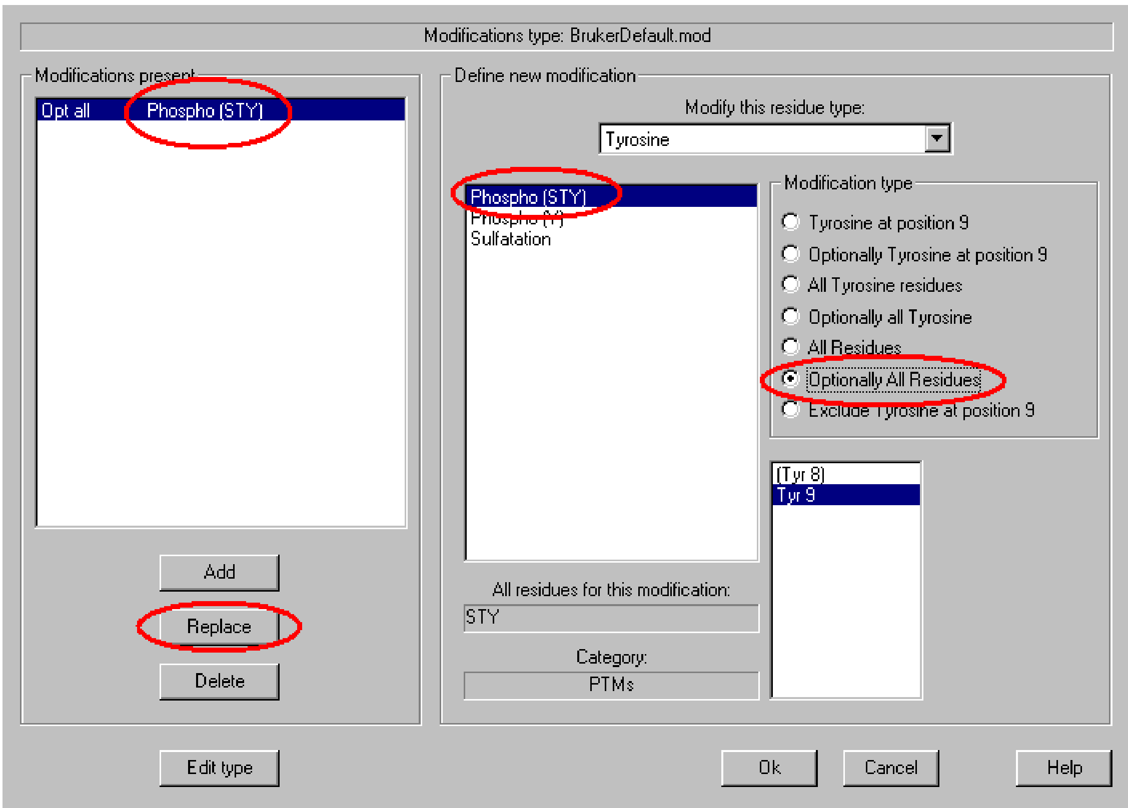


Figure C-3, Change modification

C.4 Modified Amino Acids

Now the SequenceEditor shows four modified amino acids.

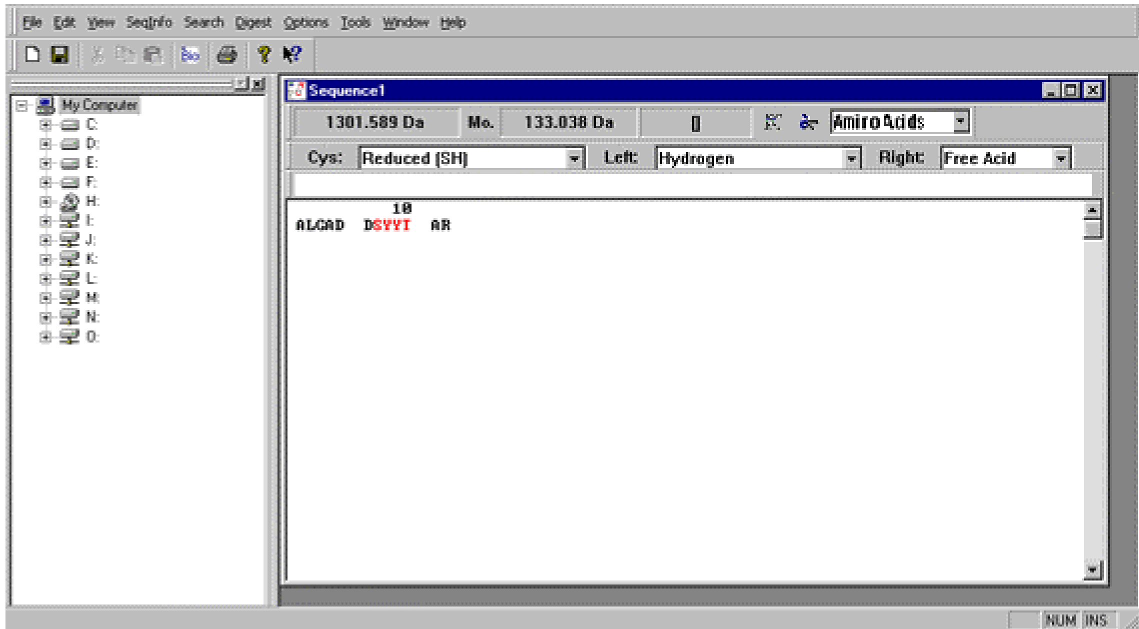


Figure C-4, SequenceEditor with four modified amino acids

C.5 Transfer to BioTools

Next click the "BioTools" button to transfer the changed modification information back to BioTools: set parameter "min" and "max" and use the "Send" button to transfer the data to BioTools (for details about the "Optional Modification Transfer Dialog" see the SequenceEditor manual).

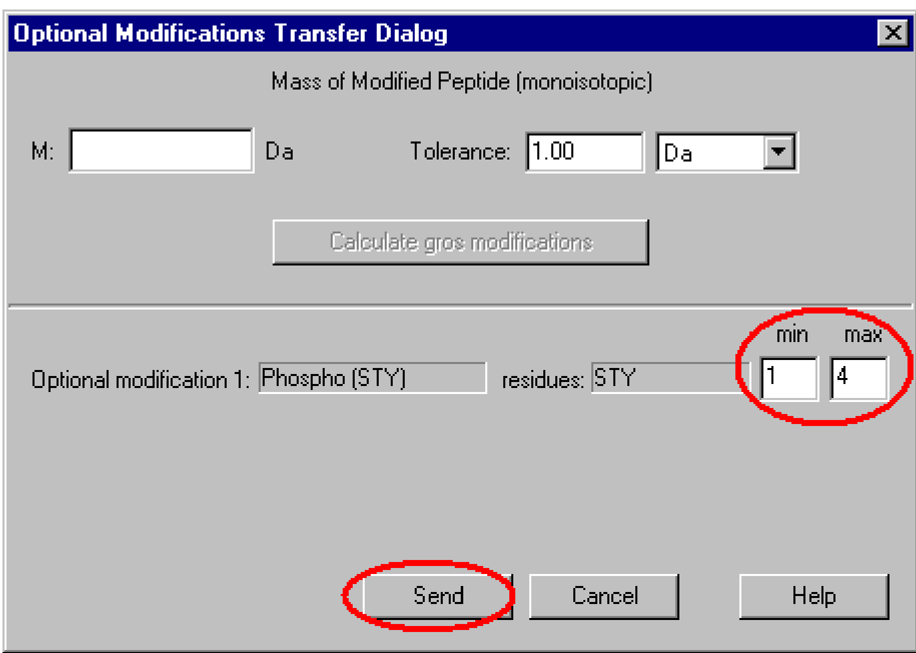


Figure C-5, Optional Modification Transfer Dialog

C.6 Results in BioTools

BioTools shows all received sequences in the treeview and matches them with the spectrum. The best match shows one modification at amino acid number 8: this is the same position you will receive if you perform the Mascot search. The BioTools score (218) for position 8 is two times higher than the score for modification at position 9 (score 100). Use this feature to find the correct position of a known modification.

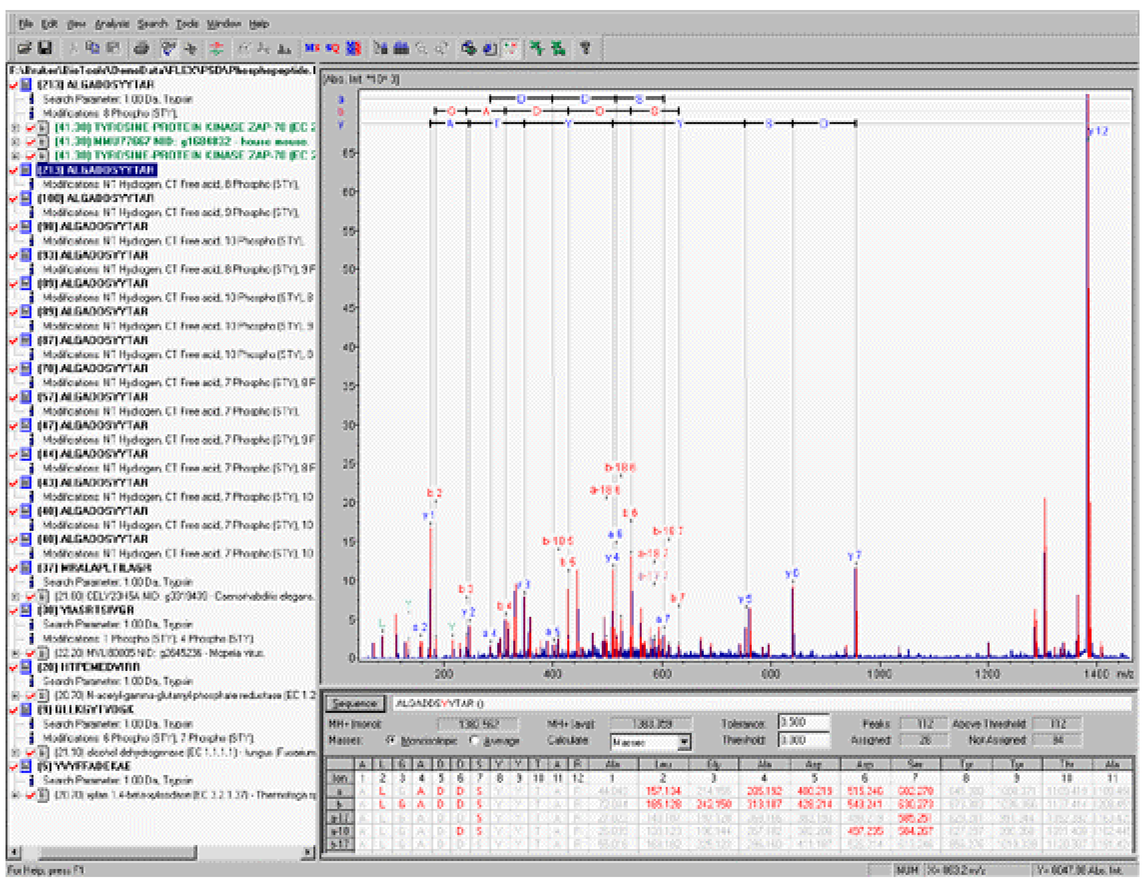
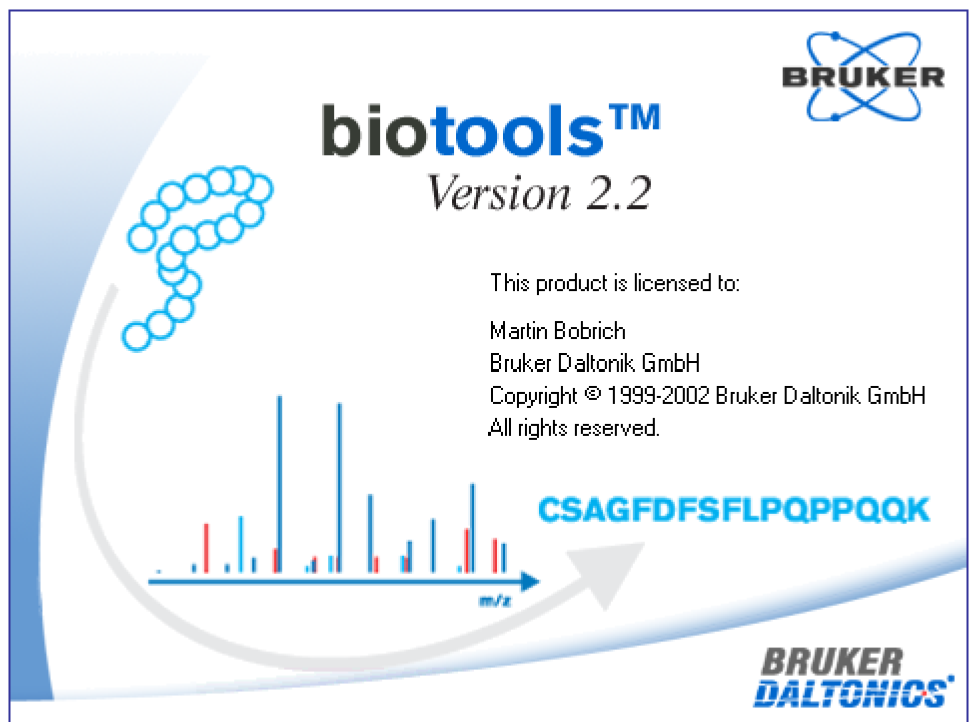


Figure C-6, Modifications

BioTools Reference Manual



(August 2002)

Copyright

Copyright 2002

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

BioTools User Manual, Version 2.2 (August 2002)
Part #: 216936
First edition: October 2000
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.

Copyright: **Bruker Daltonik GmbH**
Fahrenheitstrasse 4
28359 Bremen
Germany

Phone: +49 (4 21) 22 05-200
FAX: +49 (4 21) 22 05-103
Email: mailto:sales@bdal.de
Internet: http://www.bdal.de/

Contents

1 INSTALLATION 1-1

1.1 System Requirements 1-1
1.2 Program Setup 1-1
1.3 Starting the Installation 1-1
1.4 Starting the Program 1-8
1.5 License Manager 1-8

2 QUICKSTART 2-1

2.1 Loading processed data 2-2
2.2 Loading Spectra 2-7
2.2.1. Data from Bruker XMASS / XTOF 2-7
2.2.2. Data from Bruker DataAnalysis 2-8
2.2.3. Display of Picked Peaks 2-9
2.3 Processing of MS Data 2-10
2.3.1. DeNovo Sequencing 2-10
2.3.2. Mascot Database Query (MS/MS) 2-14

3 MENU BAR, TOOLBAR, STATUS BAR AND CONTEXT MENUS 3-1

3.1 File Menu 3-4
3.1.1. Find 3-5
3.1.2. Open Spectrum 3-7
3.1.3. Multiple open 1r 3-7
3.1.4. Close 3-8
3.1.5. Combine multiple LIFT spectra 3-8
3.1.6. Save 3-9
3.1.7. Print 3-9
3.1.8. Print Preview 3-10
3.1.9. Print Setup 3-11
3.1.10. Send 3-12
3.1.11. Last Used Data 3-12
3.1.12. Exit 3-12
3.2 Edit Menu 3-13
3.2.1. Undo 3-13
3.2.2. Cut 3-13
3.2.3. Copy 3-13
3.2.4. Paste 3-13
3.2.5. Sequence 3-13
3.3 View Menu 3-14
3.3.1. Toolbar 3-15

3.3.2. Status Bar 3-16
3.3.3. Query Results 3-17
3.3.4. View Fingerprint Results 3-18
3.3.5. View MS/MS Results 3-18
3.3.6. Matched Peaks 3-18
3.3.7. Unmatched Peaks 3-18
3.3.8. Matched and Unmatched Peaks 3-18
3.3.9. Picked Peaks 3-18
3.3.10. Coordinates 3-18
3.3.11. Grid 3-18
3.3.12. Scaling 3-19
3.3.13. Zooming 3-20
3.3.14. Undo Zooming 3-20
3.3.15. Redo Zooming 3-20
3.3.16. Distance Cursor 3-20
3.3.17. Data Cursor 3-22
3.3.18. Color 3-23
3.3.19. Display Mode 3-24
3.4 Analysis Menu 3-25
3.4.1. Check Sequence 3-25
3.4.2. DeNovo Sequencing 3-27
3.4.3. Full DeNovo Sequencing 3-34
3.4.4. ISD Data 3-42
3.4.5. Set Threshold 3-43
3.4.6. Annotation Parameter 3-44
3.4.7. Add Peaks 3-46
3.4.8. Remove Peaks 3-46
3.4.9. Peak Picking 3-46
3.5 Search Menu 3-49
3.5.1. EMBL 3-52
3.5.2. PROWL Peptide Mapping 3-53
3.5.3. Mascot 3-54
3.5.4. Search for Masses (SequenceEditor) 3-75
3.5.5. Digest (SequenceEditor) 3-76
3.6 Tools Menu 3-78
3.6.1. Authentication 3-78
3.6.2. Options 3-79
3.6.3. Maldi Spectrum Parameter 3-82
3.6.4. Modification Info 3-83
3.6.5. Formula Parser 3-84
3.6.6. Start XTOF NT 3-84
3.6.7. Start XMASS NT 3-84
3.6.8. Execute XMASS/XTOF-Commands 3-85
3.6.9. Customize 3-86

3.7 Window Menu 3-90
3.7.1. Show/Hide Treeview 3-90
3.7.2. Show/Hide Fragments 3-90
3.7.3. Show/Hide Browser Window 3-91
3.7.4. Start SequenceEditor 3-91
3.7.5. Reset Window Sizes 3-91
3.7.6. Cascade 3-91
3.7.7. Tile Horizontally 3-92
3.7.8. Tile Vertically 3-93
3.7.9. Arrange Icons 3-94
3.7.10. Active Data 3-95
3.8 Help Menu 3-96
3.8.1. License Manager 3-97
3.8.2. What's new in BioTools 2.1 3-97
3.8.3. What's new in BioTools 2.2 3-97
3.8.4. Open BioTools Manual 3-98
3.8.5. Help Topics 3-99
3.8.6. About BioTools 3-99
3.9 Menu Bar 3-100

A APPENDIX: AMINO ACID RESIDUES AND FRAGMENTATIONS .... A-I

A.1 Amino Acid Residues .... A-I
A.1.1. Single letter code .... A-I
A.1.2. Genetic Code .... A-III
A.1.3. Formulas and Molecular Weights .... A-IV
A.1.4. Chemical Structure .... A-V
A.2 Peptide Fragmentation .... A-VI
A.3 Menu and Shortcut list .... A-XI
A.4 Toolbar Reference list for BioTools .... A-XIV
A.5 Toolbar Reference list for SequenceEditor .... A-XVI
A.6 Part Numbers .... A-XVI

I INDEX .... I-I

Table of changes

Version	***Date***	***Changes***
2.0	2000-10-14	Software versions: BioTools, Version 2.0 SequenceEditor, Version 1.0
2.1	2002-01-18	Software versions: BioTools, Version 2.1 SequenceEditor, Version 2.1
2.2	2002-08-01	Software versions: BioTools, Version 2.2 SequenceEditor, Version 2.1

1 Installation

This chapter contains information on the hardware and software requirements for running BioTools. It also describes the installation procedure and how to start the program from WINDOWS Program Manager.

1.1 System Requirements

- CPU: Intel Pentium II processor or better.
- Clock: 266 MHz processor for satisfying data handling.
- Main Memory: Minimum 64 Mbytes RAM or better.
- Operating System: Microsoft Windows-NT version 4.0 or higher (Service Pack 3 or better), installed and operating Internet Explorer 4.0 or better.
- Graphic Resolution: 1024 * 768 pixel, 256 colors or better
- CD–ROM drive (4X or better)
- Ethernet connection
- Hard disk: at least 100 Mbytes of free disk space.

1.2 Program Setup

First make sure that Windows NT Version 4.0 with Service Pack 3 and the Microsoft Internet Explorer 4.0 or better is installed on your computer system. We do not recommend proceeding beyond this point when your system does not run this software.

1.3 Starting the Installation

Put the installation CD-ROM into the drive (e. g. D:). If the "Autostart" function is activated, the installation program will start automatically. Otherwise click to the WINDOWS "Run..." element.

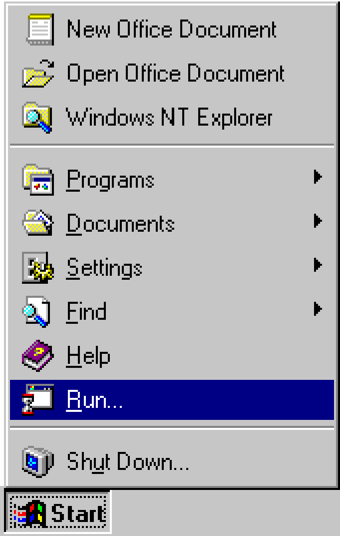


Figure 1-1, The WINDOWS „Run..." element

The dialog box shown in the next figure appears. The installation CD must be inserted into the appropriate drive (e.g. "D") and the command line "D:\SETUP.EXE" typed in. The installation system then starts as soon as the *Ok* button of the dialog box is activated.

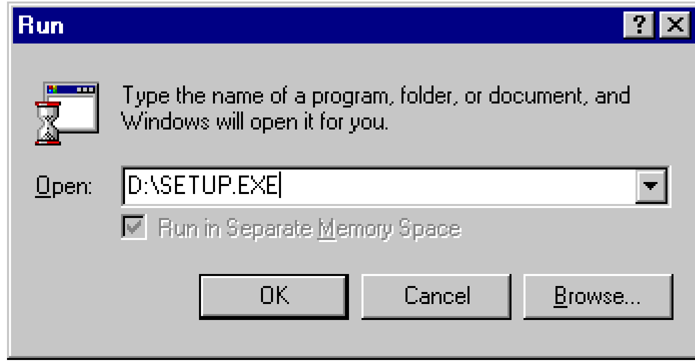


Figure 1-2, Start of the Setup program

The installation starts with a "Welcome" message followed by the software license agreement. Next follow the instructions on the screen: enter your user information, destination directory and select the components to be installed on your computer.

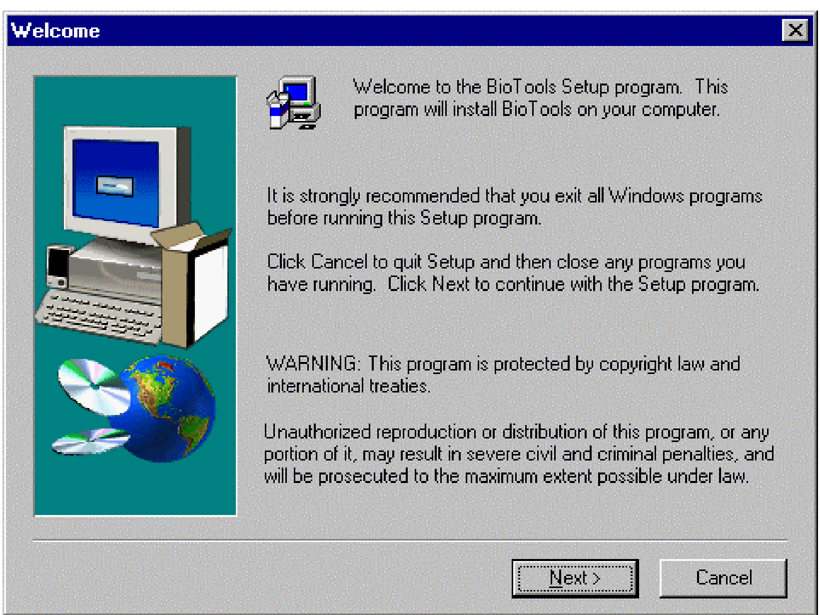


Figure 1-3, Welcome to the BioTools Setup Program

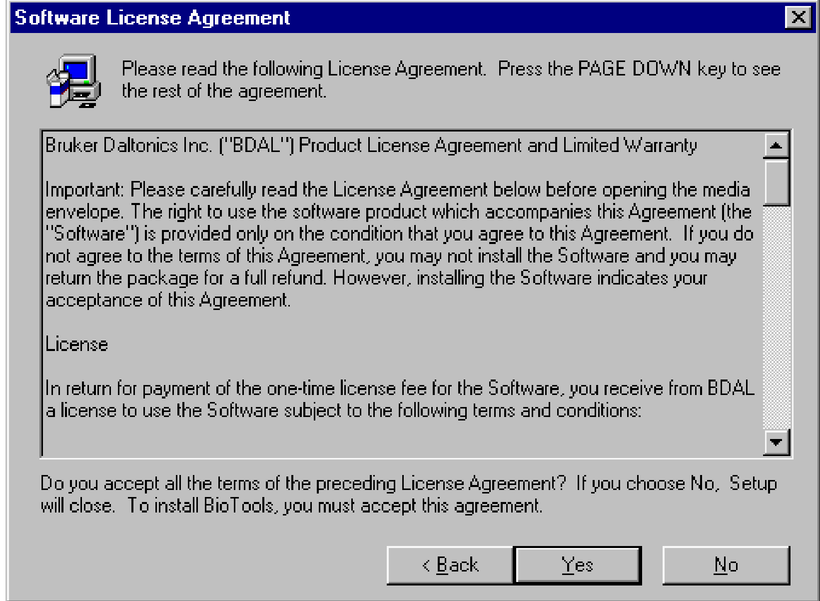


Figure 1-4, Software License Agreement

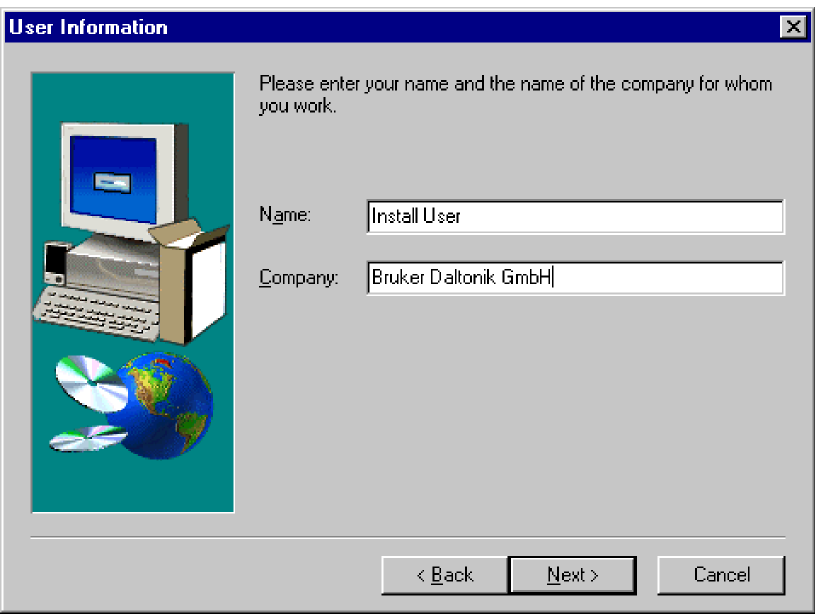


Figure 1-5, User Information

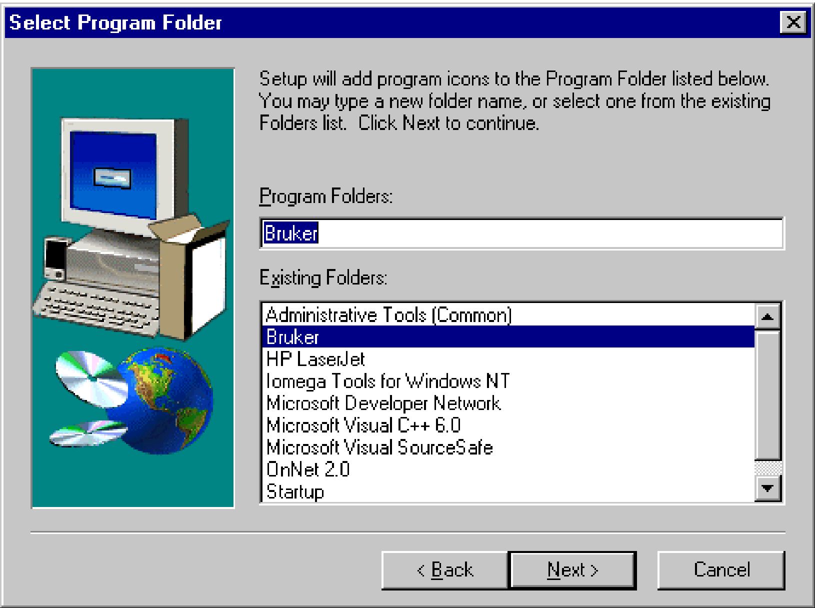


Figure 1-6, Select Program Folder

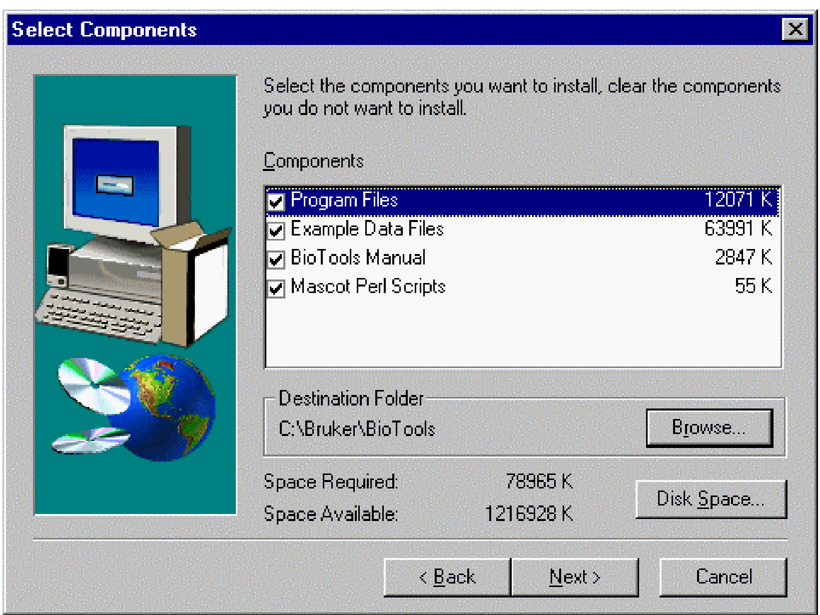


Figure 1-7, Select Components

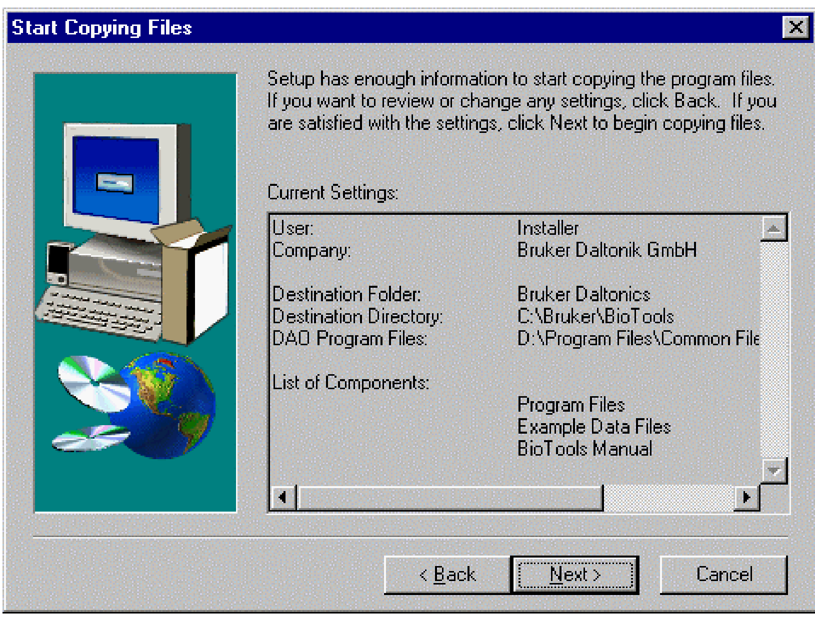


Figure 1-8, Start Copying Files

The entry "Mascot Perl Scripts" is only available if the Mascot Database is installed on your computer. Normally you will use BioTools and Mascot on different machines.

In this case you must perform the BioTools setup on your Mascot server also, but you have to select the component "Mascot Perl Scripts" only.

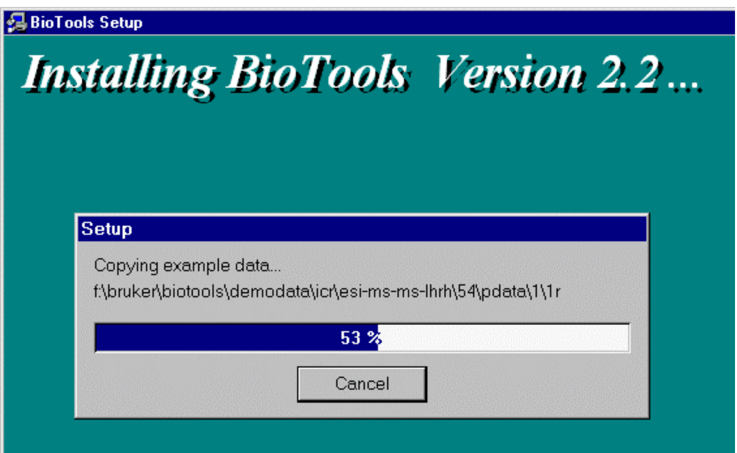


Figure 1-9, Installing BioTools

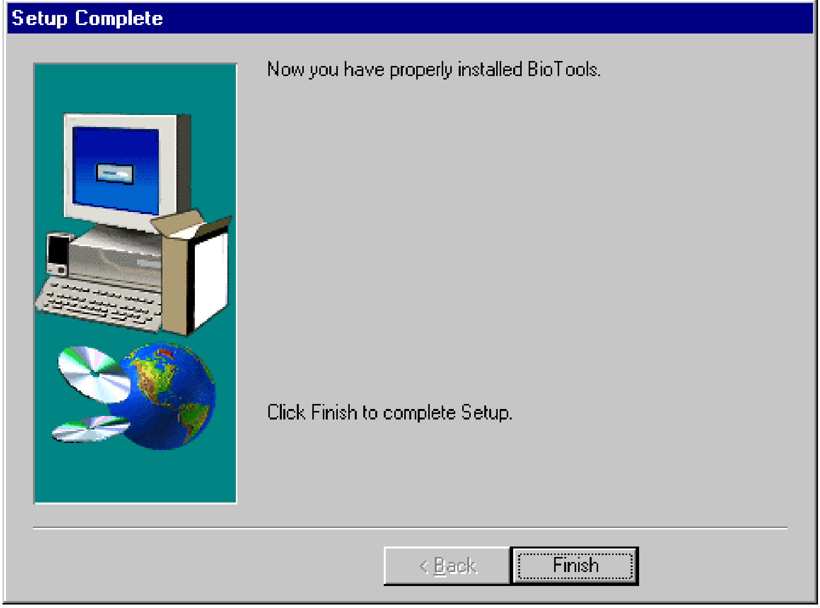


Figure 1-10, Setup Complete

On the other hand you can copy the required Perl scripts by hand from the distribution CD of BioTools (see folder "BioTools Perl Scripts for Mascot") into the Mascot directory (normally "inetpub\Mascot\cgi").

Important Note for use of remote Mascots servers:

Normally you will use BioTools and Mascot on different machines. In this case you must prepare the mascot server after finishing the BioTools installation:

- Either you can perform the BioTools setup on your Mascot server, but you have to select the component "Mascot Perl Scripts" only.
- Or you can copy the Perl scripts manually from the distribution CD of BioTools (see folder "BioTools Perl Scripts for Mascot") into the Mascot directory on the server computer (normally "inetpub\Mascot\cgi").

If you decided to install the BioTools manual and the Acrobat Reader software is not installed yet, you will find an icon "Install Acrobat Reader" in your BioTools program group. Use this icon to install the Acrobat Reader software.

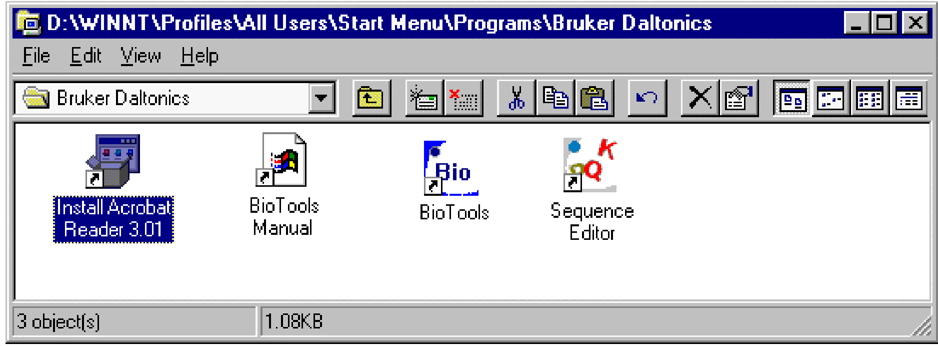


Figure 1-11, Acrobat Reader Program not yet installed

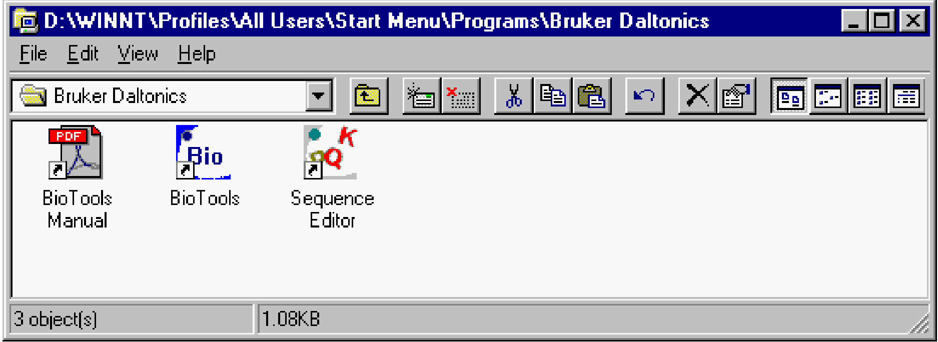


Figure 1-12, After installation of the Acrobat Reader Program

1.4 Starting the Program

Use the BioTools Icon from the program group you specified during program installation. If you receive the error message "A procedure entry point httpsendrequestExA could not be located in the dynamic linked library" during program start, the Microsoft Internet Explorer 4.0 or better is not installed on your system. The installation of the Microsoft Internet Explorer 4.0 or better will replace an existing WININET.DLL by a newer one.

1.5 License Manager

If you start BioTools the first time, you must enter a license key. The key comes together with the BioTools documentation. Use the license manager to add or remove licenses (Menu Help – License Manager).
If an invalid license key is entered, the program can not be used.

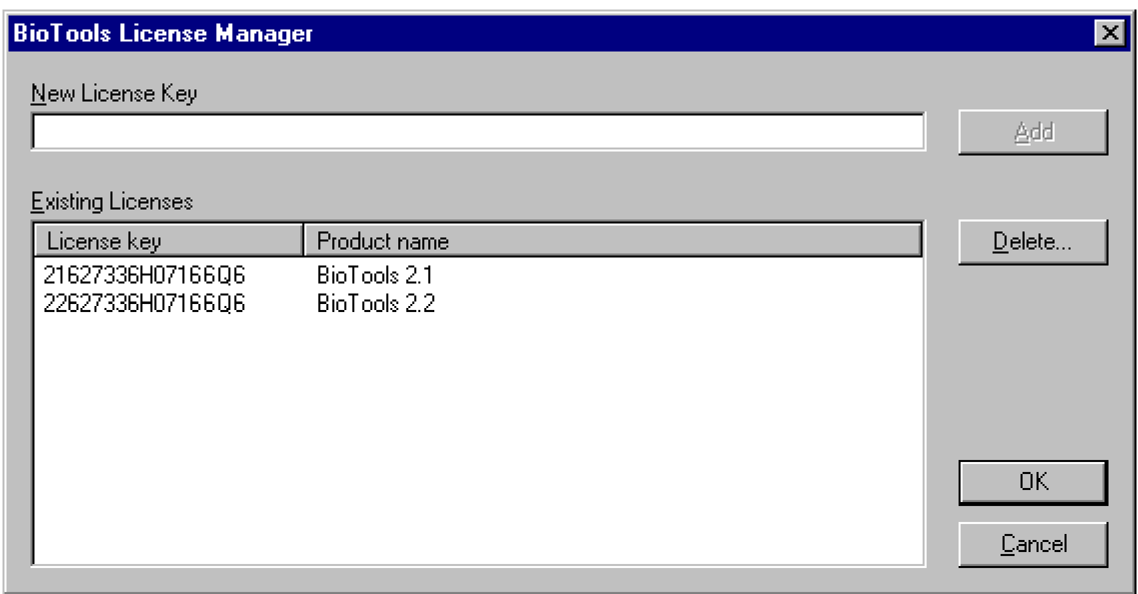


Figure 1-13, License Manager

2 Quickstart

BioTools uses processed data from different Bruker software packages. These data are spectra and peaklists, which must have been created in advance. BioTools will use the data to perform a *DeNovo* sequencing or an Internet search based on the peaklist. As a result you will receive one or more amino acid sequence which will be matched with the experimental spectrum. The best match can be used to annotate the spectrum.

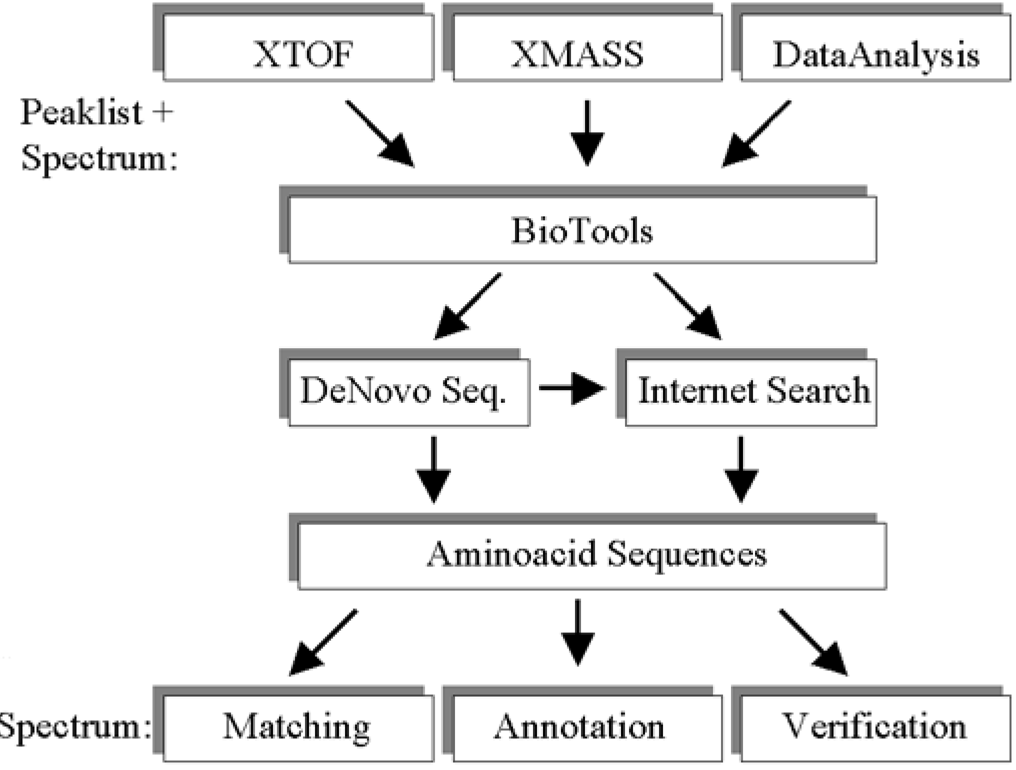


Figure 2-1, BioTools data flow

2.1 Loading processed data

To open an already processed data set, use the option "Find" from the "File" menu . The default file name is "1R" (this is processed data from XMASS/XTOF). The path after "Look in" is the folder where BioTools was saved during program installation. Subfolders will be included for the search also because the button "Include subfolders" is checked. Next click to button "Find Now" to start the search.

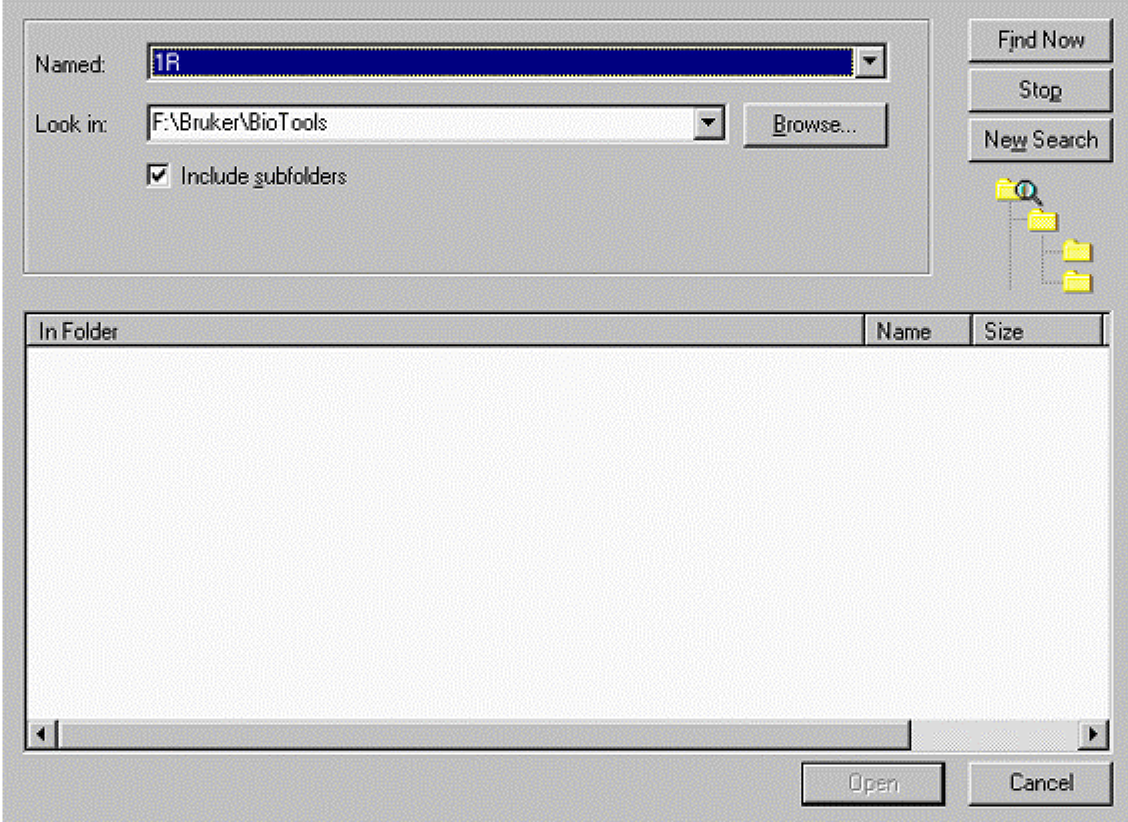


Figure 2-2, Find files

From the list of found files select the entry "ACTH1-17.FAST" and load it using button "Open". This is a MALDI-PSD spectrum and therefore an example for the analysis of MS/MS data.

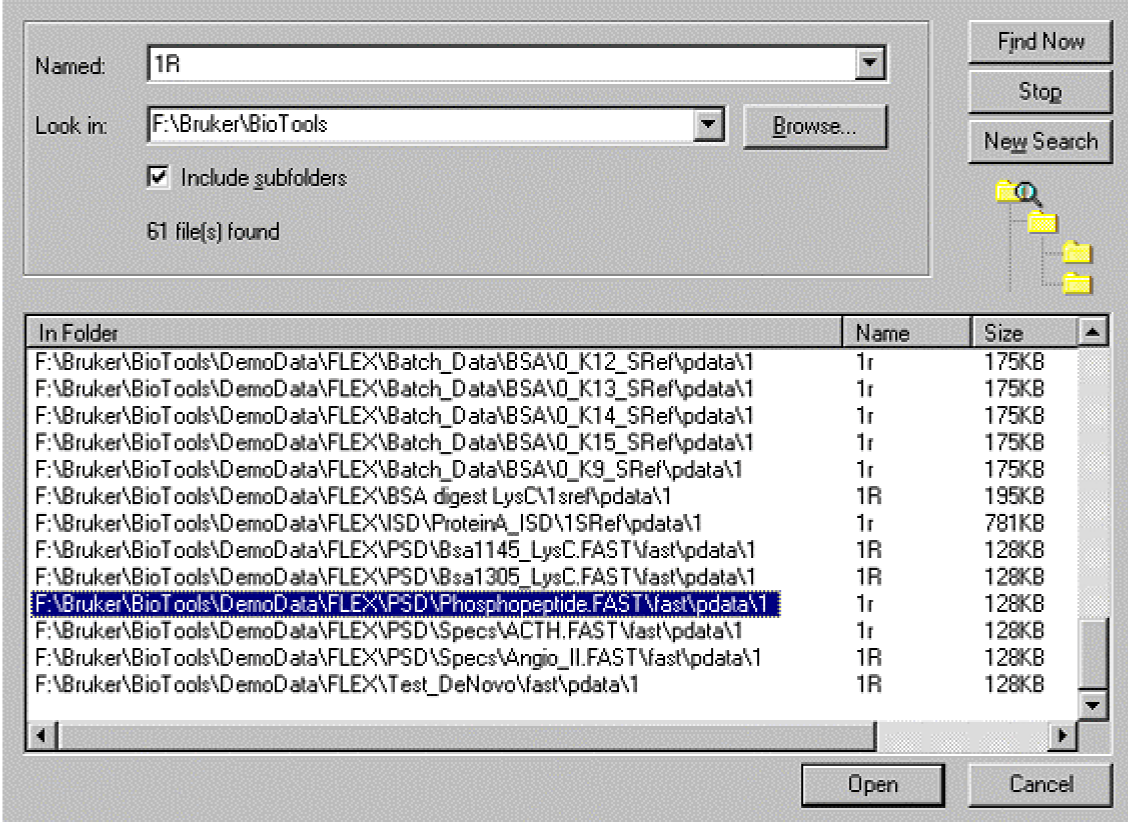


Figure 2-3, Found files

The result is shown in the next figure (in this example the background color was set to white; the default background color after program installation is gray).

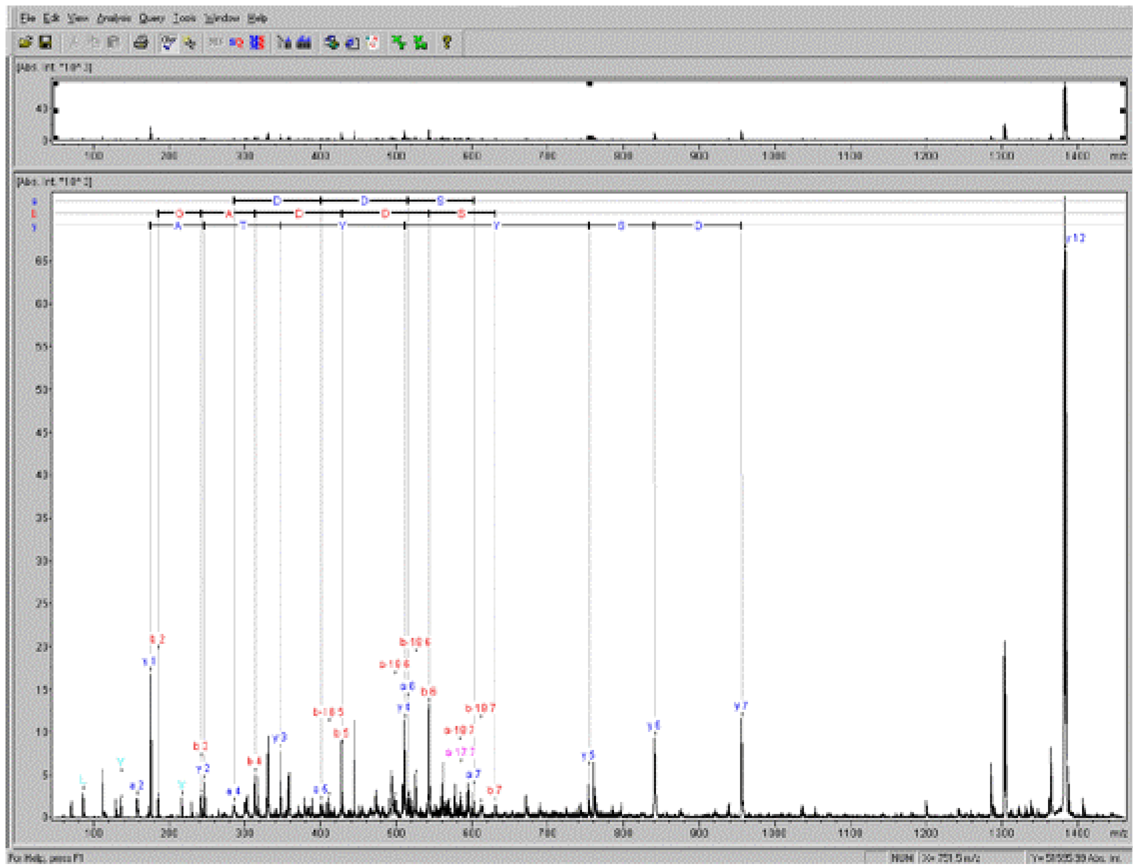


Figure 2-4, Reloaded processed data

The upper slice shows the overview window: this is always the complete spectrum. The lower slice shows the selected range of the spectrum. Use the mouse to zoom in the displayed spectrum: move the mouse cursor into the region of the spectrum window, press the left mouse button, move the mouse and finally release the mouse button. The lower slice now shows a part of the spectrum, the upper slice shows the complete spectrum and a rectangle around the zoomed area.

There are different ways to move within the spectrum:

- Move the mouse pointer into the overview window into the marked rectangle, press the mouse button, move the rectangle with your mouse to an area of interest and release the mouse button. The zoomed area will be updated immediately.
- Move the mouse pointer into the area of the x-or y-axis, press the left mouse button and move the mouse. Spectrum window and overview window will be synchronized.

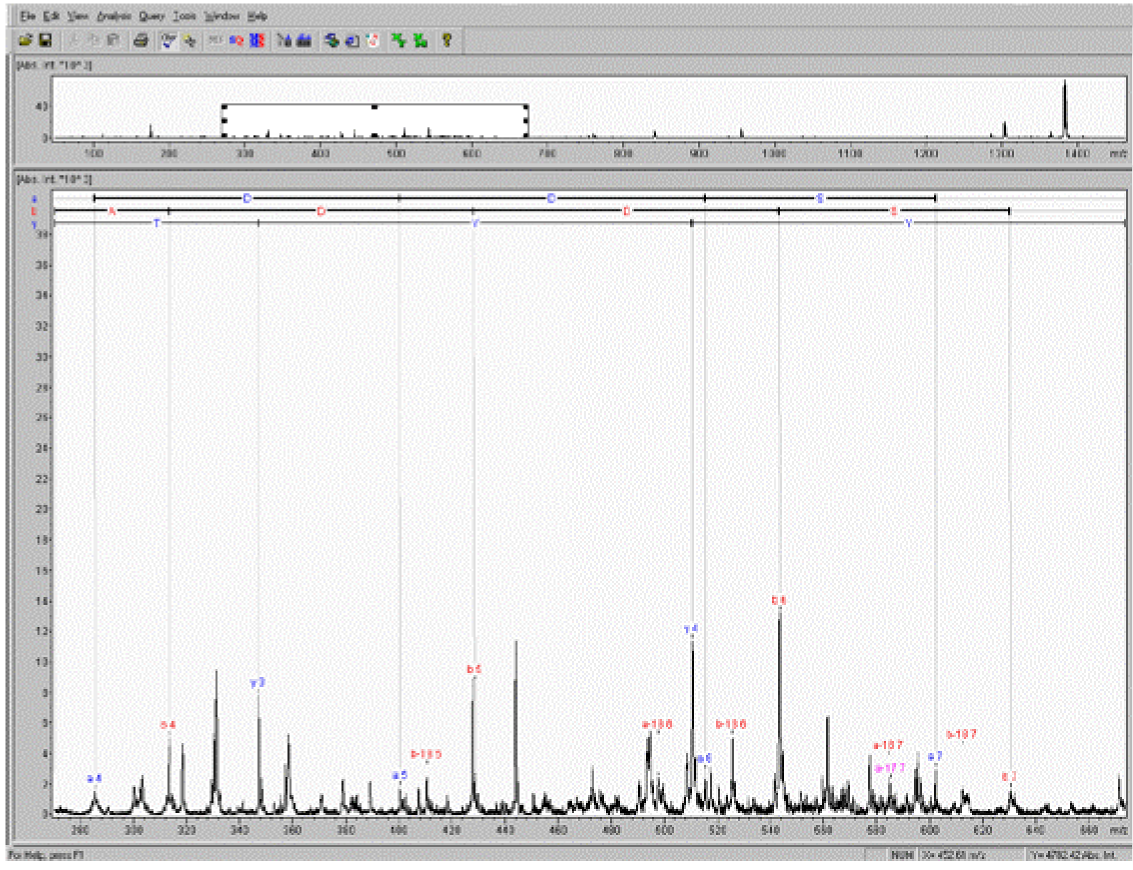


Figure 2-5, Zoomed spectrum

A double click with the left mouse button within the spectrum window will reset the display (a more detailed description of all features of this window will be given later in this manual).

To change the ratio between overview window and spectrum window move the mouse cursor to the small area between both windows. The shape of the cursor will change to an up/down arrow if you have reached the right position. Now press the left mouse button and move the mouse upward to hide the overview window.

Probably you may wonder where the information of annotation comes from. From the toolbar select the "Show/Hide Fragments" button .

The display will be split into two parts: the upper still shows the annotated spectrum, the lower gives information about the sequence and calculated fragments. The text field to the right of the "Sequence" button, displays the amino acid sequence (optionally followed by an additional comment). This sequence is used to calculate the sum-

formula, mono-isotopic, average parent mass and fragment ions (there will be a detailed description how these fragments are calculated later in this manual). The "Sequence" button allows editing of the sequence and of modifications using the ***SequenceEditor***, which is described in the respective manual. Calculated fragment ions are matched to picked peaks in the spectrum . In case of a **match** the single letter symbol of the amino acid and the corresponding mass is **displayed in red color**, **otherwise in gray color**. Matching takes place with a given tolerance of permissible differences between calculated and found peak masses. A threshold may be set to ignore small peaks from background noise.

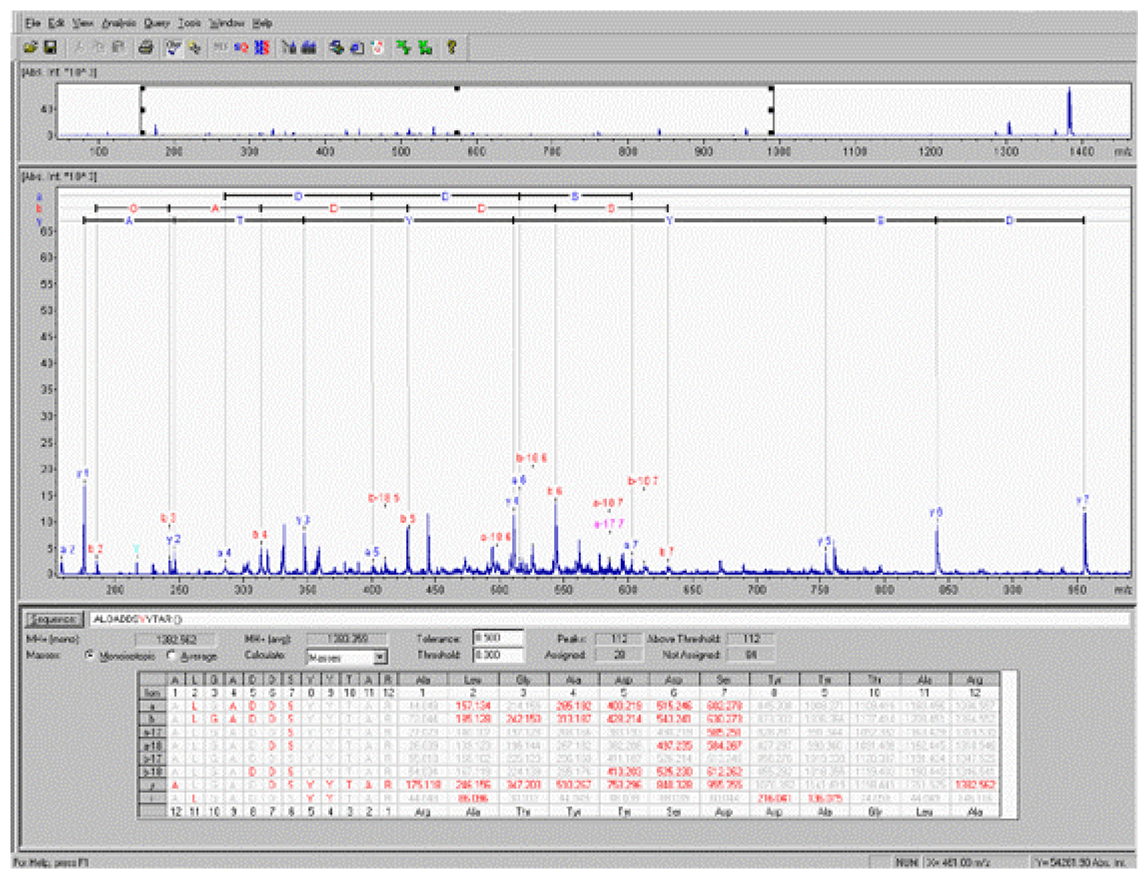


Figure 2-6, Annotated spectrum and calculated fragments

2.2 Loading Spectra

BioTools uses data that has already been processed either by Bruker XMASS/XTOF (Version 5.1 or higher) or Bruker DataAnalysis (Version 2.0 or higher).

2.2.1. Data from Bruker XMASS / XTOF

2.2.1.1. XTOF Windows NT

Using the Windows NT version of XTOF is the easiest way to receive data for processing with BioTools. Start XTOF, open a spectrum, perform peak picking using "label region" and then click the button "Bio" to start the program BioTools. BioTools will automatically load the corresponding spectrum together with the picked peaks.

2.2.1.2. XTOF Unix

Open a spectrum and perform peak picking using "label region". Use the "Bio" button to transfer the data back to BioTools.

For XTOF Unix 5.1 and higher: Use the "Bio" button to store the current peak list as an XML-file.

For XTOF Unix 5.0 and lower: Use the **P2BT** command from the command line level in order to save the peak list in all formats required for the full functionality available in BioTools. If you do not find the command on your system please write an email to DSU@bdal.de and specify your XTOF version. We will email the command to you.

The peak list is then automatically loaded into BioTools together with the spectrum when it is opened manually.

2.2.2. Data from Bruker DataAnalysis

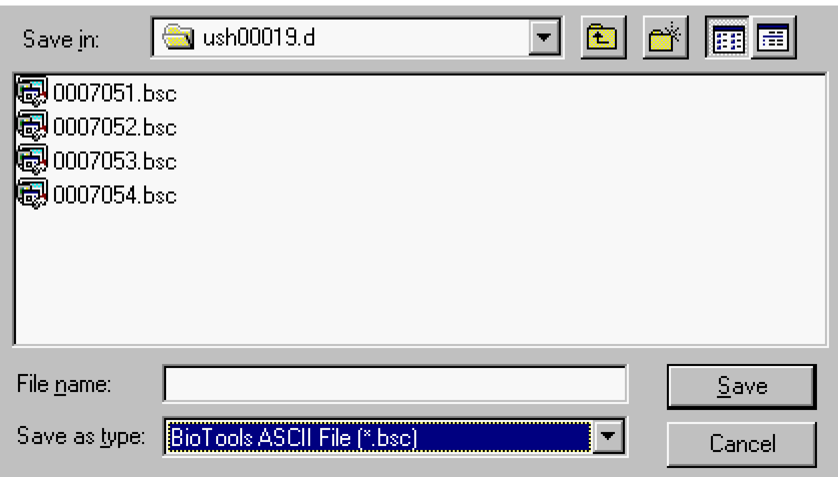


Figure 2-7, Export of data from Bruker DataAnalysis

To export data from DA 2.0 and higher, use "Export – Mass Spectrum" from the "File" menu. The data format must be "BSC". Export the data into the same folder where the other files of the data set are stored.

More detailed information is in the tutorial, Using BioTools for esquireSeries data .

2.2.3. Display of Picked Peaks

To display peaks that have been previously picked by Bruker XTOF or Bruker DataAnalysis 2.0 move the mouse into the spectrum window of BioTools, press the right mouse button and check the entry "Picked Peaks". Picked peaks are displayed as red histogram bars in the spectrum.

To hide the picked peaks uncheck this entry.

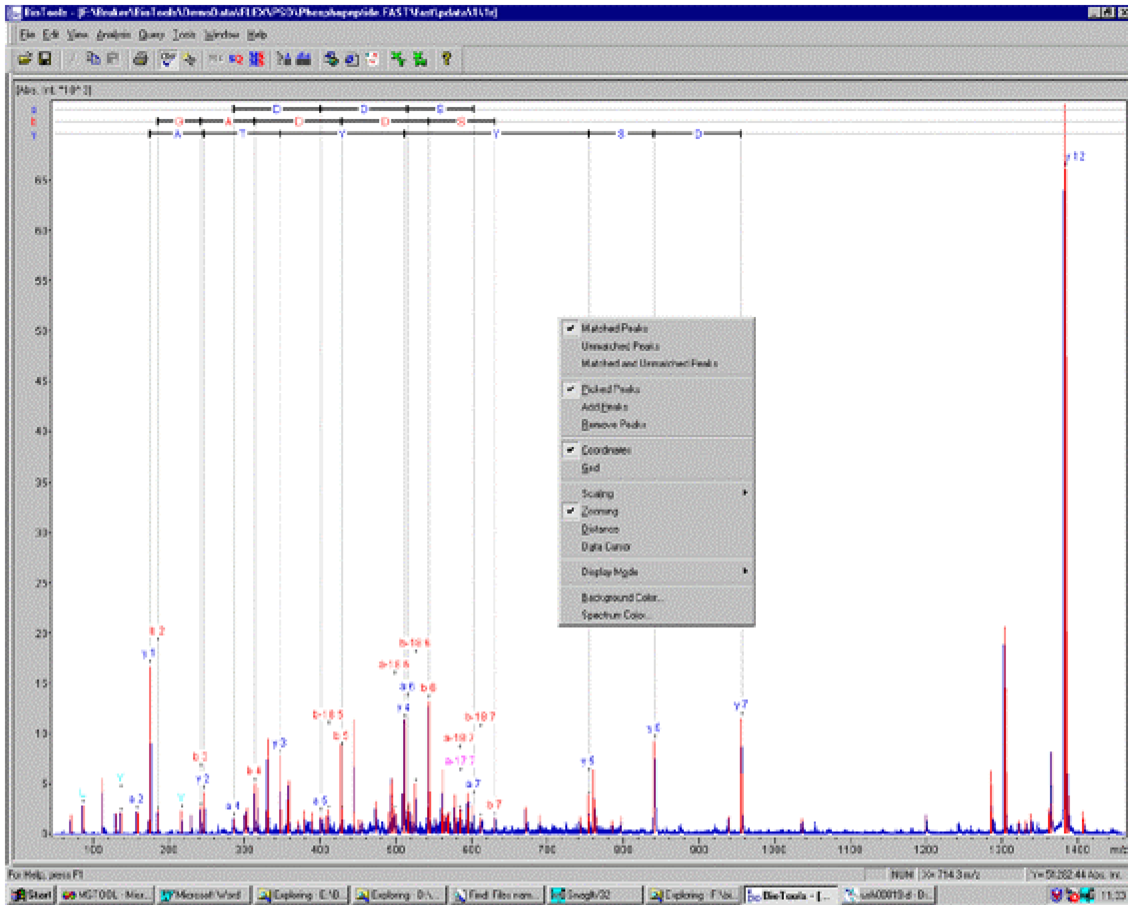


Figure 2-8, Spectrum with picked peaks

2.3 Processing of MS Data

2.3.1. DeNovo Sequencing

To open a data set, use again the option "Find" from the "File" menu. If you did use this option previously the list of files that were found before is displayed again. Otherwise set entry "Named" to "1R". The path after "Look in" should be the folder where BioTools was saved during program installation. Subfolders must be included for the search (check button "Include subfolders"). Next click to button "Find Now" to start the search. From the list of found files select the entry "Test_DeNovo.FAST" and load it using the "Open" button.

Next select button "*DeNovo* Sequencing", button "Show/Hide Treeview" (if the treeview is shown) and button "Show/Hide Fragments" (if the calculated fragments are not shown). The result is shown in the next figure.

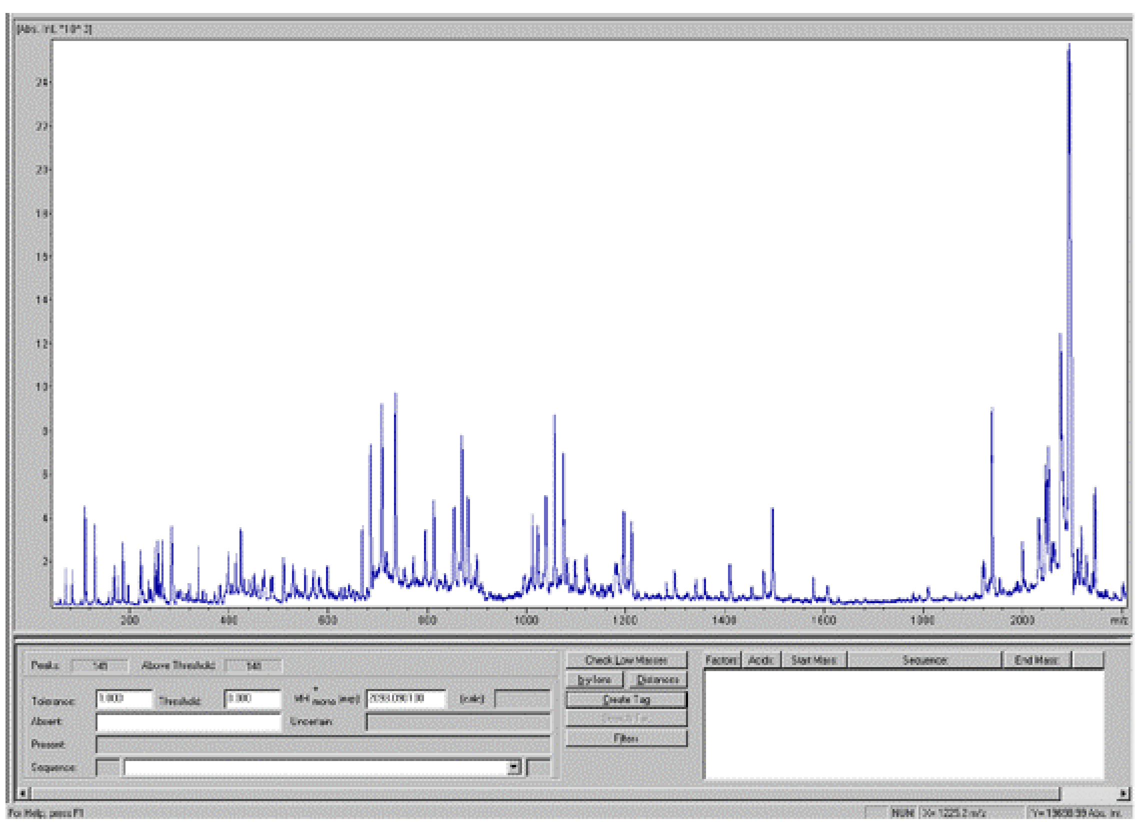


Figure 2-9, Start of *DeNovo* sequencing

In the "Absent" text field type in "**I Q**" if you want to avoid redundant sequence tag suggestions for the isobaric residue pairs I and L, and K and Q. Click the buttons "Create Tag" and then "Search Tag".

Note ***The search with sequence tags in PeptideSearch is only useful when a MASCOT search failed and the assumption is, the reason may a modification in the peptide. Peptide search allows for error-tolerant searches by reduction of the required match regions (standard: 1 and 2 and 3). 1 and 2 allows an error C-terminal of the sequence tag, 2 and 3 N-terminal of the tag and 1 and 3 allows an error within the tag sequence.***

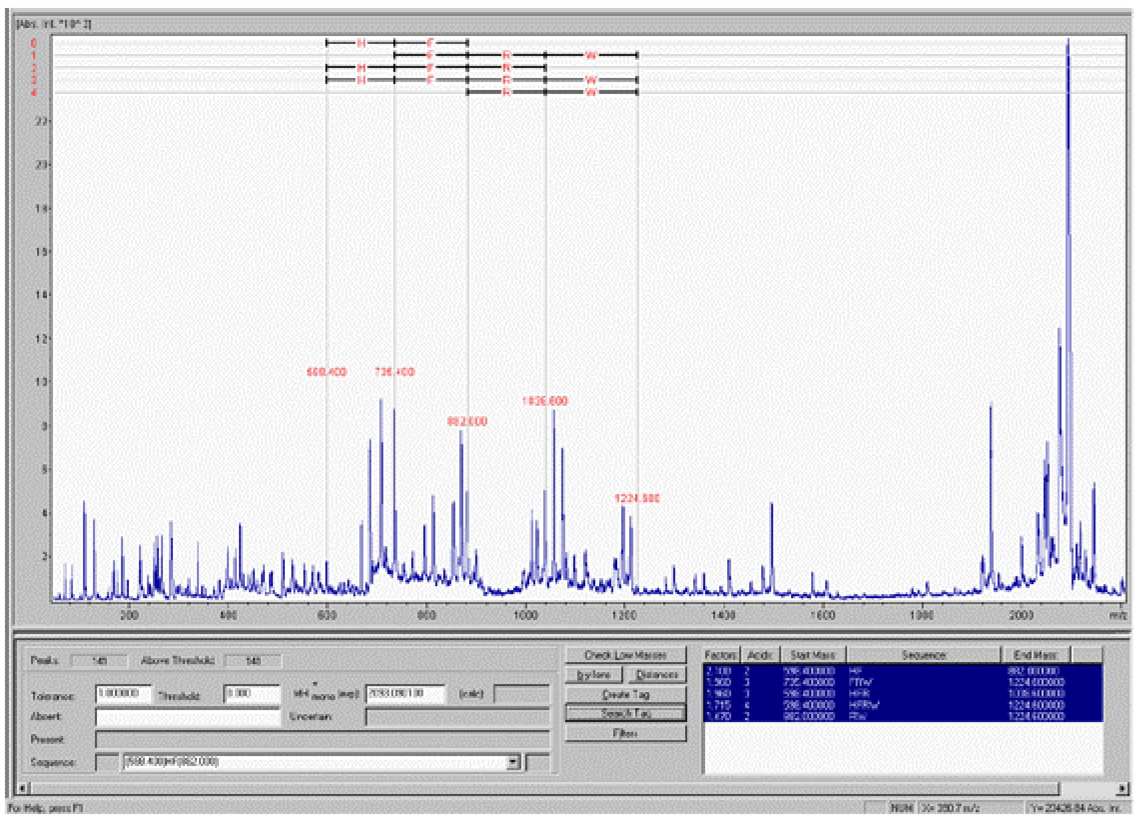


Figure 2-10, Create Tag result

URL: http://193.175.249.95:80/CGI/PPG.PeptSearchSequenceTags.acgi

Protein mass range from (kDa): 0 to (kDa): 300

Cleavage agent: Trypsin

Cysteine is: Cys

Oxidized Methionine

Peptide mass (neutral): 2092.082275 Monoisotopic mass

Mass accuracy: 0.3 Da

Peptide sequence tag: (598.400)HF(882.000)

Match regions: 1 and 2 and 3

Pattern match search by: B-type sequence ions

Edman type search by: N-terminal sequence extension

Allowed number of errors: 0

Cleavage specifities: N-terminal specifity C-terminal specifity

Results per page: 100

Copy Save Start Exit

Figure 2-11, Sequence Tag search parameter

The "Sequence Tag Search" dialog shows search parameters and the "best" found sequence tag resulting from the peak list. A click to the button "Start" will connect you to the Internet (provided your Internet connection is installed properly). After some time the results of the query will be displayed in a window of the Microsoft Internet Explorer. Use the button "Get Results" (on the right side of the result window) to transfer the query results to your PC. The obtained sequences (in this example there are 25 hits, however all found sequences are identical) will be added to the treeview on the left side of the screen. Leave the result window (use the "Exit" button), close the "Sequence Tag Search" dialog and switch back to the "Check Sequence" mode . You will receive the annotated spectrum.

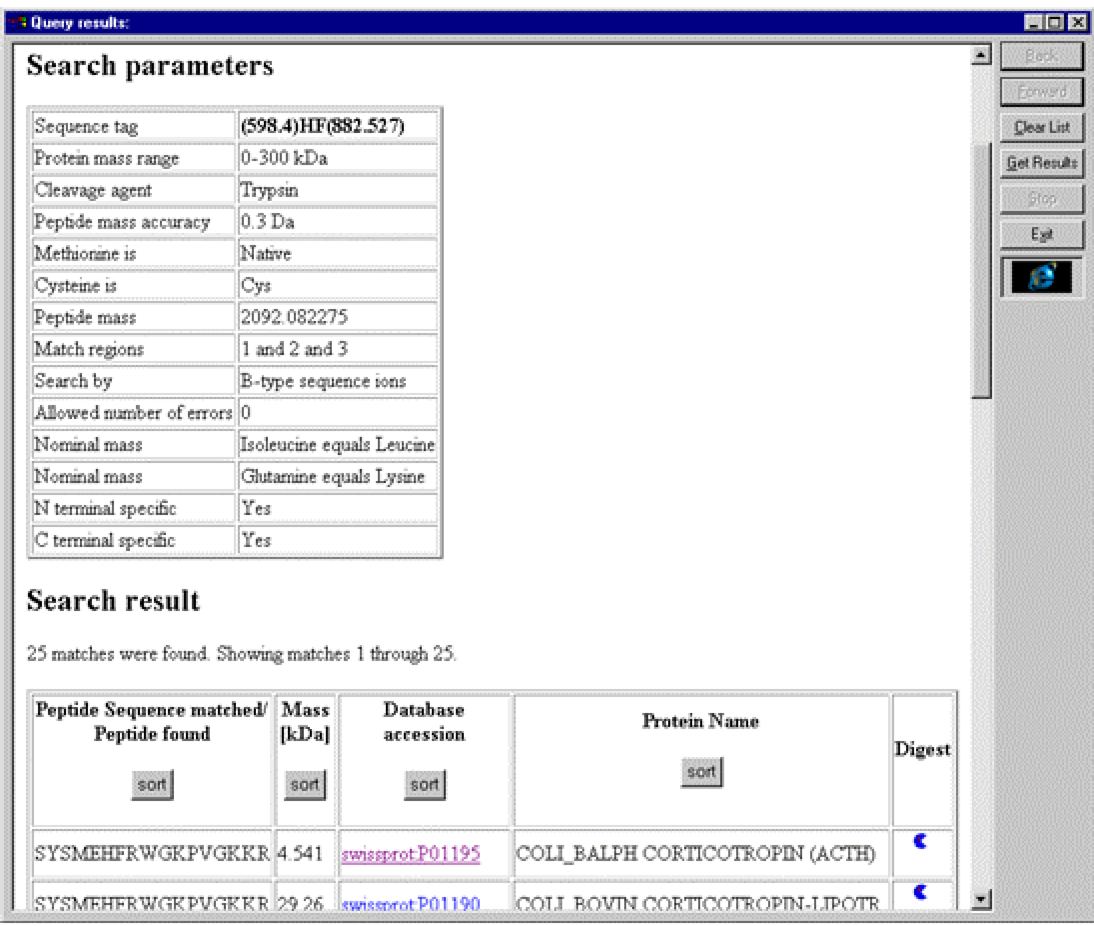


Figure 2-12, Sequence Tag search results

2.3.2. Mascot Database Query (MS/MS)

In all cases, in which the presence of reference sequence information in protein sequence databases can be expected, it is useful to perform a sequence database search. To open a test data set, use the option "Find" from the "File" menu. If you did use this option previously the list of files that were found before is displayed again. Otherwise set entry "Named" to "1R". The path after "Look in" should be the folder where BioTools was saved during program installation. Subfolders must be included for the search (check button "Include subfolders"). Next click to button "Find Now" to start the search. From the list of found files select the entry "Bsa1305_LysC.FAST" and load it using the "Open" button. Next select button "Check Sequence", button "Show/Hide Treeview" (if the treeview is not shown) and button "Show/Hide Fragments" (if the calculated fragments are not shown). The result is shown in the next figure.

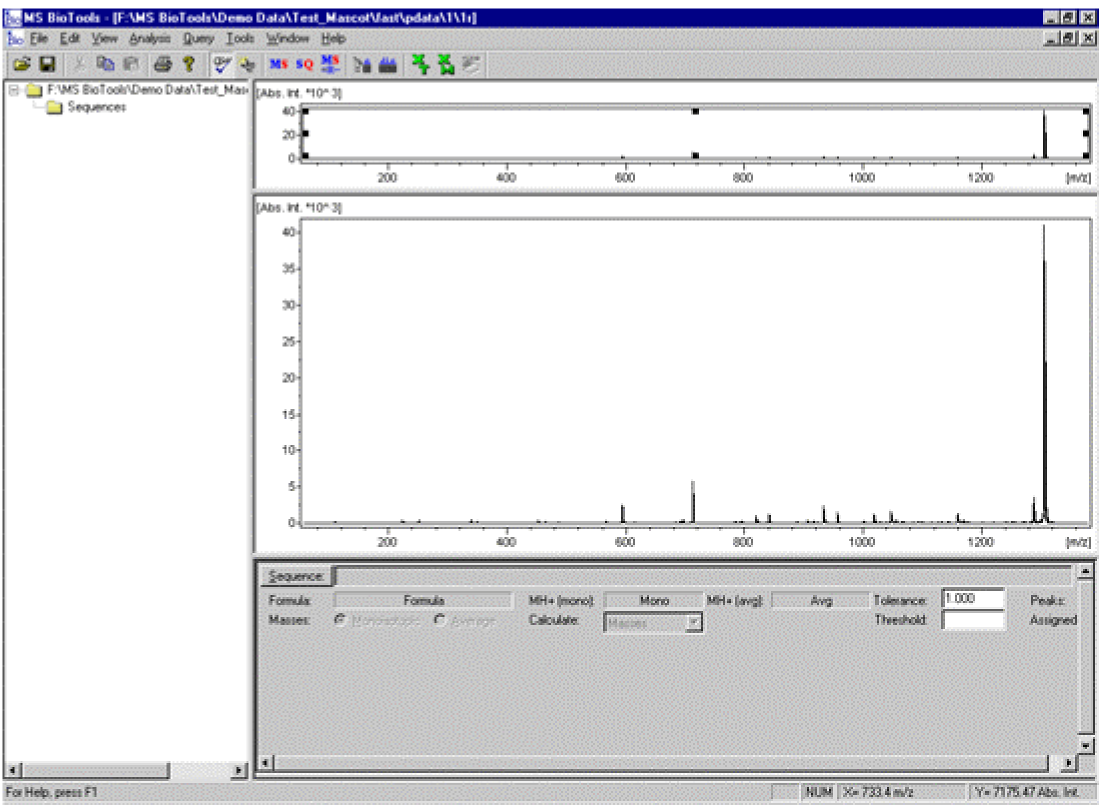


Figure 2-13, Start of Mascot search

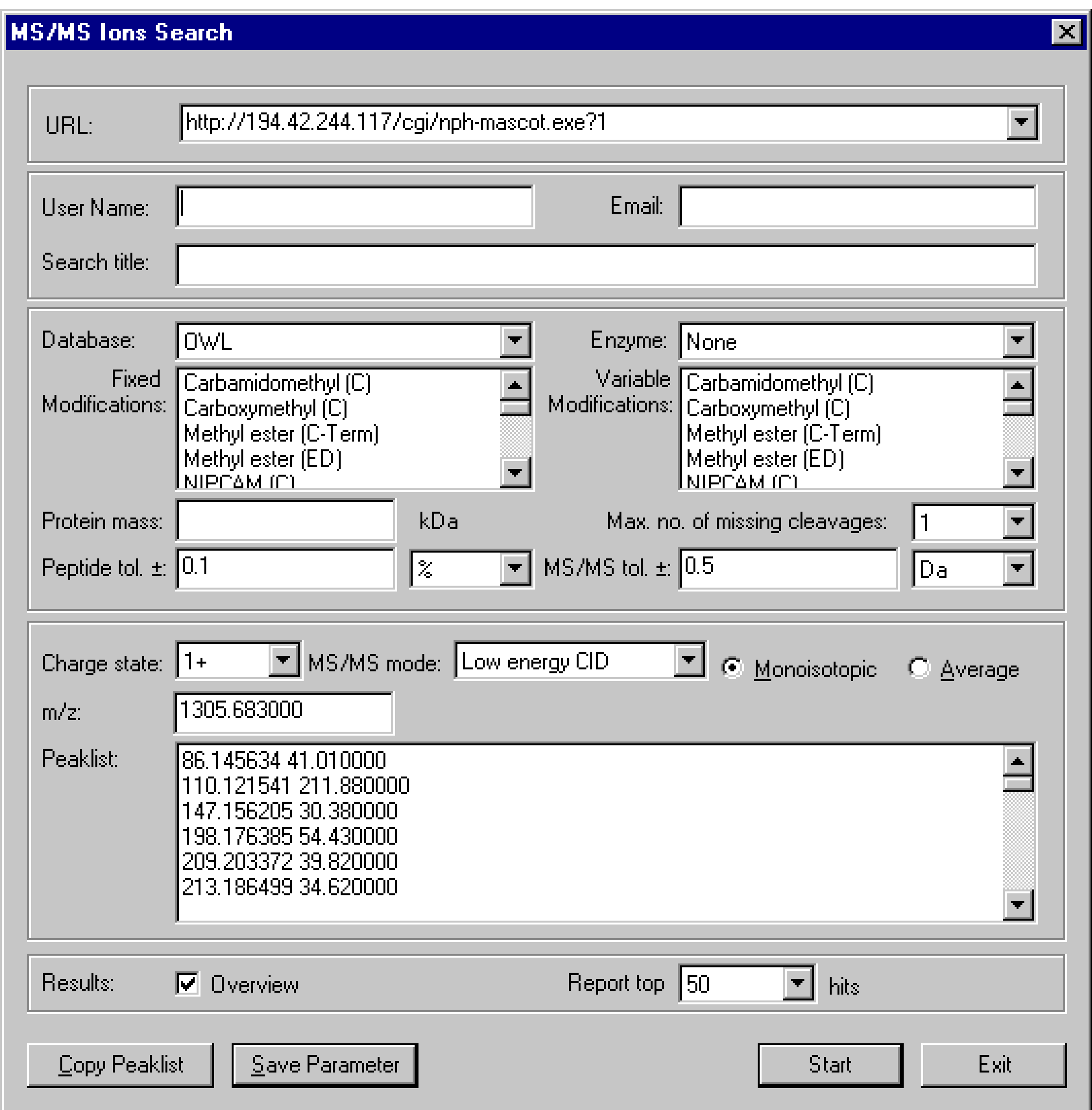


Figure 2-14, MS/MS Ion Search Parameter

Next click to the button "Mascot MS/MS Ion Search" and start the query using the "Start" button with the parameters: Enzyme **Lys C**, Fixed Modifications **Carbamidomethyl (C),** Variable Modifications **none**, Peptide Tol. **50 ppm**, MS/MS Tol. **0. 5 Da**, Charge state **+1**, MS/MS mode **PSD**, **monoisotopic**, Report Top **10** hits. For details of Mascot query parameters see http://www.matrixscience.com/, follow the link

to "Mascot" - "MS/MS Ion Search" – "Search Form" and select any underlined entry for detailed information. After a short period, the results are shown in the browser window. To transfer the results use the "Get Results" button. Leave the results window (use the "Exit" button), close the "MS/MS Ion Search" dialog and you will receive the annotated spectrum.

Note ***The MASCOT top score protein name is color-coded in green in the treeview. The Bruker MS/MS score is applied to each peptide sequence and all sequences are sorted according to the Bruker score (in brackets left of the peptide sequence).***

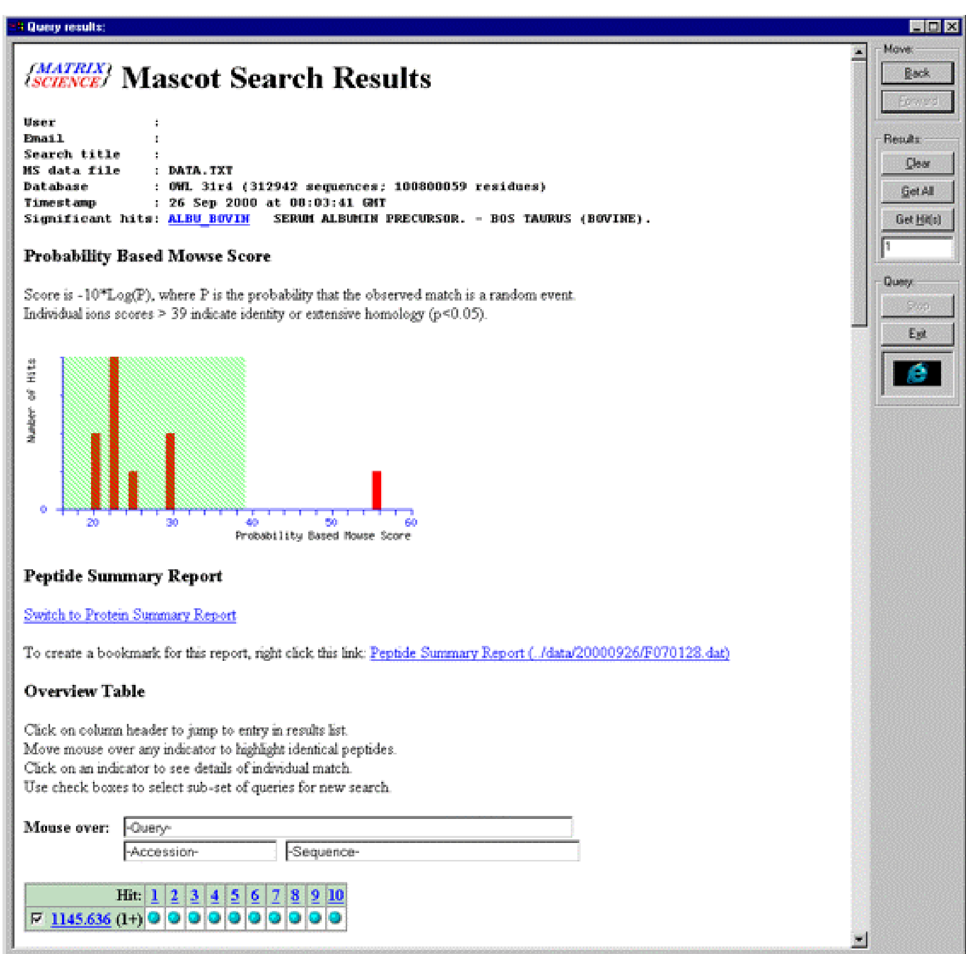


Figure 2-15, MS/MS Ion search results

3 Menu Bar, Toolbar, Status Bar and Context Menus

The BioTools window is separated into several smaller windows shown in Figure 3-1. Each window (see chapter 3.7), the toolbar (see chapter 3.3.1) and the status bar (see chapter 3.3.2) can be hidden. The Windows can also be arranged by setting the cursor on the border between two windows and move the border with the mouse during the left mouse button is held.

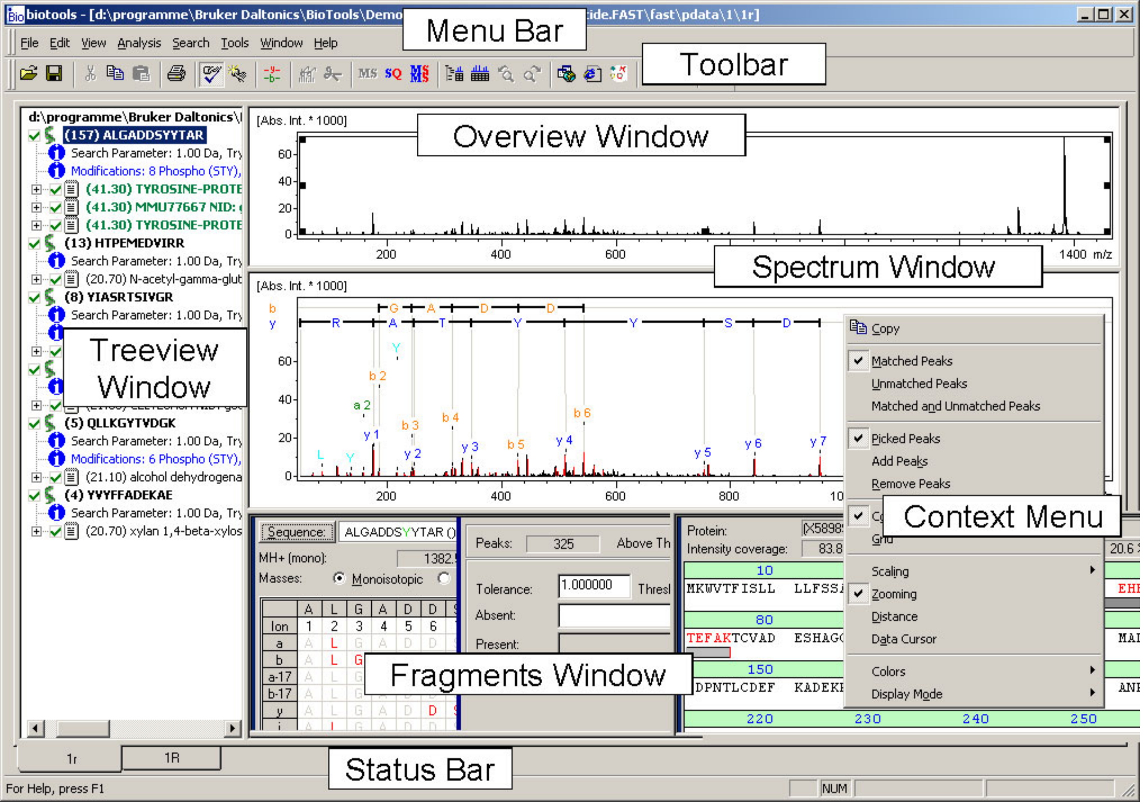


Figure 3-1, BioTools window

Some menu options are also available from the toolbar and with shortcuts (combination of Ctrl-key, and/or Shift-key and letter-key).
A detailed list of the toolbar buttons is given in chapter 3.3.1, the implemented shortcuts are shown in the respective menu. The contents of the menus and the toolbar depends on whether a spectrum is loaded (Figure 3-2) or not (Figure 3-3).

A detailed description of the treeview for fingerprint data is in the BioTools tutorial Sequence Database Searches from MALDI Peptide Mass Fingerprints .

Figure 3-2, Menu bar and toolbar (spectrum loaded)

If no spectrum is loaded, only the menus *File* (Figure 3-4), *View* (Figure 3-5), *Search* (Figure 3-6) and *Help* (Figure 3-7) are available. In the toolbar only Open, About and the connections to other programs (SequenceEditor, Internet browser) are activated (Figure 3-3).

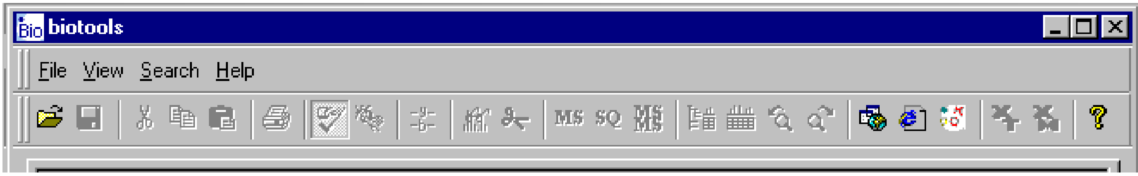


Figure 3-3, Menu bar and toolbar (no spectrum loaded)

In the menu *File* the four previous used files are also listed. A detailed description of the options *Find* (section 3.1.1), *Open Spectrum* (section 3.1.2), *Combine Spectra* (section 3.1.5), *Print Setup* (section 3.1.9) and *Exit* (section 3.1.12) of the menu *File* can be found in the respective sections of chapter 3.1 File Menu.

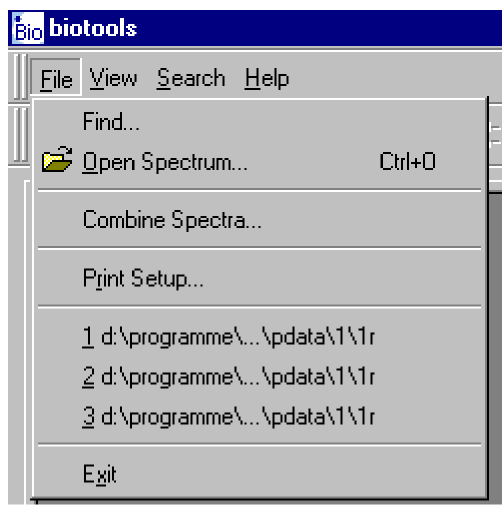


Figure 3-4, File menu (no spectrum loaded)

With Toolbar and Status bar in the menu *View* the upper toolbar and lower status bar can be shown or hidden.

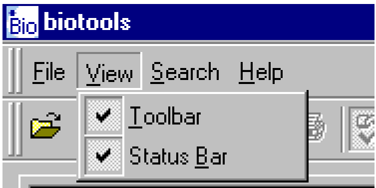


Figure 3-5, View menu (no spectrum loaded)

Figure 3-6, Search menu (no spectrum loaded)

In the menu *Help* the following options can be chosen:

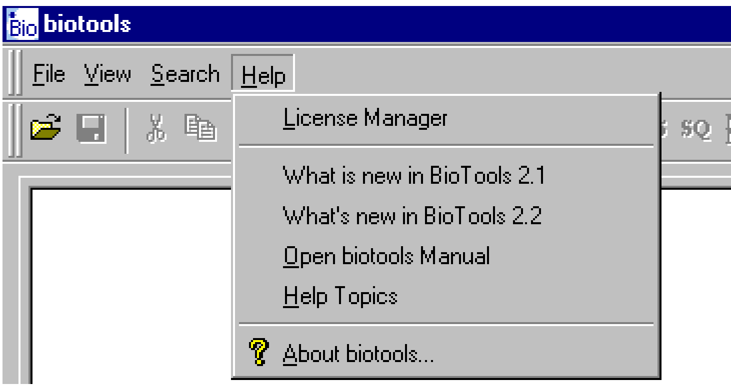


Figure 3-7, Help menu (no spectrum loaded)

3.1 File Menu

The content of the *File Menu* depends on whether a spectrum is loaded or not. If no spectrum is loaded, only the options *Find*, *Open*, *Print Setup* and *Exit* are available, also the four previous used spectra are listed. If one spectrum is loaded or more spectra are opened, the options of the *File menu* changes as shown in Figure 3-8.

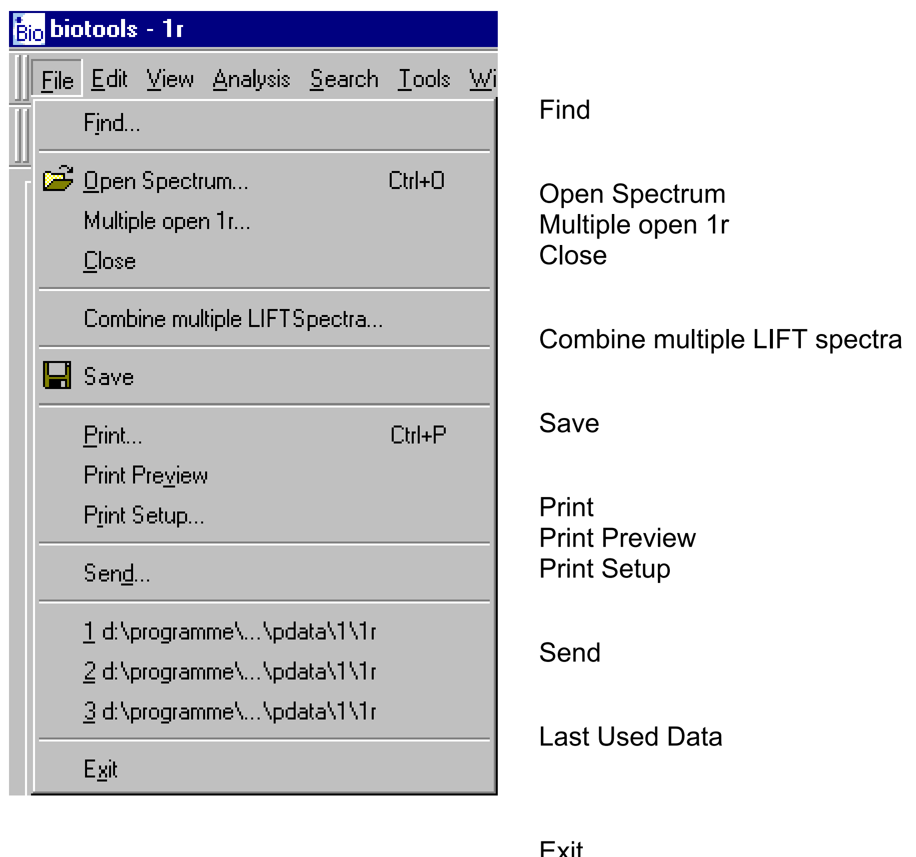


Figure 3-8, File menu with all options

3.1.1. Find

Because of the complex structure of, e.g., XMASS data directories it is often difficult to find the data. Use this entry to browse through a complete hard disk drive – even a drive in your network. First select the data type after entry "Named":

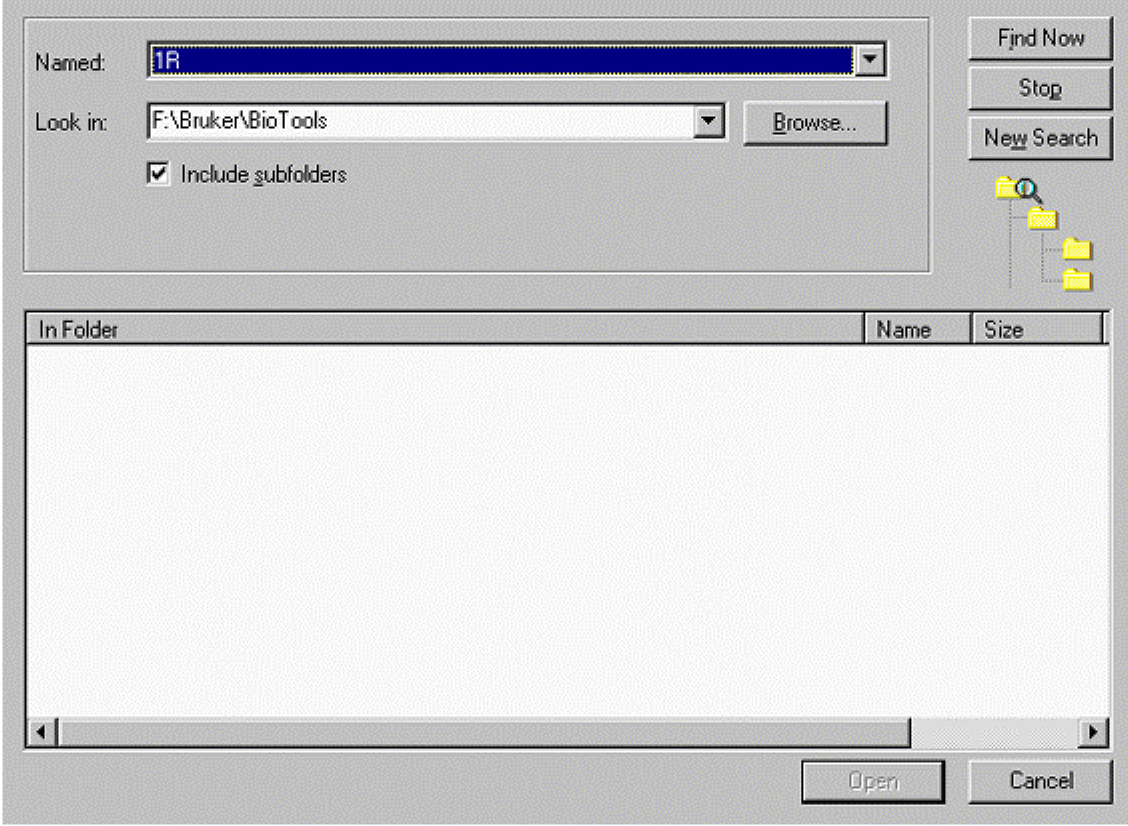


Figure 3-9, Find Files

Available data types are:
1R (processed spectra from XMASS/XTOF)
XY (deconvoluted spectra from XMASS/XTOF)
.BSC (processed data from Bruker Data Analysis 2.0)
.CSV (processed data from Bruker Data Analysis up to 1.6)
.MGF (Mascot Generic File)

Next specify the start directory of the search: either enter the complete path name after " Look in: " (e. g. "F:\DATA") or use the button Browse. This button opens a new dialog to select a disk drive and a folder.

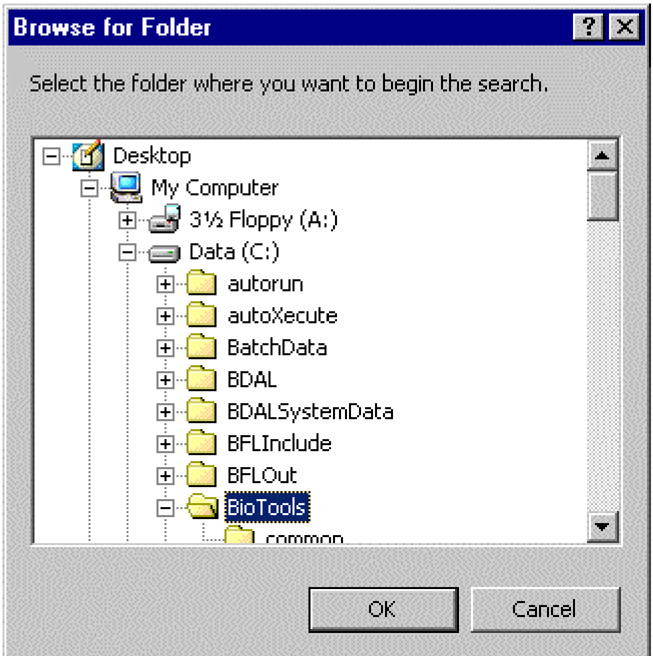


Figure 3-10, Selecting a start directory

If subfolders should be used for the search also select the checkmark beside " Include subfolders " and then start the search using the button Find Now. To interrupt the search use button Stop. To start a new search use button New Search. Use the standard MS Windows functionality to select one or more entries from the list of found files (mark the desired files by holding the Shift-key or Ctrl-key). Use button Open to load the selected data, use button Cancel to abort loading data files.

If you use the option *File - Find* during the same run of the program again the list of previous found files will be displayed again.

3.1.2. Open Spectrum

With this option an existing data file can be opened directly by double-clicking on it.

Available data types are:
1R (processed spectra from XMASS/XTOF)
XY (deconvoluted spectra from XMASS/XTOF)
.BSC (processed data from Bruker Data Analysis 2.0)
.CSV (processed data from Bruker Data Analysis up to 1.6)
.MGF (Mascot Generic File)

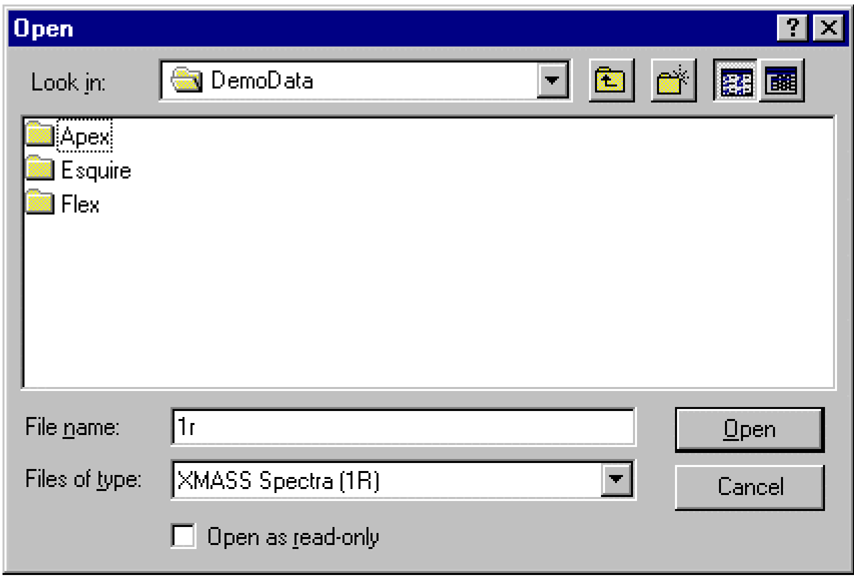


Figure 3-11, Select a spectrum

3.1.3. Multiple open 1r

Use this option "Multiple open 1r" entry from the menu "File" to open more than one already processed 1r data sets.

Select from the following "Browse for Folder" dialogue the desired directory. **All** 1r files in this folder and all subfolders will be opened.

Press the "ESC" key to interrupt the opening process. This may be helpful if you selected a folder with a huge number of 1r files.

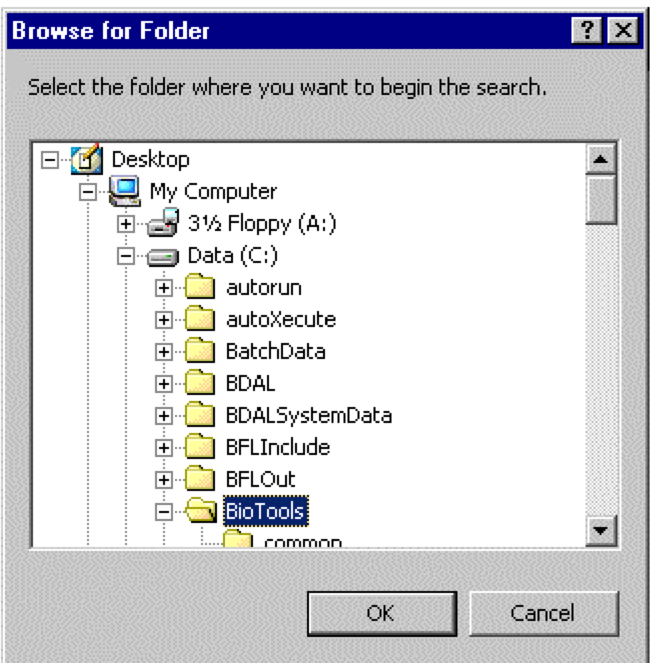


Figure 3-12, Selecting a start directory for multiple open 1r

3.1.4. Close

This menu option closes the active data file.

3.1.5. Combine multiple LIFT spectra

Multiple LIFT-TOF/TOF MS spectra can be combined and submitted to MASCOT as a single search providing dramatically increased search specificity.

Select all the spectra you want to combine. Either first use the Multiple open 1r option and open all the spectra or use the "Add" button from the "Combine spectra" dialog. You must select exactly one fingerprint spectrum (Type TOF) and an arbitrary number of MS/MS spectra (type PSD). Select the corresponding line from the list and click to the "Remove" button to remove an unwanted entry. The resulting combined MGF-file will be found in the same directory where the TOF spectrum is stored. The created MGF-file will be loaded automatically.

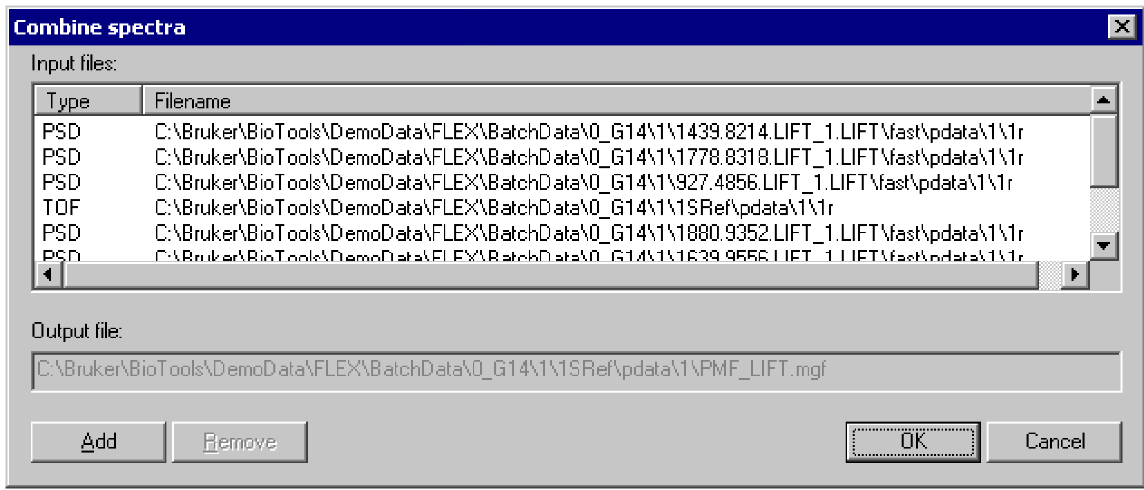


Figure 3-13, Combine spectra dialog

3.1.6. Save

This menu option saves the active data file.

3.1.7. Print

This menu option opens the Print window to choose printer and properties, print range and quantity of copies.

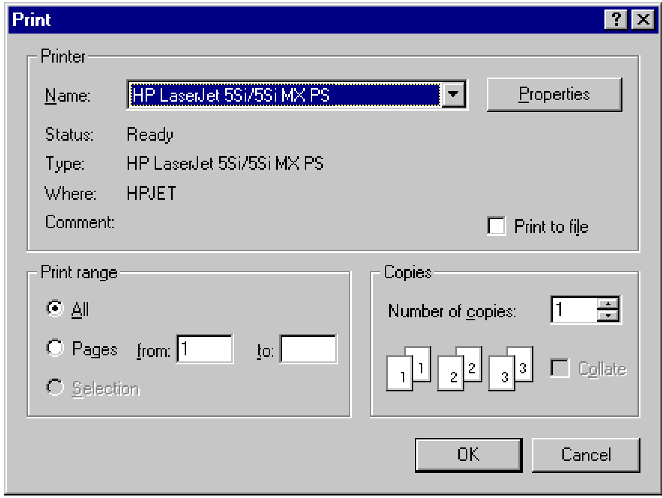


Figure 3-14, The print dialog

3.1.8. Print Preview

This menu option opens the print preview window. It shows how the printout will look and how many pages will be printed.

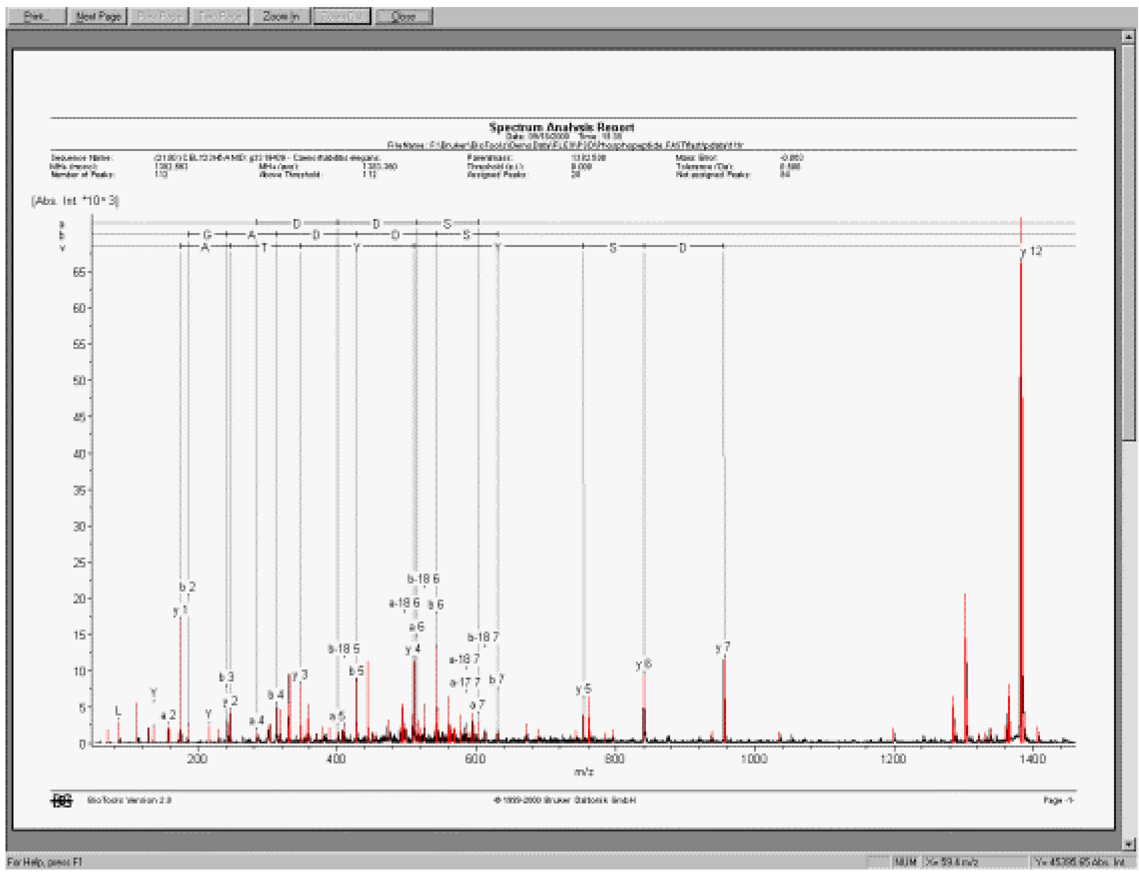


Figure 3-15, Print preview

3.1.9. Print Setup

This menu option opens the Print Setup window to choose printer and its properties, paper format and orientation.

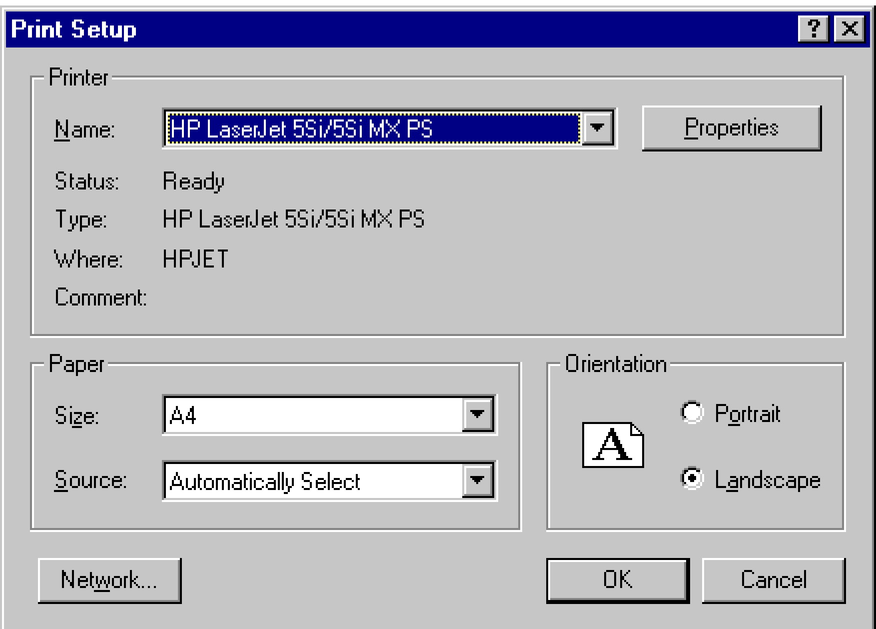


Figure 3-16, Print setup

3.1.10. Send

Opens an email program (if properly installed) to send the active spectrum to an email address as an attached file.

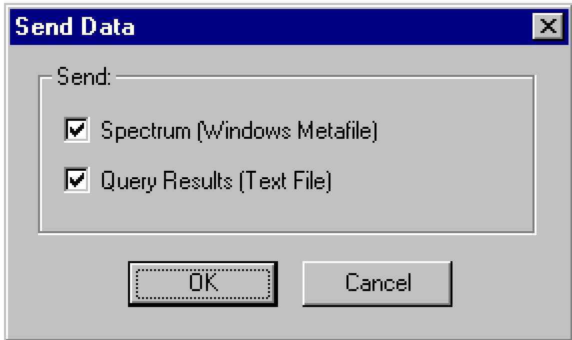


Figure 3-17, Send data dialog

3.1.11. Last Used Data

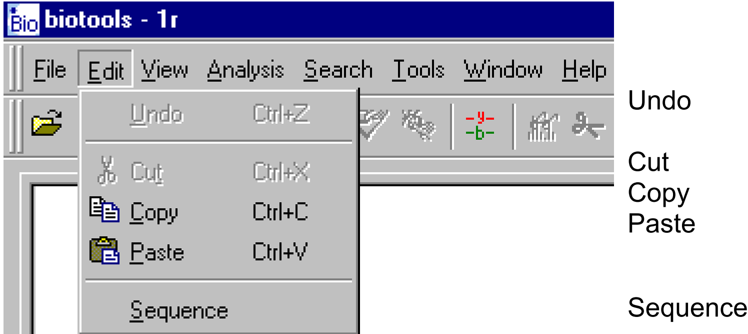
Gives a list of the four previously loaded spectra.

3.1.12. Exit

This menu option closes all opened files and finally the Bruker BioTools program.

3.2 Edit Menu

The *Edit Menu* is used to work with the clipboard. Its contents depends on previous actions, e. g., nothing is in the clipboard the option *Copy* is gray. It is important for working with other programs (e. g., table calculating program, word processing program).



Undo

Cut
Copy
Paste

Sequence

Figure 3-18, Edit menu

3.2.1. Undo

The recently last actions can be canceled / undone.

3.2.2. Cut

The marked sequence, number or text is deleted and copied into the clipboard.

3.2.3. Copy

The marked sequence, number or text is copied into the clipboard.

3.2.4. Paste

The content from the clipboard is inserted into the actual position.

3.2.5. Sequence

With this option the program SequenceEditor is opened and a sequence can be edited and modified (see Start SequenceEditor).

3.3 View Menu

With the *View Menu* the view of the windows can be arranged and the functions of the cursor can be changed.

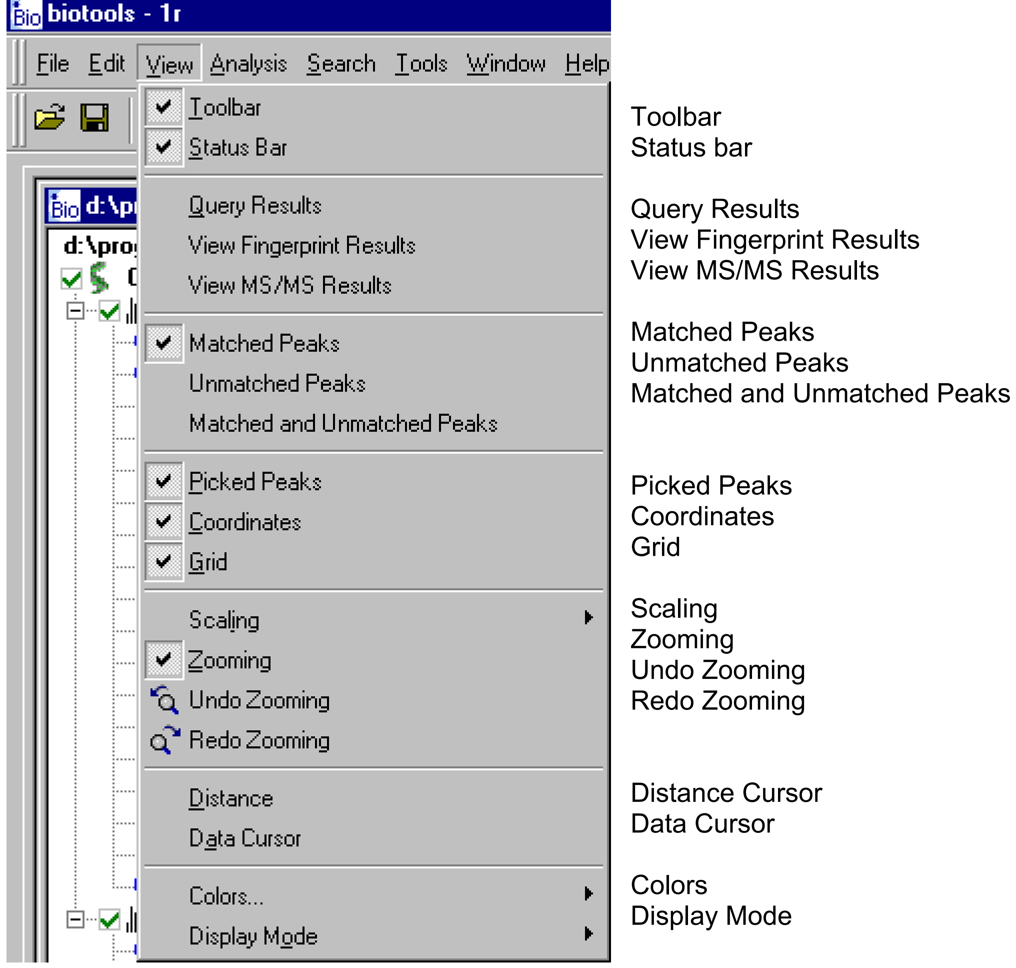


Figure 3-19, View menu (standard)

3.3.1. Toolbar

By activating this option the upper toolbar is shown (standard) or hidden. The toolbar can also be moved with the mouse. Clicking on the background of the toolbar with the left mouse button and move the toolbar with held mouse button to the desired position.

Figure 3-20, Toolbar

Toolbar button	*Menu option*	Shortcut / Function key	Description
	File – Open Spectrum	Ctrl + O	Opens a file manager
	File – Save	Ctrl + S	Saves the state of the active data file
	Edit – Cut	Ctrl + X	Deletes and copies into clipboard
	Edit – Copy	Ctrl + C	Copies from clipboard to cursor position
	Edit – Paste	Ctrl + V	Pastes from clipboard to cursor position
	-	-	Prints the active data file immediately in accordance to the Print Setup
	Analysis – Check Sequence	-	Changes the fragment window to check sequence mode
	Analysis – DeNovo Sequencing	-	Changes the fragment window to *DeNovo* sequencing mode
	Analysis – Annotation Parameter	-	Opens the annotation options dialog box
	Search – Search Mass Search SequenceEditor	-	Opens the Search for mass dialog box in SequenceEditor (*Search – Mass Search*)
	Search –Digest SequenceEditor	-	Opens the Perform Digest dialog box in the SequenceEditor (*Search – Perform Digest*)
	Search – Mascot Peptide Mass Fingerprint	-	Opens the internet search via Peptide Mass Fingerprint
	Search – Mascot Sequence Query	-	Opens the internet search via Sequence Query
	Search – Mascot MS/MS Ion Search	-	Opens the internet search via MS/MS Ion Search

	Window – Show/Hide Treeview	-	Show or hide the treeview window
	Window – Show/Hide Fragments	-	Show or hide the fragment window
	View – Undo zooming	-	The previous zoom action is undone
	View – Redo zooming	-	The previous zoom action is redone
	Search – Mascot Batch	-	Opens the Mascot Batch Mode window
	Window – Show/Hide Browser Window	-	Starts the Internet Browser to get Query results
	Edit – Sequence…	-	Loads a sequence into the Sequence Editor and starts this program
	Tools – Start XTOF NT	-	Starts the XTOF NT program
	Tools – Start XMASS NT	-	Starts the XMASS NT program
	Help – About BioTools	-	Opens About BioTools window

3.3.2. Status Bar

By activating this option the lower status bar is shown (standard) or hidden. In the left bottom corner a short help text corresponding to the cursor position and actions is given. The next boxes show the activated caps lock and the numeric function of the numeric block. In the right corner the x-position and y-position (or x-distance) of the cursor (if not deactivated) is given.

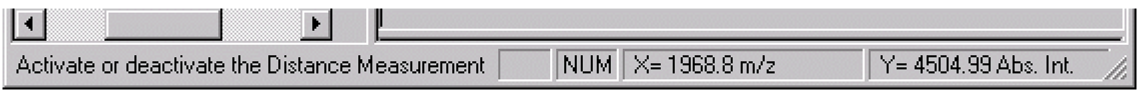


Figure 3-21, Status bar

3.3.3. Query Results

This menu option opens the query results window and shows the previous results for the active data file, provided a Get Results operation was performed at that time.

Clear	Deletes all imported entries in the treeview
Get All	Imports the whole result list into the treeview
Get Hit(s)	Imports the selected candidate sequences into the treeview ("1,4-7" imports the sequences 1,4,5,6 and 7)

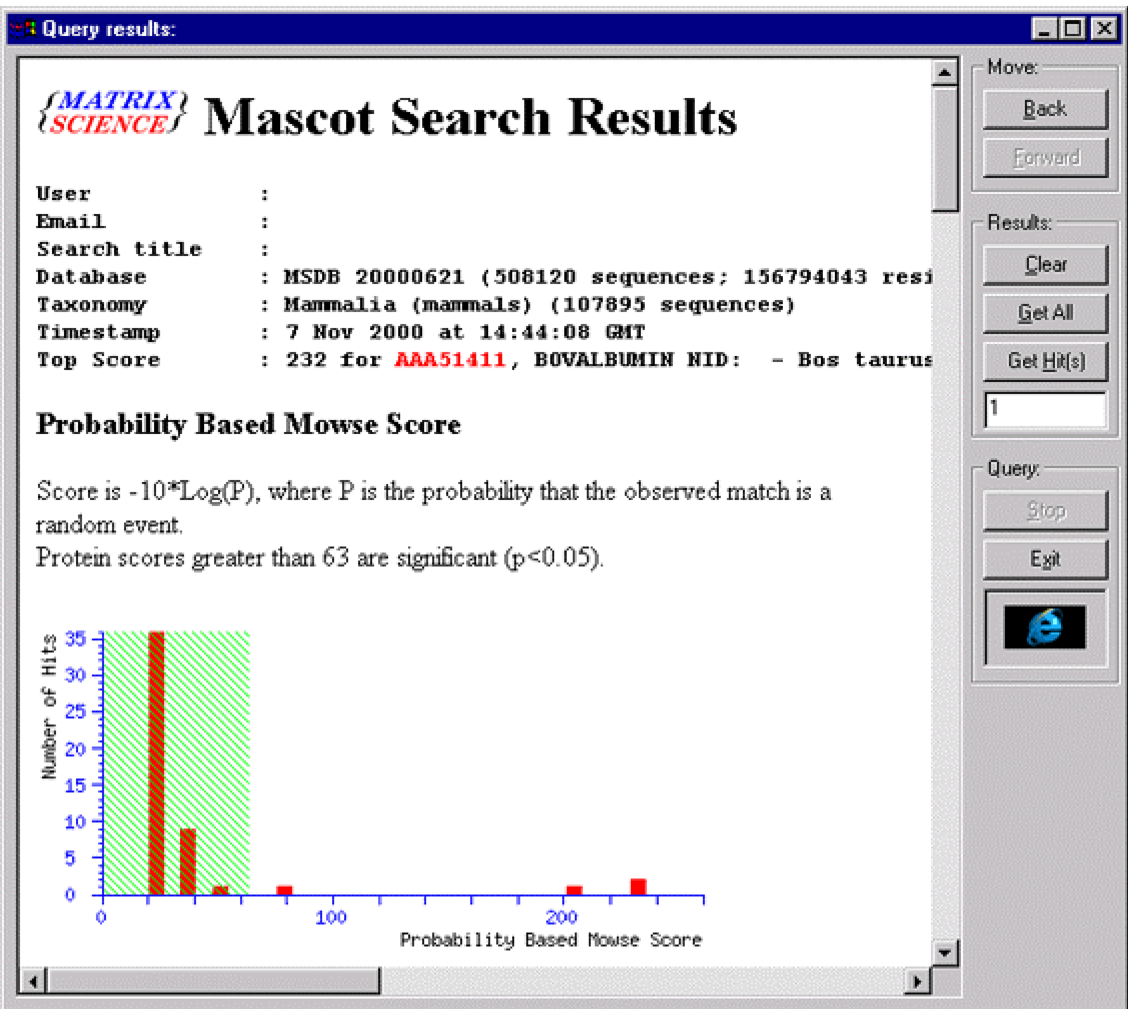


Figure 3-22, View of Query Results window

3.3.4. View Fingerprint Results

Changes the Fragments window to the fingerprint results.

3.3.5. View MS/MS Results

Changes the Fragments window to the MS/MS results.

3.3.6. Matched Peaks

By choosing this option all matched peaks will be labeled with mass + sequence position (or fragment type in case of MS/MS), and the parent mass, respectively.

3.3.7. Unmatched Peaks

By choosing this option all unmatched peaks - which are not used in the sequence - will be annotated with their respective mass.

3.3.8. Matched and Unmatched Peaks

By choosing this option all matched and unmatched peaks will be labeled with mass + sequence position (or fragment type in case of MS/MS), and the parent mass, respectively. Matched and Unmatched masses are displayed in colors.

3.3.9. Picked Peaks

By choosing this option all peaks of the peak list are shown in the spectrum window in a line spectrum.

3.3.10. Coordinates

By activating this option the x-coordinate and y-coordinate (or distance) are shown in the lower status bar. This depends whether the cursor has been set from zoom cursor (standard, chapter 3.3.11) to distance cursor (chapter 3.3.12), data cursor (chapter 3.3.13) or – by deactivating all cursors – to move cursor.

3.3.11. Grid

By activating this option a grid is shown in the spectrum window. The width of the grid is set automatically corresponding to an optimal view in the spectrum window and not to overload the spectrum window.

3.3.12. Scaling

With this option the scaling of the spectrum window can be changed simply by function keys, shortcuts and manually.

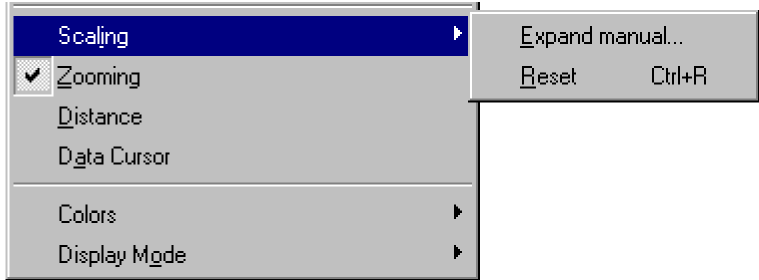


Figure 3-23, Scaling options and function keys

Options	Shortcut / Function key	Description
Expand manual	-	Opens the Manual Scaling window to enter the x-range and y-range manually Figure 3-2)
Reset	Ctrl + R	Moves to total overview

By choosing the option *Expand manual* the Manual Scaling window pops up. The mass range and/or intensity range can be entered.

Click on the [OK] button to accept the entered values for the spectrum window.

Click on the [Reset] button to get the whole spectrum back into the spectrum window.

Click on the [Cancel] button to leave that window and to get back to the previous range.

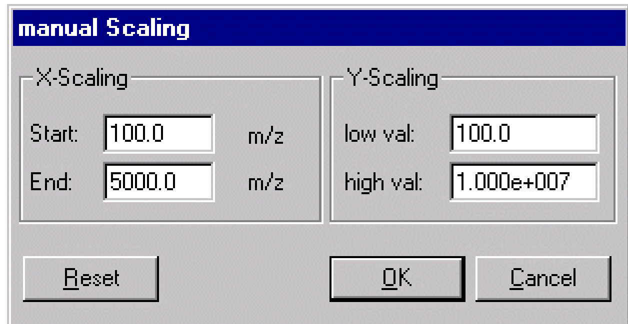


Figure 3-24, Manual Scaling window

3.3.13. Zooming

This option is normally activated; the cursor shows a zoom glass (Figure 3-25). To zoom in an area, mark the area with the zoom cursor while holding the left mouse button. Release the left mouse button when the area is marked, it will be zoomed to this area. To get back to the previous view with the mouse cursor zoomed area click once in the spectrum window, to get the whole spectrum double-click in the spectrum window.

By deactivating this option, the cursor changes to a move cursor (Figure 3-26).

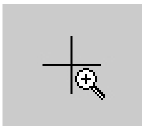


Figure 3-25, Cursor with zoom glass

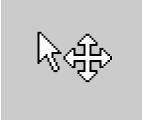


Figure 3-26, Cursor as move cursor

The cursor changes back to the zoom or move cursor when it has been set to distance cursor (chapter 3.3.12) or data cursor (chapter 3.3.13) and the right mouse button is pressed to end this two special function of the cursor.

3.3.14. Undo Zooming

With this option the last zoom factors and views will be undone.

3.3.15. Redo Zooming

With this option the previous undone zoom factors and views will be redone.

3.3.16. Distance Cursor

This function is helpful to mark a distance between two peaks.

By activating this option the cursor changes to a double-arrow (Figure 3-27) and two vertical lines appear. Simultaneously the y-coordinate disappears and a dX-distance is shown in the status bar(Figure 3-28). The polarity of the dX-distance depends on which line the cursor is positioned.

Now move the vertical line on the right on the desired peak, which is the higher one from both. If the cursor is on the correct position, click the left mouse button. The cursor jumps to the second line, which is on the left side from the first one. Now move

the line on the desired peak, which is the lower one from both. The position of the actual line and the distance to the other line is shown in the status bar.

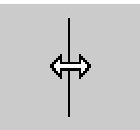


Figure 3-27, Distance cursor

Figure 3-28, X-coordinate and distance

By clicking the left mouse button the cursor jumps between the two lines back and forth to correct their position.

To end this function click the right mouse button, the cursor changes back automatically to zoom or move cursor.

3.3.17. Data Cursor

This function is helpful to show the intensity of a peak in the status bar.

By activating this option the cursor changes to a T-cursor (Figure 3-29).

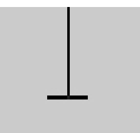


Figure 3-29, Data cursor

Now move the vertical line across the curve on the desired peak. The intensity of the actual position is shown in the status bar (Figure 3-30).

Figure 3-30, Data cursor on peak

To end this function click the right mouse button, the cursor changes back automatically to zoom or move cursor.

3.3.18. Color

The color for Background, Spectrum, Peak Picking and Annotation can be chosen (Figure 3-31) from a list of custom colors (Figure 3-32). They can also be defined manually (Figure 3-33).

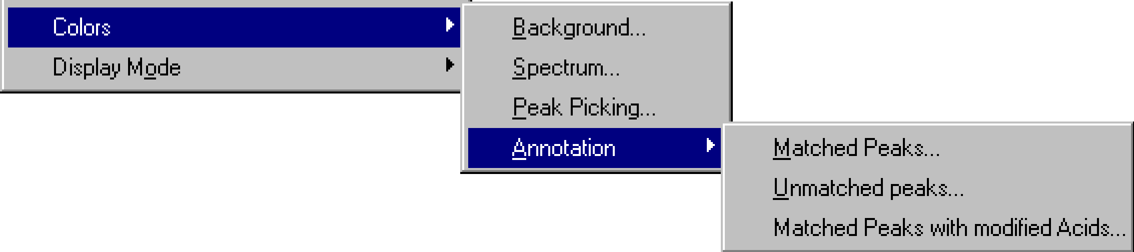


Figure 3-31, Colors options

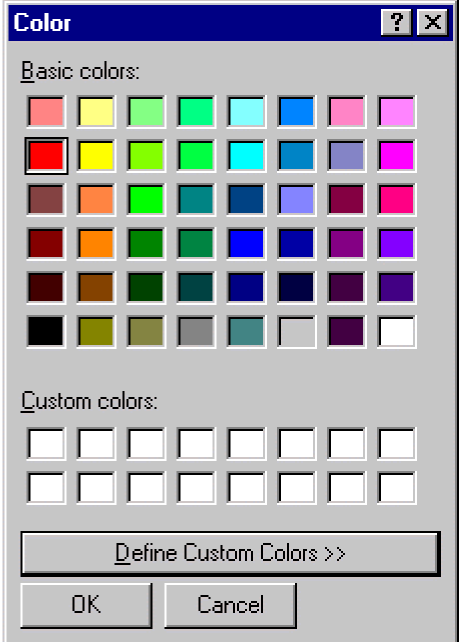


Figure 3-32, Choosing a color

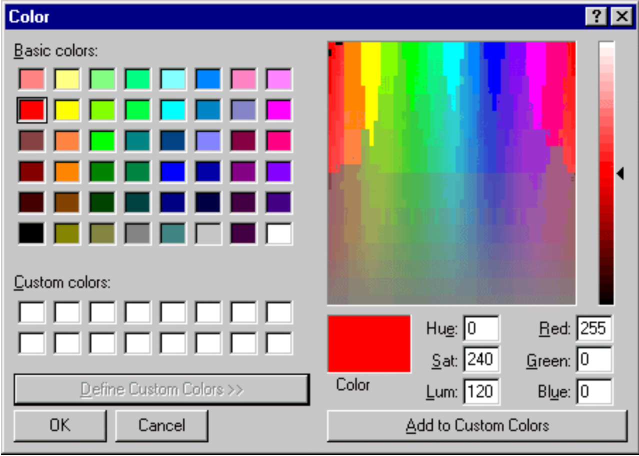


Figure 3-33, Choosing or defining a color

3.3.19. Display Mode

With this option the display of the data set on the screen can be modified in different ways.

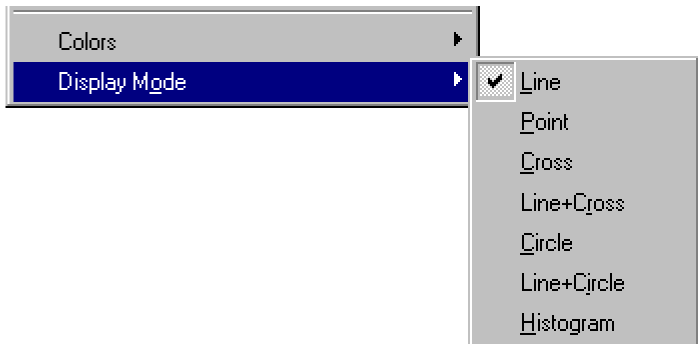


Figure 3-34, Display Mode options

Options	**Display**
Line	Data points are connected with a line
Point	Data points are shown as points
Cross	Data points are shown as crosses
Line+Cross	Data points are connected with a line and shown as crosses
Circle	Data points are shown as circles
Line+Circle	Data points are connected with a line and shown as circles
Histogram	Data points are shown as vertical lines

3.4 Analysis Menu

With the *Analysis Menu* the sequence form can be chosen and shown in the fragments window, the list of ions can be selected, and the threshold and the art of annotation can be modified.

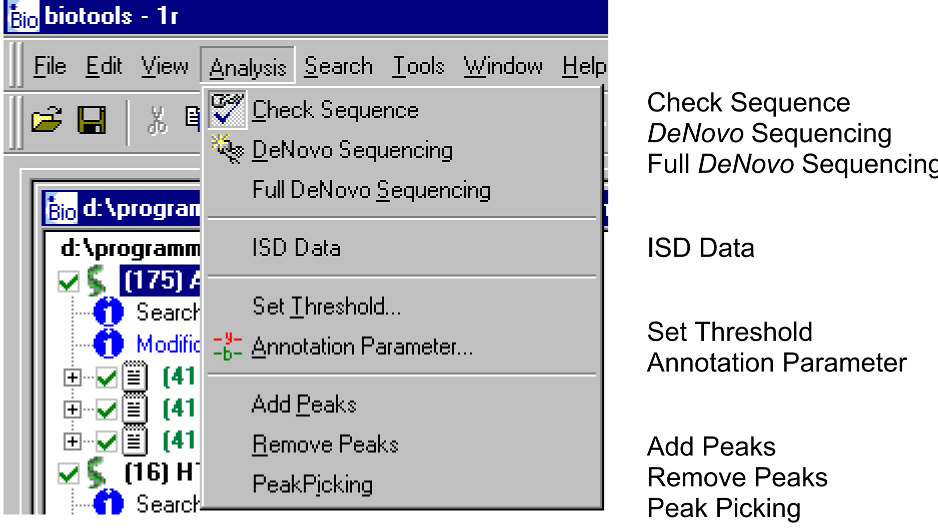


Figure 3-35, Analysis menu

3.4.1. Check Sequence

With the option *Check Sequence* a sequence can be chosen, their amino acids are shown in the fragments window and the data will be annotated in the spectrum window corresponding to this sequence.

Sequence: SYSMEHFRWGKPVGKKR ()

MH+ (mono): 2093.086 MH+ (avg): 2094.464 Tolerance: 0.500 Peaks: 140 Above Threshold: 87

Masses: Monoisotopic Average Calculate: Masses Threshold: 1000.000 Assigned: 27 Not Assigned: 60

	S	Y	S	M	E	H	F	R	W	G	K	P	V	G	K	K	R	Ser	Tyr	Ser	Met	Glu	His	Phe	Arg	Trp
Ion	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	3	4	5	6	7	8	9
a	S	Y	S	M	E	H	F	R	W	G	K	P	V	G	K	K	R	60.044	223.108	310.140	441.180	570.223	707.282	854.350	1010.451	96.5
b	S	Y	S	M	E	H	F	R	W	G	K	P	V	G	K	K	R	88.039	251.103	338.135	469.175	598.218	735.277	882.345	1038.446	24.5
y	S	Y	S	M	E	H	F	R	W	G	K	P	V	G	K	K	R	175.118	303.213	431.308	488.330	587.398	684.451	812.546	869.567	55.6
	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Arg	Lys	Lys	Gly	Val	Pro	Lys	Gly	Trp

Figure 3-36, Fragments window with Check Sequence parameter and annotation

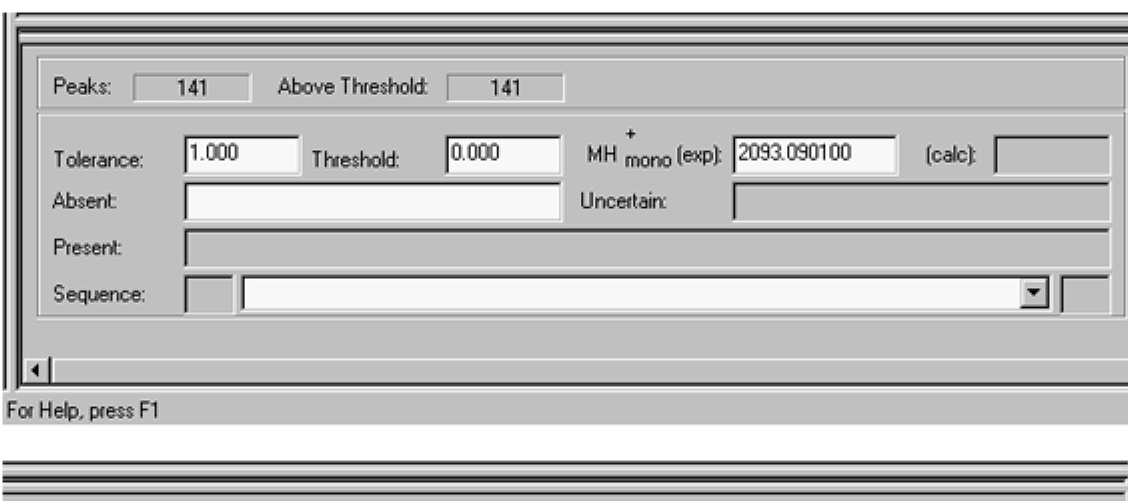
Contents of the fragments window:

Sequence Button	sends the active sequence to the SequenceEditor. The sequence and its title (in brackets) is shown in the field to the right
Symbol	If there is a fault in the sequence, a message will be pop up after opening. In addition a symbol is placed in the title.
MH^+ (mono)	The monoisotopic mass is shown
MH^+ (avg)	The average mass is shown
Masses	(Monoisotopic / Average) Choose whether monoisotopic or average mass is used for calculation
Calculate	Choose how the result of calculation is shown below None only the list of proteins is shown Masses proteins and their masses are shown Errors [Da] the calculation faults are given in [Da] Errors [ppm] the calculation faults are given in [ppm]
Tolerance	Enter the tolerance (in m/z) of calculation, which calculation results will be excepted
Threshold	Enter the threshold manually to exclude peaks below threshold; alternatively enter the threshold with the Threshold option (chapter 3.4.5)
Peaks	Number of total recognized peaks in the original data file
Above Threshold	Number of peaks above the entered threshold
Assigned	Number of peaks, which are assigned to a possible sequence
Not Assigned	Number of peaks, which are not assigned to a sequence; The sum of Assigned and Not Assigned peaks is the number of peaks shown in the Above Threshold box

To view an already existing sequence via the Treeview window, select the desired sequence by clicking on it with the left mouse button.

3.4.2. DeNovo Sequencing

Under the option *DeNovo* sequencing, tools are available to extract sequence information from MS/MS data. Depending on the data quality and type, these can be either rather complete sequence suggestions or only "seed" sequences or sequence tags, which require further interactive work.



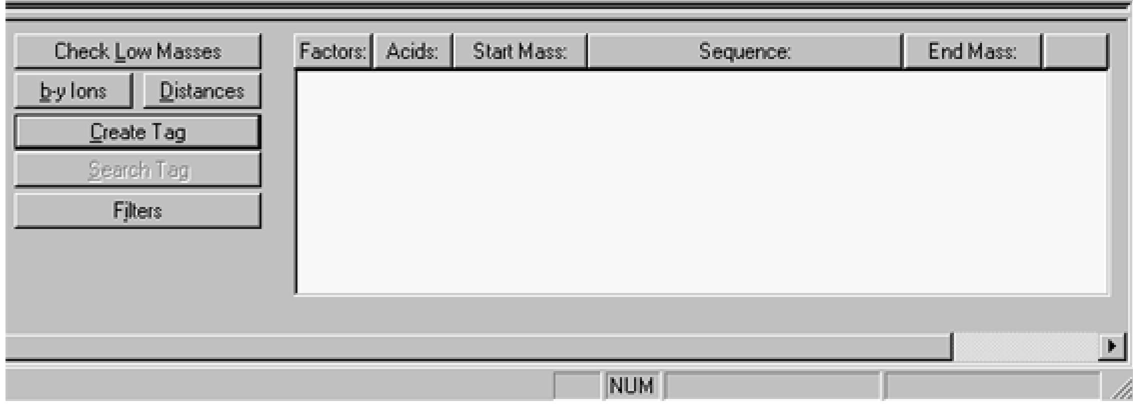


Figure 3-37, Fragments window with *DeNovo* Sequencing parameter (window is split in two for better view**)**

Contents of the fragments window:

Peaks	Number of total recognized peaks in the data file; these peaks are shown in the spectrum window by activating Picked Peaks
Above Threshold	Number of peaks above the entered threshold
Tolerance	Enter the tolerance (in m/z), which calculation results will be expected
Threshold	Enter the threshold manually to exclude peaks below the threshold; alternatively enter the threshold with the Threshold option (chapter 3.4.5)
MH^{+}_{mono}(exp)	The monoisotopic parent mass is displayed and can be edited
MH^{+}_{mono} (calc)	The calculated mass is displayed here
Absent	Shows the amino acid residues suggested to be absent by the Low Masses button operation
Uncertain	Shows the amino acid residues suggested to be uncertain by the Low Masses button operation
Present	Shows the amino acid residues suggested to be present by the Low Masses button operation
Sequence (scroll)	Shows the results of Create Tag arranged by decreasing probability: (start mass) amino acids (end mass)
Check Low Masses	Searches for existing low masses (i-type ions) in the data file, the absent, uncertain and present low masses are shown in the field Absent , Uncertain and Present Low masses can be entered in the absent field manually, their color can be modified by changing it via the menu *Analysis - Annotation - Parameter* and the color of the i-ions.

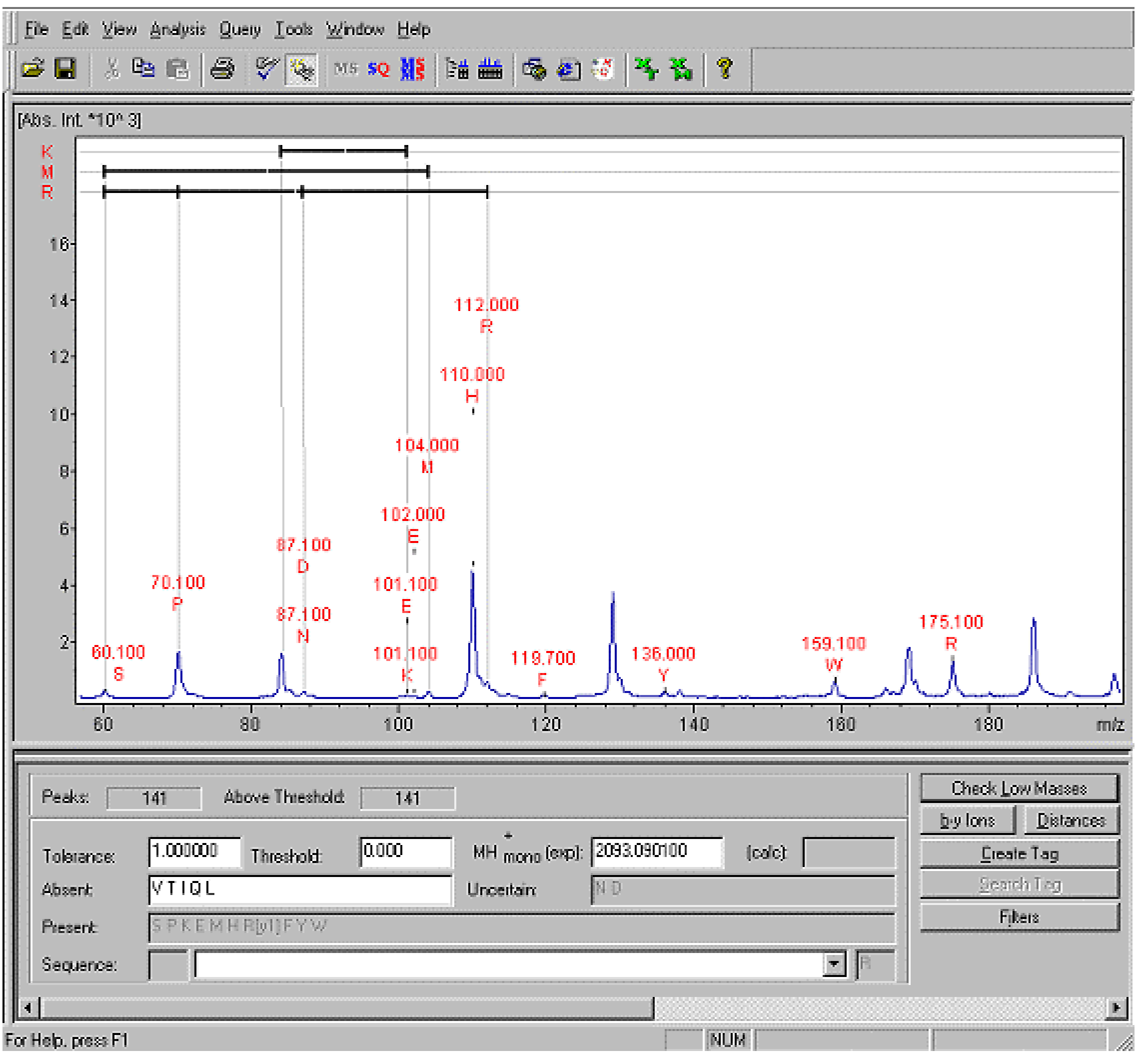


Figure 3-38, Found low masses

b-y Ions

Searches for pairs of b and y ions in the data file, which result from the dissociation of the same peptide bond; the results are shown in the spectrum window. **Important**: The mass error must be set 2x as high as the actual mass error for a single fragment ion.

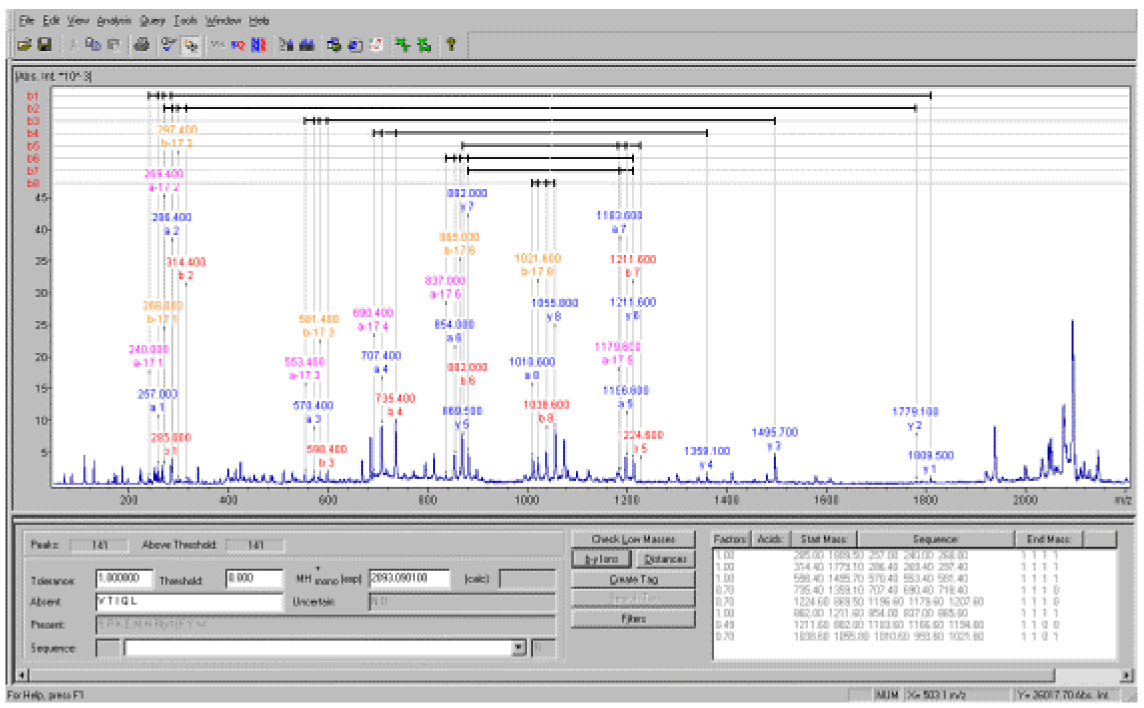


Figure 3-39, View of b-y ions

Distances

DeNovo Sequencing is done with the Distance Algorithm (for esquireSeries, MALDI-ISD and APEX).
BioTools can distinguish between MS and MS/MS data automatically, which activates the relevant display, calibration and analysis options. ISD data, however, are not recognized as MS/MS data automatically and must be specified manually. Enable this entry to perform PSD data processing for ISD data (e. g. for *DeNovo* sequencing). See also the tutorial on the analysis of ISD data.

Create Tag

Searches for sequences matched to the b-y ions; the results are shown in the spectrum window and listed in the Sequence scrollbar (decreasing probability).

The different colors shows either the sequence has been found only in one calculation direction (from low to high or reverse) or in both directions.
The colors can be changed in *Analysis - Annotation - Parameter* : the color of the b-ions (one direction) and y-ions (both directions).

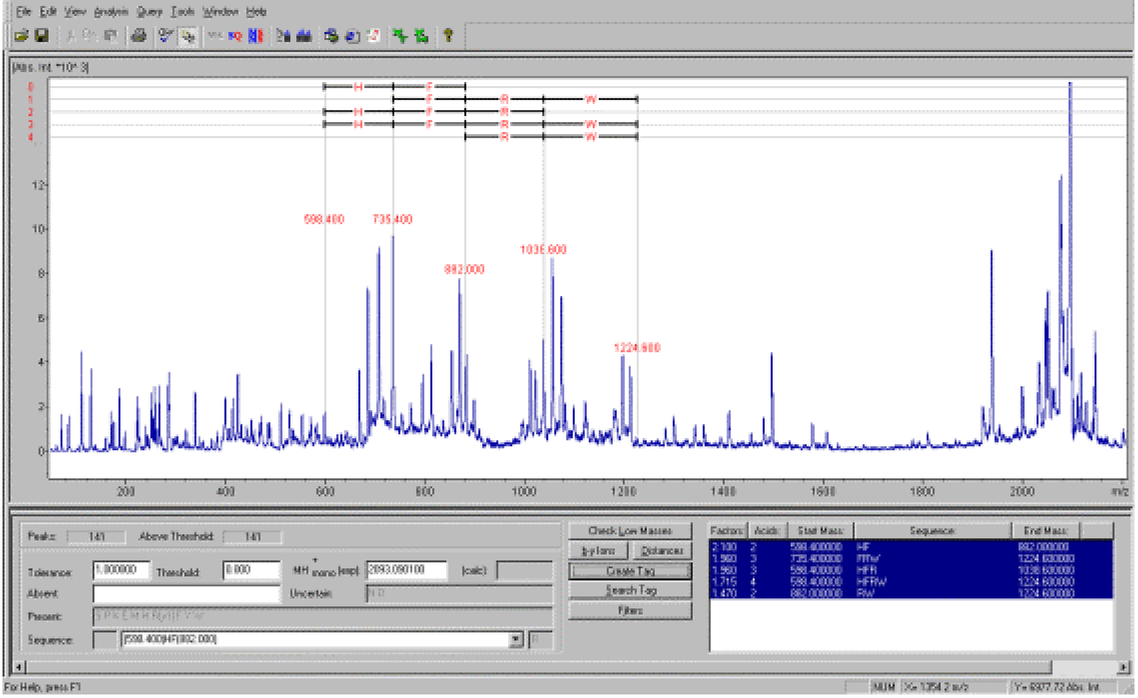


Figure 3-40,   Created tag

Starts the sequence tag search window for search via Internet (only available if a tag has been found) on PeptideSearch at EMBL.

URL: http://193.175.249.95:80/CGI/PPG.PeptSearchSequenceTags.acgi

Protein mass range from (kDa): 0 to (kDa): 300

Cleavage agent: Trypsin
Cysteine is: Cys
Oxidized Methionine

Peptide mass (neutral): 2092.082275 Monoisotopic mass
Mass accuracy: 0.3 Da
Peptide sequence tag: (598.400)HF(882.000)
Match regions: 1 and 2 and 3

Pattern match search by: B-type sequence ions
Edman type search by: N-terminal sequence extension
Allowed number of errors: 0
Cleavage specifities: N-terminal specifity C-terminal specifity
Results per page: 100

Copy Save Start Exit

Figure 3-41, Start of the search tag after creating a tag

Procedure:

To start *DeNovo* sequencing, enter the tolerance, the threshold and – if not done automatically – the monoisotopic parent mass. After this click the Check Low Masses button to get information about existing amino acid residues (PDS only!), the results can be changed by entering additional or deleting absent residues. I and Q should be generally included to avoid redundant sequence assignments.

For PSD data, press the b-y ions, the tag results and the search buttons.

Filters

Opens the filters dialog to set the valence of the ions and the minimum and maximum values of the search .

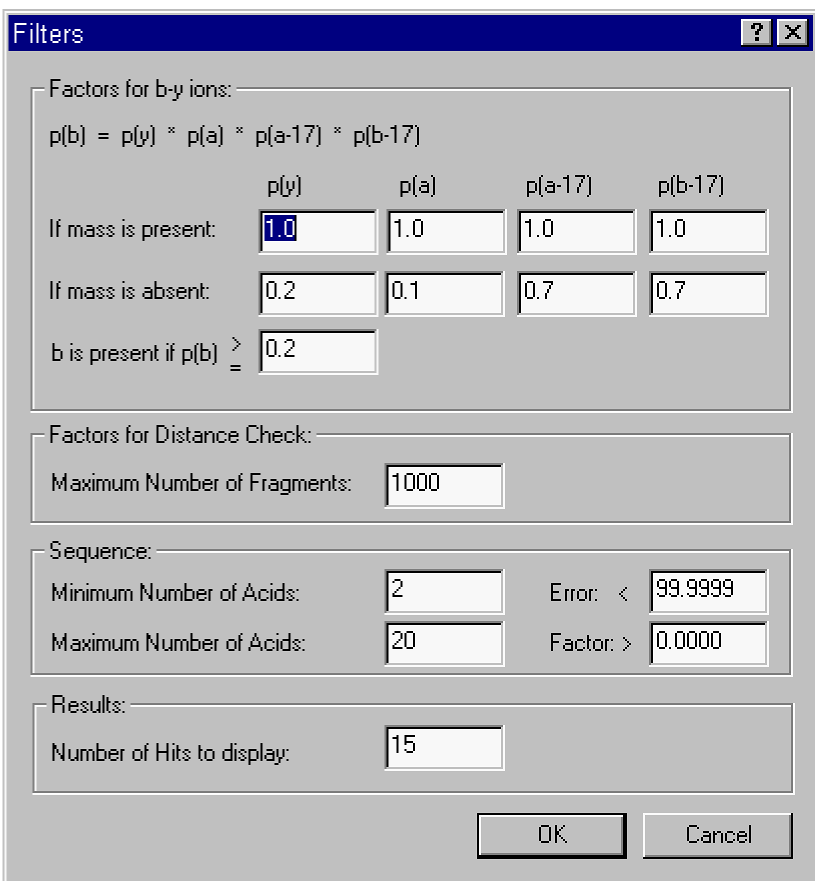


Figure 3-42, Filters dialog window with settings for PSD. esquireSeries settings should be 1.0 for "If mass is absent:" p(a), p(a-17) and p(b-17)

3.4.3. Full DeNovo Sequencing

Description of procedure

The DeNovo sequencing procedure used is based on the algorithm developed by Ishikawa, Niwa and Sasagawa and performs a stepwise build-up of the amino acid sequence beginning with a partial sequence of 3 amino acids. The sequence score is calculated according to a table optimized for the selected spectrum type. If that score is greater than the product of *(Partial) Stringency of calculation* * *SequenceLength* the partial sequence is stored for use in the next iteration of the algorithm, otherwise discarded. A sensible choice for the *(Partial) Stringency of calculation* parameter is therefore dependent on the scoring table for the peaks. If its value is low enough sequences are always returned even if there are very few peaks matched. A default value of 5 is appropriate in most cases, sometimes lower values are necessary to move the algorithm successfully past the initial stage, especially when there are few peaks in the spectrum at low masses. The internal scoring of sequences has been modified recently, each ion is now scored independently, the total score is then the sum of all ions found individually. This could be used in the future to make the scoring table user-defined. As a consequence of this change the **Stringency of calculation** should be within a range of 0 to 20, negative values are not longer necessary for this parameter since the resulting DeNovo scores are higher than before.

If more than *Number Of Seq. Candidates* (default value: 100) partial sequences have been obtained after the previous step only the best (highest internal score) *Number of Seq. Candidates* results will be used in the subsequent extension of the partial candidate sequences. A high *Number of Seq. Candidates* slows down the speed of calculation but ensures more variety in the resulting sequences.

If the mass of a growing peptide is close to the *Parent Mass* (+-*Peptide tolerance* given) the score of that sequence is calculated and if found to be above *(Full) Stringency of calculation* * *SequenceLength* the now complete sequence is stored as one of the final results. The *MS/MS tolerance* is the tolerance used for deciding whether the calculated peaks from the theoretical spectrum is found in the experimental Peak list.

An internal hint can be created (button **Define sequence tag**) from the experimental peaks based on the sequence tag functionality already available in BioTools. A separate dialog is used that gives the choice of either automatic search based on b/y ion pairs (same as already implemented in BT 2.1) or manual definition of an internal hint including the start mass. If available, an internal hint can be used by selecting it in the list box ‘Select hint sequence tag’.

The automatic search for present/absent amino acids based on their Immonium ions (button: **Low mass ion info**) has been limited to 6 acids: Y, F, W, H, L, and V, which are thought to be most reliable. Sequencing from both N-and C termini is done by default with merging of the best results.

The resulting sequences from the procedure are then searched for common substring s. A dialog box ('DeNovo Results') displays these short subsequences with a relative score based on their occurrence in the resulting sequences after the sequencing has finished. This dialog also contains a list box with all returned full sequences ordered according to BioTools score. The presence of checked short subsequences in the full sequences is indicated by a variation in background color of the amino acid residues concerned. The idea behind this was to find short significant sequence 'tags' that are common to most resulting sequences and are probably also contained in the correct sequence. One of these short subsequences can then also be used as internal hint for a subsequent DeNovo sequencing step or a tag search in a database (EMBL PeptsearchSequenceTag).

The DeNovo settings dialog is available for documents containing MS/MS data from the **Analysis** menu as **Full DeNovo Sequencing**. If the old DeNovo mode is active, a switch to the Check Sequence mode (see menu Analysis - Check Sequence) is performed to enable the viewing of sequence annotations etc. If the user activates a different spectrum during the calculation, the sequencing procedure is aborted. While the sequencing is running it is not possible to close any document or shut down the application before terminating it with the button '**Cancel Calculation**'.

For a more detailed description see the tutorial Full DeNovo Sequencing with BioTools.

3.4.3.1. DeNovo settings smart

Calculation parameters:

- The **Parent Mass** is set correctly for TOF-PSD data, for .bsc (Esquire) and ICR this value has to be set/corrected manually.
- The **tolerances** for **peptide** parent mass/**MSMS** peaks should be known by the expert user, useful defaults are: TOF-PSD 0.3/0.5 Da, Esquire 0.4/0.25 Da, ICR 0.2/0.01 Da. These values including the unit for peptide tolerance are saved with the document.
- The **Stringency of calculation** parameter refers to the internal scoring used by the sequencing algorithm; useful defaults are set for each spectrum type. A default of 5 is recommended, lower values tend to return more sequences while a higher threshold will reduce the number of results. The **Stringency of calculation** parameter (this is in fact the value for the final full sequence, the corresponding one for the partial sequence is calculated automatically from this value) is saved with the document.
- The **Spectrum Type** determines the ion types that are used internally for the scoring of the partial sequence. The internal scoring is optimized for each spectrum type.

- The **Advanced Settings** button brings up the values for **Number of partial sequence candidates** (100 is good for most applications) on a separate dialog. This value is saved with the data.

Sequence hints (always read from N-terminus):

- The **Low Mass Ion Info** button does an automatic search for the immonium ions of 6 Amino Acids: Y F W H L V, and for C terminal Amino Acids (all), the result of this is entered into the edit field for Absent/Present/C term hints. The user is advised to check the result of this analysis and remove uncertain entries before calculation.
- **Sequence Hints** can be given for the **N** (only one) and **C termini** (up to 4 different ones separated by blanks, e g K R). The resulting sequences will then contain these hints, when correct they improve the quality of the results and shorten the calculation time.
- Residues that must be **Absent amino acids** (are excluded beforehand from the sequencing) or **Present amino acids** (resulting complete sequences are checked after completion of sequencing) in the sequences can be entered separated by blanks. I and Q (if the MSMS tolerance is above 0.2) are entered here my default.
- The use of an internal hint is done by selecting it in the list box **Select hint sequence tag** if available. No internal hint is used when the default string 'none' is selected.
- The [Define_sequence_tag] button creates a modal dialog for definition of internal hints.
- The [Define_sequence_tag] button brings up the database tag search dialog (URL: http://194.94.45.86/CGI/PPG.PeptSearchSequenceTags.acgi at EMBL) with the internal hints from the settings dialog available (only works if internal hints are available in the settings dialog). This feature has been available since BT 2.1 as 'Search Tag'.
- The [Shrink/Expand] button hides/shows 4 list boxes for further selections.
- The [Remove unchecked] button removes unchecked entries from the BioTools treeview.
- The [Calculate] button starts the calculation and also does some checks on inconsistent settings (e g same Amino Acid in Absent and Present fields), if inconsistencies are found a message is displayed and the calculation aborted.
- The Calculation can be aborted while the [Cancel caluculation] button is active.

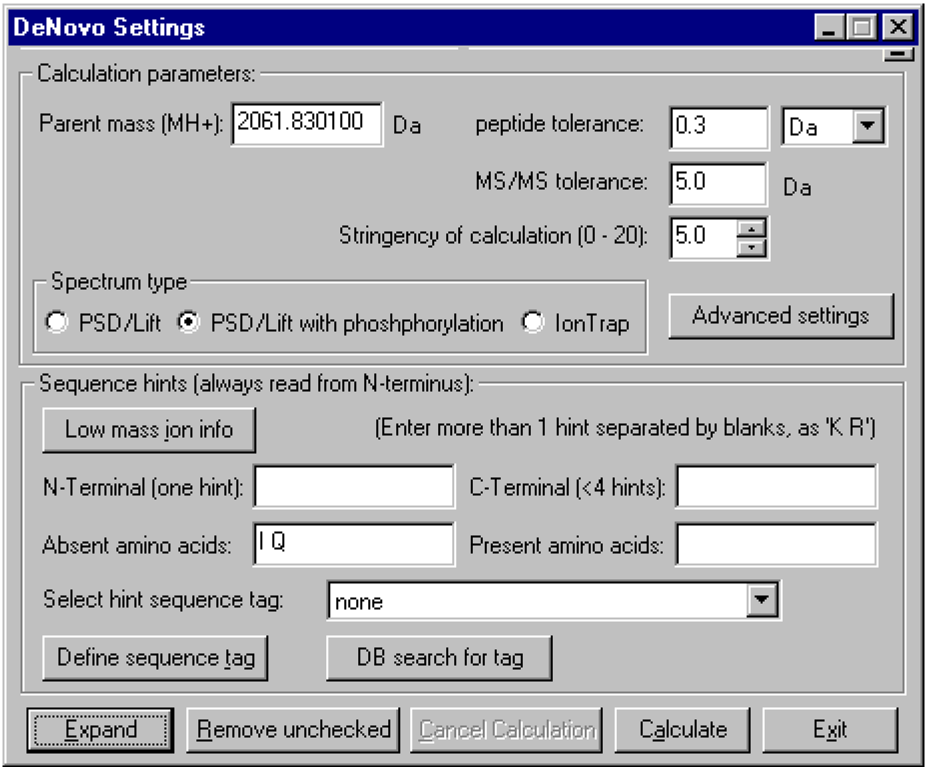


Figure 3-43, Full DeNovo settings smart

3.4.3.2. DeNovo settings details

The **Shrink/Expand** button hides/shows 4 list boxes for further selections.

- Multiple selections can be made for **Optional Modifications** in the expanded view (unmodified Amino Acids are also included in the sequencing) and **Fixed Modifications** (unmodified Amino Acids are excluded).
- One selection can be made from **Fixed N** or **C term modifications**.

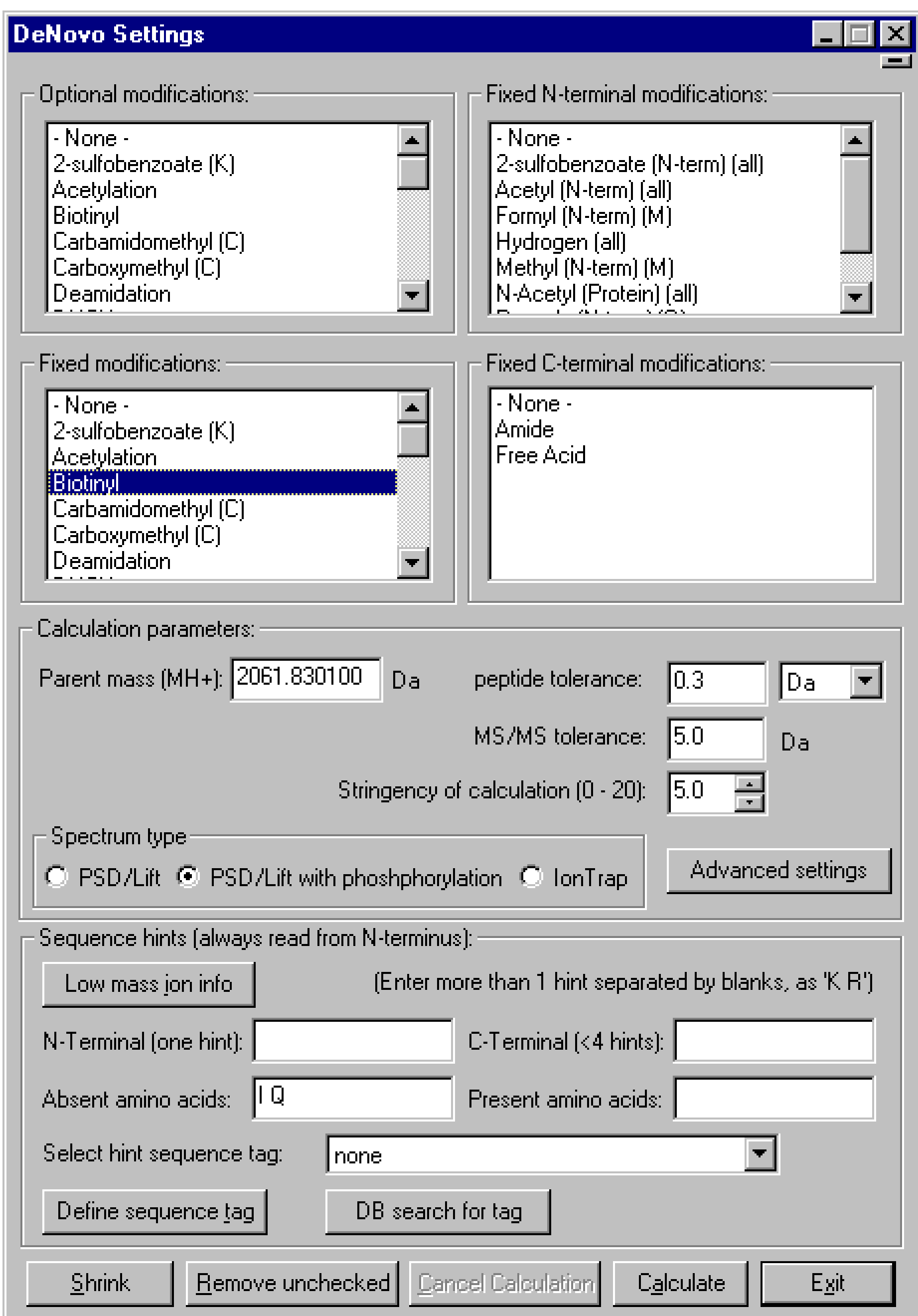


Figure 3-44, Full DeNovo settings details

3.4.3.3. DeNovo - advanced settings

The **Advanced Settings** contains the values for **Number of partial sequence candidates** (100 is good for most applications) on a separate dialog. This value is saved with the data.

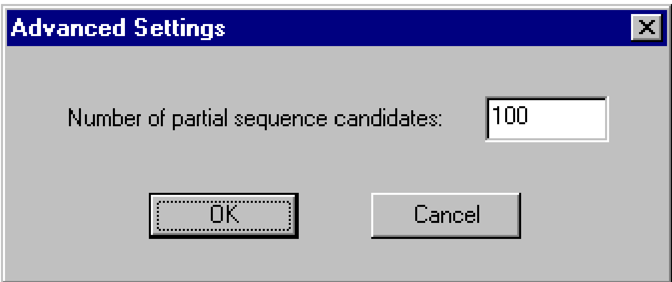


Figure 3-45, Full DeNovo advanced settings

3.4.3.4. DeNovo - define sequence tag

The **Define sequence tag** button creates a modal dialog for definition of internal hints.

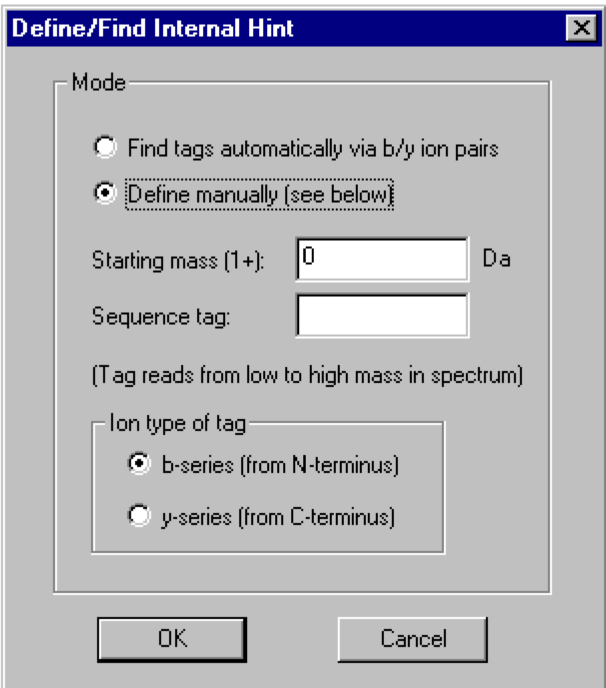


Figure 3-46, Full DeNovo – internal hints dialog

- The **Find tags automatically via b/y ion pairs** radio button option makes available the same functionality as the Create Tag button on the old DeNovo View (based on the availability of b-y ion pairs, the search is done up to the tolerance value specified by the MS/MS tolerance field on the settings dialog).
- The **Define manually** radio button option allows the user to enter start mass and tag for use in either b or y direction (specified with the **Ion type of tag** radio button (If y-direction is specified the tag and the start mass are then recalculated using the parent mass as values for the corresponding b-series.). The start mass and the tag are always read from left to right in the spectrum display. It is important that the start mass is accurately known (depending on the tolerances set). The start mass value should be MH+, i e the mass of the expected b or y ion up to the start of the internal hint tag specified.

3.4.3.5. DeNovo - DB search for tag

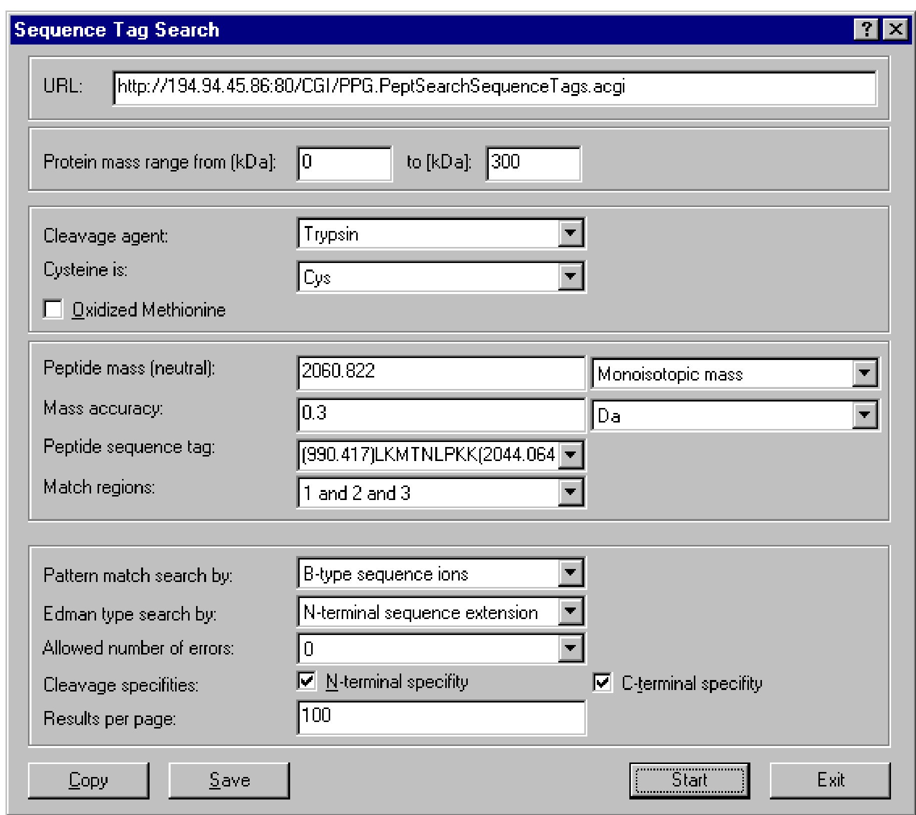


Figure 3-47, Full DeNovo DB search for tag

3.4.3.6. DeNovo - result analysis

The Result Box is now also modeless enabling the user to zoom into the spectrum and examine closely the annotation of the generated sequences. A sequence is activated by clicking on it in the bottom List Box.

While the Result Box is up, the DeNovo Settings Box is deactivated and minimized.

The occurrence of the fragments checked above in the sequences below is marked by a colored background for the acid residues concerned. The sequences are aligned according to the fragment selected with the highest score. The order of the resulting sequences is according to the BioTools score shown to the left of the sequence.

- The **Select tagged** button selects all full sequences that contain at least one of the fragments check marked in the List Control above.
- The **Define Internal Hint** button is used to create an internal hint for a subsequent sequencing step or Tag Database search. For the definition of a new internal hint exactly one fragment and exactly one sequence containing this fragment have to be selected, pressing the Define Internal Hint button then writes that hint to the left of the button. Upon leaving the dialog with the New Analysis button this hint is copied into the Combo Box of Internal Hints on the DeNovo Settings dialog and selected for the next sequencing step. If the selections are inappropriate upon clicking this button, a Message Box displaying a description of how to define a hint is displayed.
- The **Accept** button closes both the Result and Settings dialogs and copies the selected (check marked) sequences to the Tree View.
- The **New Analysis** button closes the Result dialog, reactivates the Settings dialog and copies the selected (check marked) sequences to the Tree View.
- The **Cancel** button closes both the Result and Settings dialogs.
- The **Copy/MS Blast** button copies the check marked sequences joined by a minus sign into the clipboard, the sequences are also formatted (Trypsin digest) for use with the EMBL MS Blast search
- The **Open MS Blast** button opens the web browser installed on the system and navigates to the current URL http://dove.embl-heidelberg.de/Blast2/msblast.html for EMBL MS Blast homology searches. (See that site for further information on the use of MS Blast for homology searches). This URL can be modified by including an entry in section [DeNovo] under ‘URL for MS Blast’ in the MSTool.ini file.

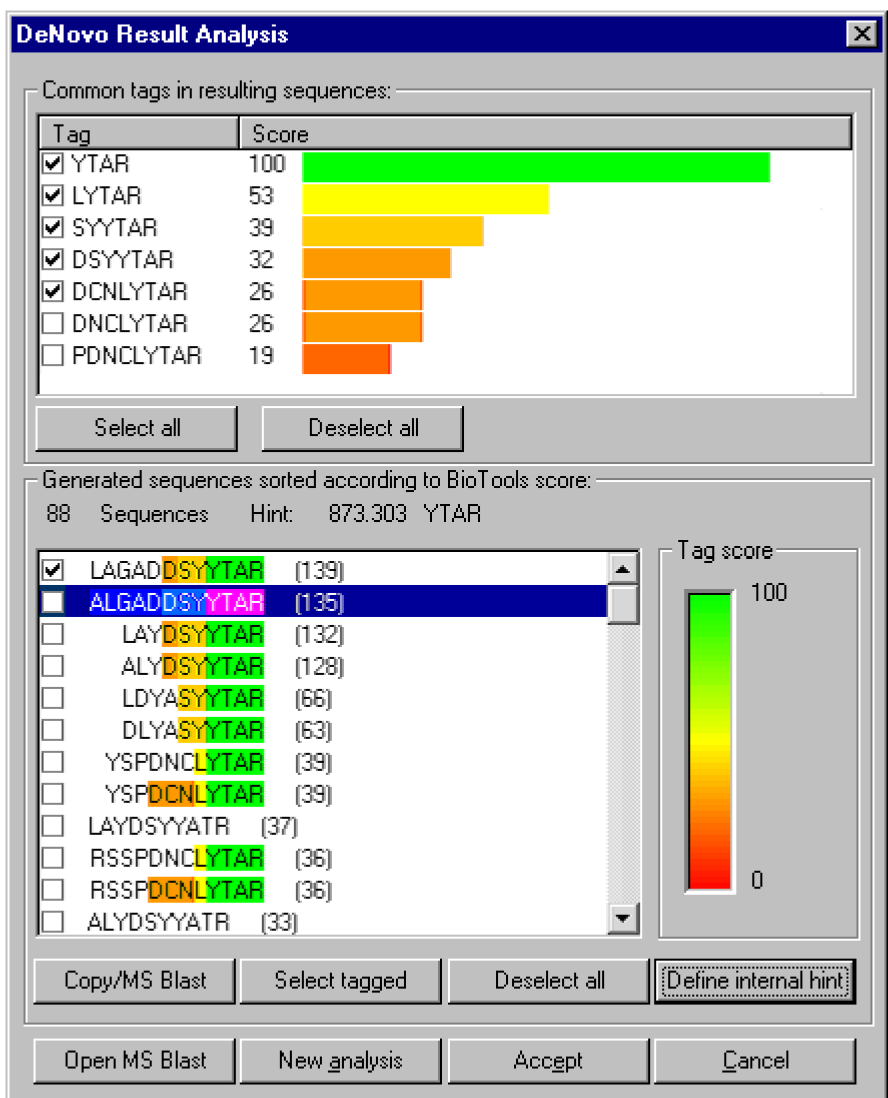


Figure 3-48, Full DeNovo result analysis

3.4.4. ISD Data

BioTools has to distinguish between different data types: TOF data to perform MS fingerprint searches and PSD data to perform MS/MS searches. ISD Data however are stored as TOF-data, but must be processed as PSD data. Enable this entry to perform PSD data processing for ISD data (e. g., for *DeNovo* sequencing).

3.4.5. Set Threshold

With this option an intensity threshold can be set for all visualizations and calculations, including the peak list restriction for MASCOT searches. After selection, the cursor in the spectrum window changes according to Figure 3-49 and can be set manually with the cursor.

Simultaneously, the threshold level is shown in the fragments window (Figure 3-50) and in the status bar. To terminate threshold mode click the right mouse button). The threshold can also be typed directly into the respective text field.

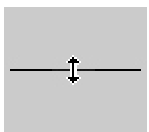


Figure 3-49, Threshold cursor

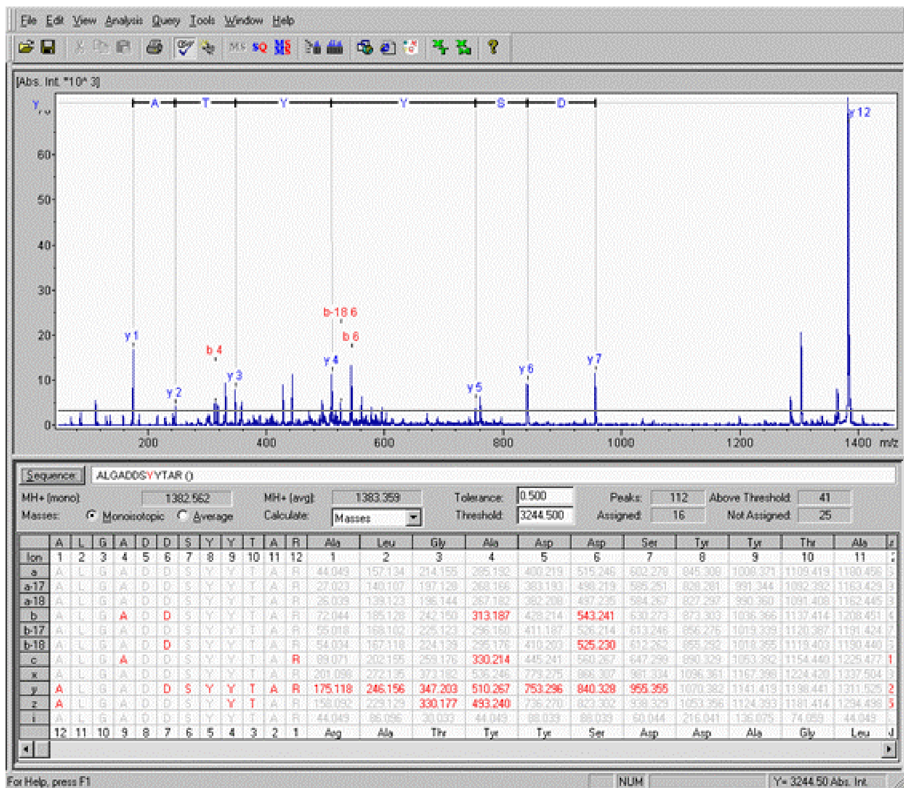


Figure 3-50, Threshold level in spectrum and fragments window and status bar

3.4.6. Annotation Parameter

Here, the annotation in the spectrum window can be modified, automated or manually set (Figure 3-51). This option is always set to automatic in conjunction with *DeNovo* sequencing.

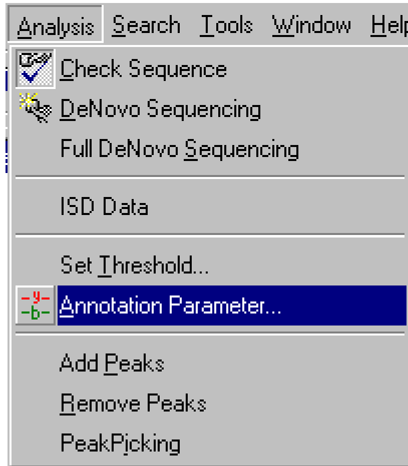


Figure 3-51, Annotation parameter in the analysis menu

With this parameter the view of the annotation objects shown in the spectrum window can be arranged. The annotation mode, the kind and the colors of the annotation for each ion type can be configured (Figure 3-52).

Ion list

The list of ion types considered for fragment ion calculations can be selected.

Annotate:

Choose for each ion type by clicking on it and activate the respective check box underneath label, mass or sequence or click the check box under the list window. To simplify the annotation, mark more ions by clicking on them keeping the Shift key or Ctrl key pressed and activate the respective check box near label, mass or sequence. If an annotation is not available it is grayed.

Label	The ion type of a peak will be shown (e.g., b5)
Mass	The m/z value of the ion will be shown
Sequence	Continuous sequence tags within the ion series will be shown in the upper part of the spectrum window

Colors:

Choose the color for each ion by clicking on it and change the color. Ignore improper color display in this box it will be shown in the right color in the spectrum.

Mode:

Choose the kind of annotation depending on the annotation speed.

No Annotation	No annotations are shown
Fast	High speed annotation object overlap possible
Optimized	Slower annotation, overlaps of objects are avoided

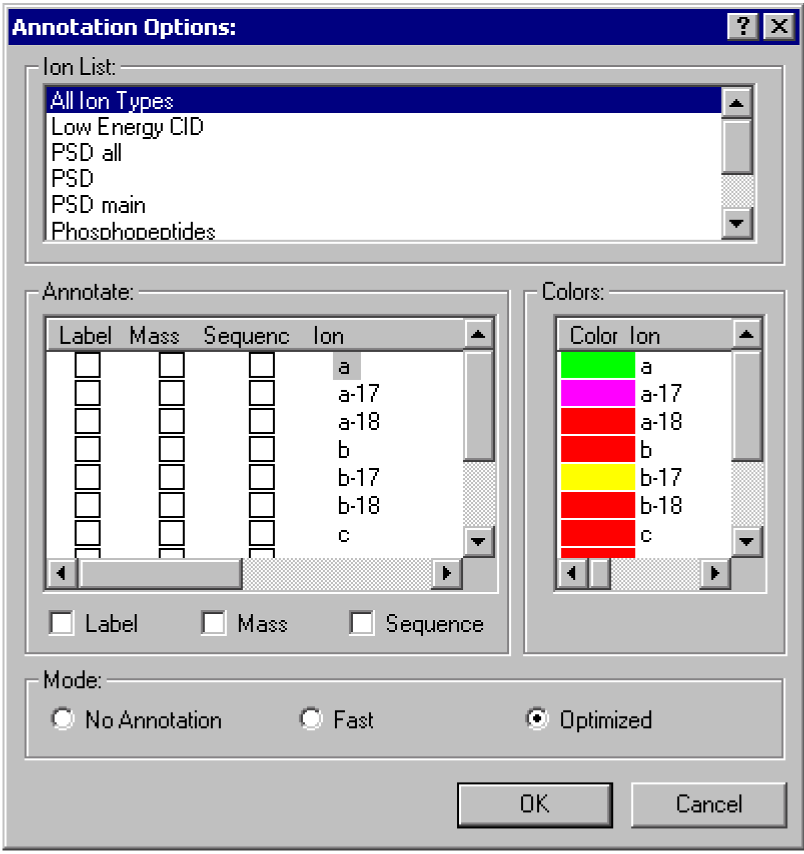


Figure 3-52, Annotation Options window

3.4.7. Add Peaks

Adds peaks to the existing peak list (**XMASS/ XTOF only**!). However, this should be done only for quick data evaluation, e.g., in SequenceEditor or MASCOT work. It allows only a simple maximum data point peak picking and does not update the peak list for Mascot batch mode searches! Peak picking should still be done with the data processing programs.

Click right mouse button in spectrum window to exit "Add Peaks" mode. Any zoom box operation in the spectrum window will also revert back to the zoom mode. Only zoom box operations in the overview spectrum will leave the "Add Peaks" mode activated.

3.4.8. Remove Peaks

Deletes peaks from the existing peak list (**XMASS/ XTOF only**!). However, this should be done only for quick data evaluation, e.g., in SequenceEditor or MASCOT work. It does not update the peak list for Mascot batch mode searches! Peak picking should still be done with the data processing programs.

Click right mouse button in spectrum window to exit "Remove Peaks" mode. Any zoom box operation in the spectrum window will also revert back to the zoom mode. Only zoom box operations in the overview spectrum will leave the "Remove Peaks" mode activated.

3.4.9. Peak Picking

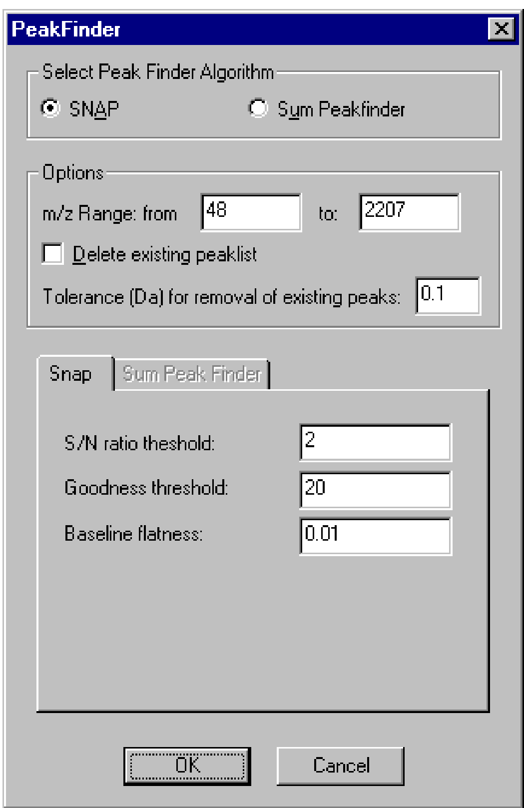
For TOF data BioTools can perform peak picking with the SNAP or Sum Peak Finder algorithms available from the Numerical Toolbox. This feature is available for suitable data sets from **Analysis – Peak Picking** or from the context menu in the spectrum via the right mouse button.

With this feature the annotation of selected ranges of a spectrum can be easily improved by picking additional peaks using less selective parameters.

- The '**Select peak finder algorithm**' distinguishes between the SNAP (most sophisticated, takes isotope patterns into account) and Sum peak finder (very fast) algorithms.
- The **Options** include the **m/z range** where peaks should be searched (this is set to the current display range of the spectrum by default), a check box ('**Delete existing peaklist**') to toggle the replacement of existing peaks (by default, peaks are always added to the existing peak list), if a new peak is found to be closer than the value of **Minimal mass difference between peaks** (default 0.1 Da) it is not added to the peak list to avoid multiple peaks with the same mass.
- SNAP or Sum parameters.

3.4.9.1. SNAP

The **SNAP** tab contains three parameters:



SNAP Peak finder dialog

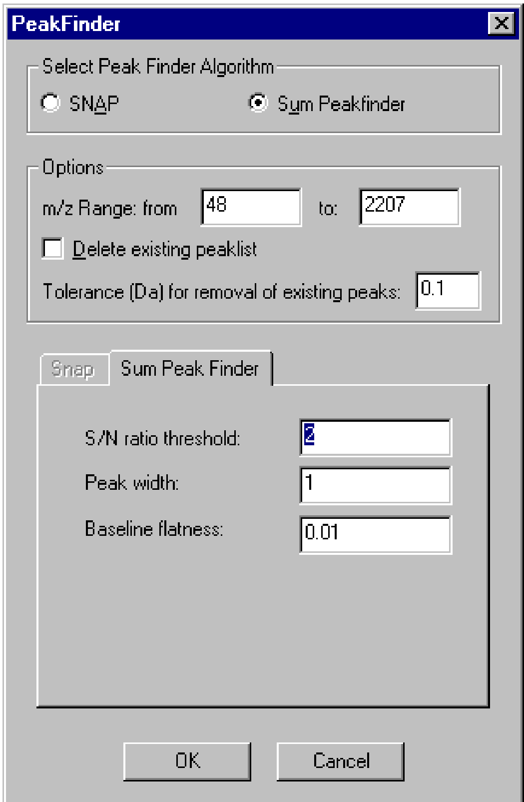
S/N ratio signal-to noise ratio, NTB script: Threshold for the intensity. This value is also used in other ways to select efficiently peaks worth to be analyzed. Values between 2.0 and 4.0 are reasonable for most cases. A lower value also means longer calculation times as more peaks have to be analyzed. If you have a large Goodness-Threshold it doesn't make sense to use a very small ThreshSN value. The noise is defined locally. The larger the flatness of the baseline the more segments will be used for the baseline and noise estimation. This helps to model variable noise levels as long as the peak density isn't too high.

Goodness threshold A higher value makes the method more selective. NTB script: The goodness is positive, but isn't limited. The useful threshold depends on the kind of spectrum and the level of quality you expect for the patterns. A level of 20.0 might be a reasonable starting point but you should definitely test that on your own for your specific analytical problem!

Baseline flatness NTB script: Flatness of a spectrum is used for baseline calculations. Here the whole spectrum needs to be separated into segments whose number is allowed to vary from 1 to 100. Flatness is a measure of how strong the baseline varies. Flatness of 0.0 means that the baseline does not vary at all or only to a very small degree. In this case just the whole spectrum is considered as the only segment. For a flatness value of 1.0, the spectrum is separated into 100 segments.

3.4.9.2. Sum Peak Finder

The **Sum Peak Finder** tab has the three parameters: S/N ratio (signal-to noise ratio), **Peak width** (NTB script: This should be approximately the peak width. It is used to set the number of points to be used to calculate the Savitzky Golay polynomials. If it isn't set or it is = 0.0 an internal estimation is done once for the whole spectrum, which might be preferable if you really don't have a reasonable guess.) and baseline flatness.



Sum Peak finder dialog

3.5 Search Menu

With the *Search Menu* the search function for protein identification can be selected. Use this entry to start a query on the Internet or on your local Intranet (e. g., if you have installed the Mascot system on a local machine). The results of the query will be displayed in the built-in web browser. Before you can use this option make sure that you have a working Internet connection.

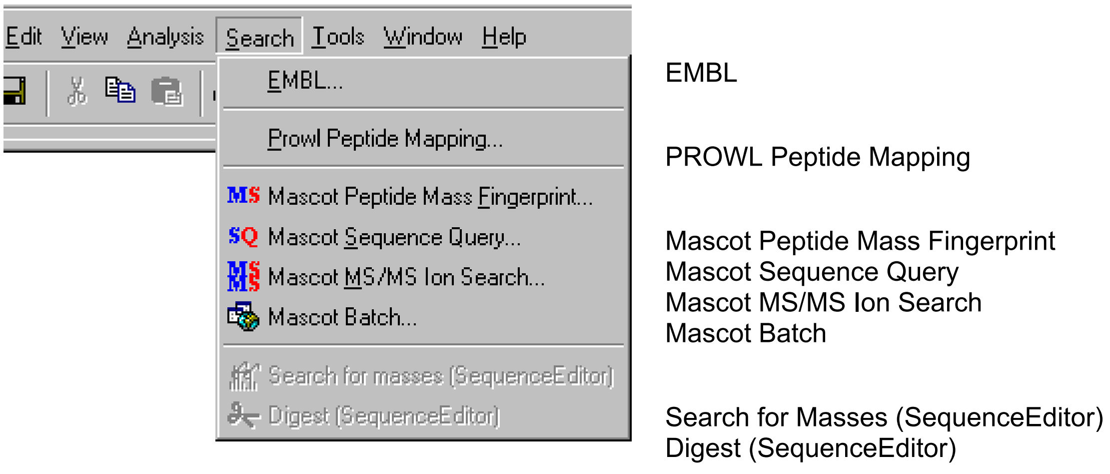


Figure 3-43, Search Menu

The program shows the input masks of EMBL, Prowl and Mascot and fills the input forms with useful default values and parameters taken from the MS – spectrum (e. g., peak list and parent mass). A description of the meaning of the fields of the input mask is given in the tutorial for Sequence Database Searches from MALDI Peptide Mass Fingerprints (see chapter F.2.1., Define the Search). For details of the meaning of these fields see the information given at the corresponding Internet addresses (see below).

BioTools submits the input data directly to the Internet address shown after "URL:" (e. g., in case of Mascot database search to:

http://194.42.244.117/cgi/nph-mascot.exe?).

However from time to time this address or even the input mask may change.

Figure 3-44, URL for Mascot database search

If the address (in this case 194.42.244.117) has changed, enter the new address and store it using the "Save" or "Save Parameter" button. If you do not know the address or other required parts of the address use your Internet Browser (Microsoft Internet Explorer or Netscape Navigator) and go to the homepage, e.g.

http://www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html .

Next follow the link to the input mask for the search using a list of masses.

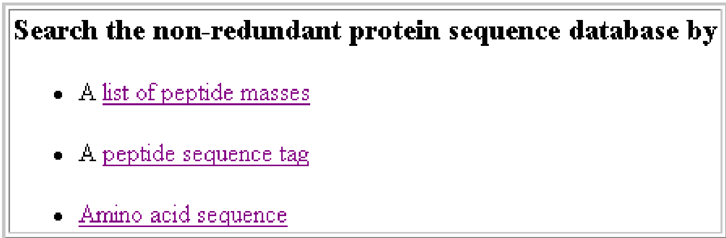


Figure 3-45, Link to Search using a list of peptide masses

When the input mask is displayed in your browser window, use the right mouse button within the input form. From the popup menu select entry "View Source". Using the Microsoft Internet Explorer open a text editor (e. g. Notepad).

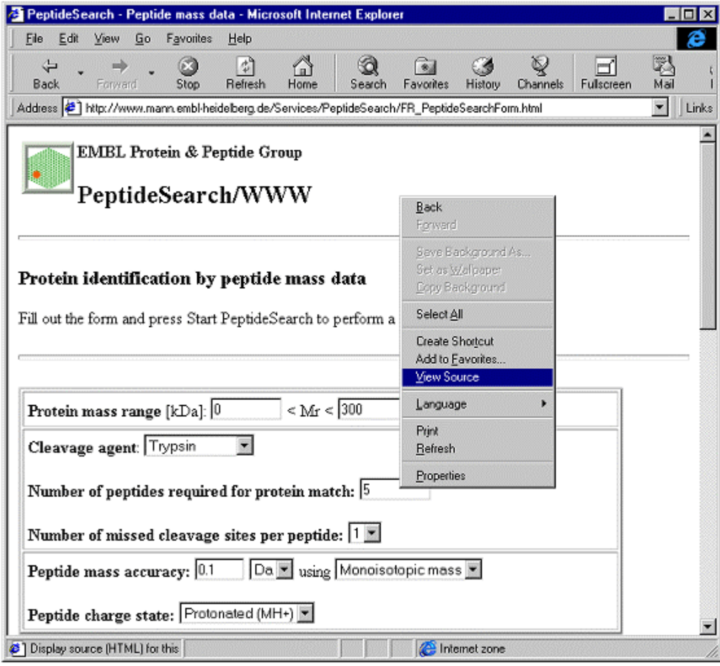


Figure 3-46, Popup menu to display the source of a HTML page

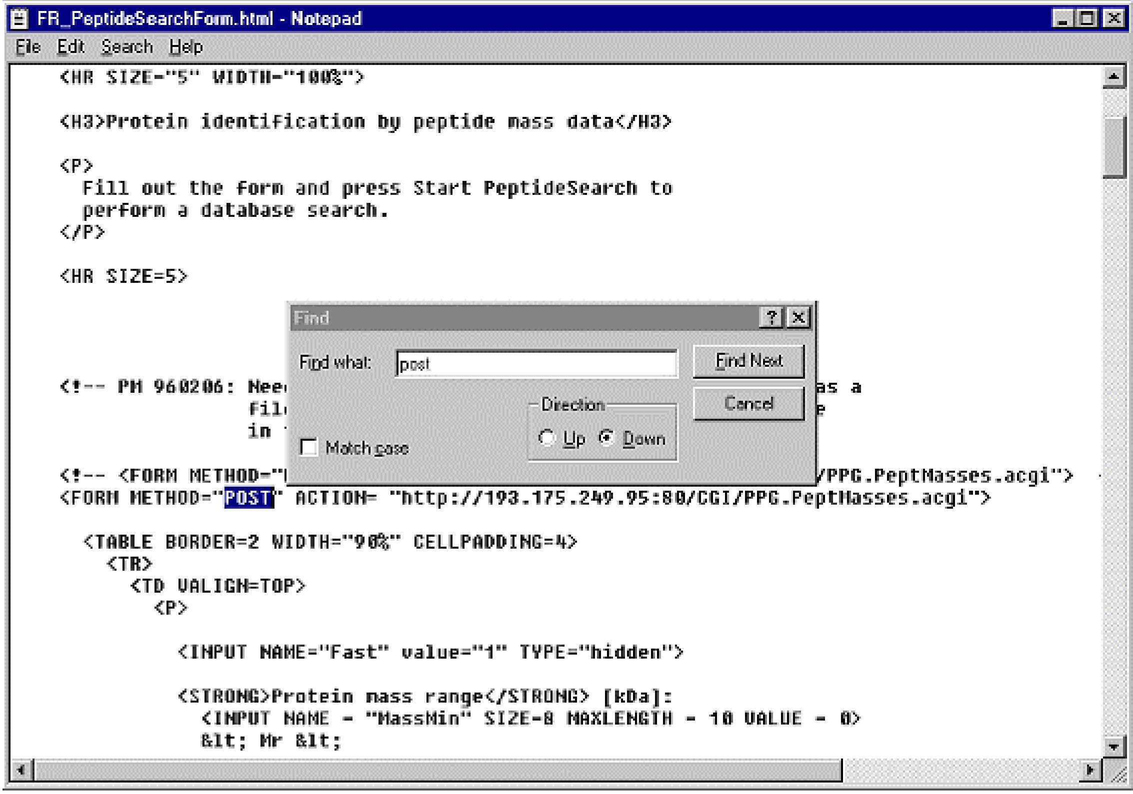


Figure 3-47, Source code of an HTML page

Use the "Find" option from the "Edit" menu and search for "post". There you will find an address or parts of it that is used for the submission of your input data. Enter this address in the input mask of the BioTools window and save the data (see above).

3.5.1. EMBL

For a description of Protein identification by peptide mapping or peptide sequencing check the following address:

http://www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html

The original input form can found at:
http://www.mann.embl-heidelberg.de/Services/PeptideSearch/FR_PeptideSearchForm.html

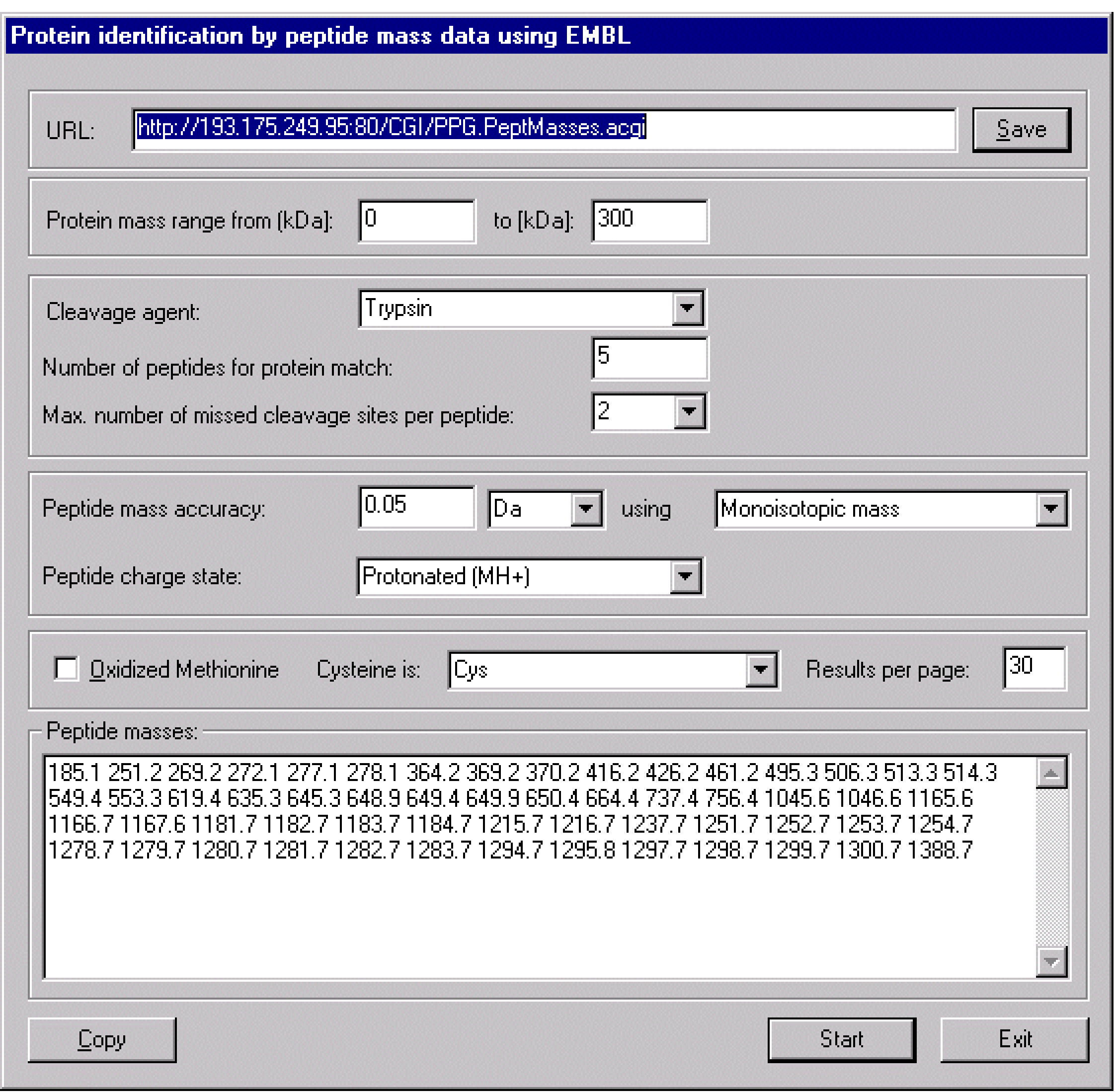


Figure 3-48, EMBL window

3.5.2. PROWL Peptide Mapping

For a description and an example of Protein identification by comparison of a peptide map check the following address:

http://prowl.rockefeller.edu/cgi-bin/ProFound?INTRO

The original input form can found at:

http://prowl.rockefeller.edu/cgi-bin/ProFound

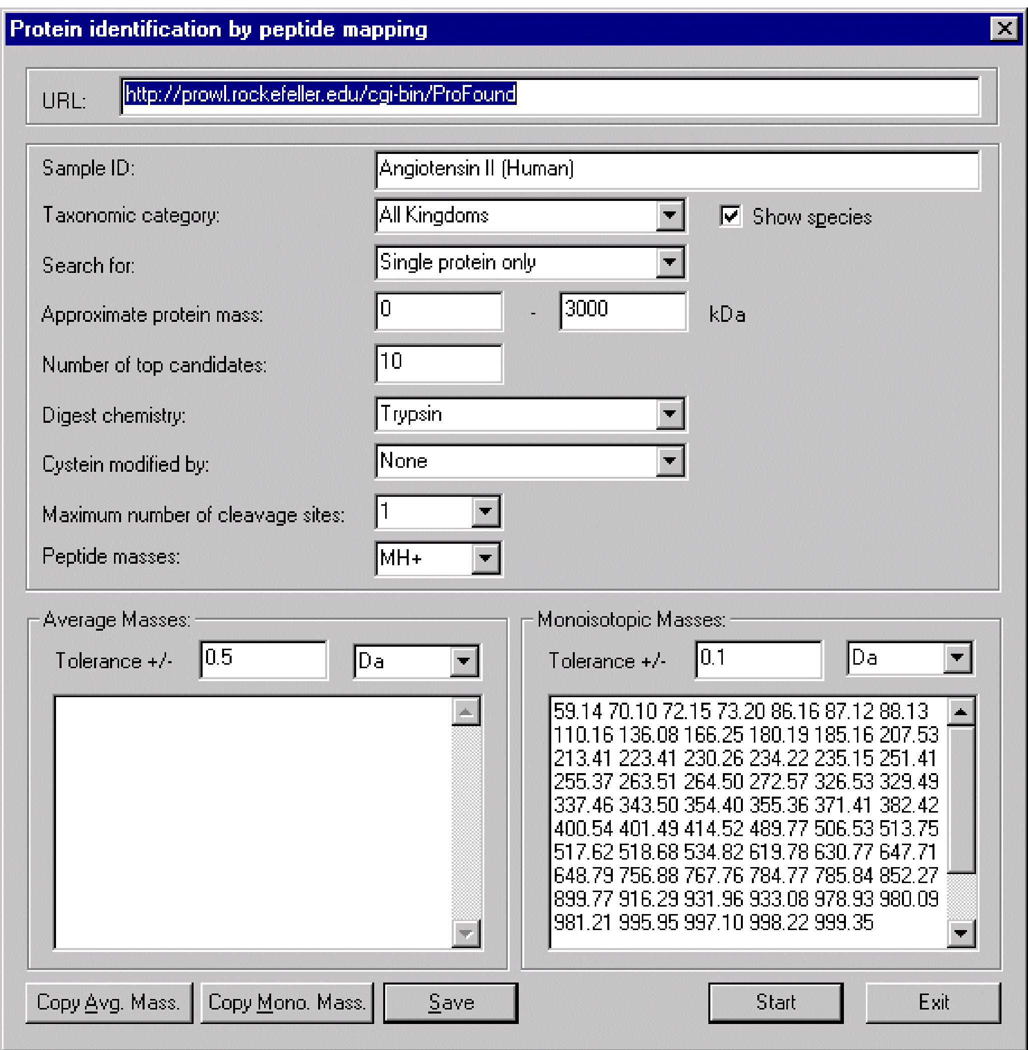


Figure 3-49, PROWL Peptide Mapping window

3.5.3. Mascot

For a detailed description and examples of Protein identification check the following address:

http://www.matrixscience.com/

The Mascot software is also available for an Intranet running on different platforms. Contact Bruker Daltonik for further information.

http://www.bdal.de/

3.5.3.1. Mascot Peptide Mass Fingerprint

The original input form can be found at:

http://www.matrixscience.com/cgi/search_form.pl?SEARCH=PMF

Typically, **fixed modifications** include the known chemistry, such as reductive disulfide cleavage and carbamidomethylation.

Unknown chemistry, such as the artifact Methionine oxidation or phosphorylation can be specified as **variable modifications**.

Important is the **peptide tolerance** (or mass error) which can be a major source of frustration due to failed identifications! So: be sure about your data quality.

Typically, the precursor **protein mass** does not need to be specified.

The number of **missed cleavages** (or partials) accounts for tolerated internal cleavage sites in matching peptides. This number should be set to 0 or 1, since higher values reduce the specificity of the search as extensive use of **variable modification** does. If higher values seem to be required on a routine basis, you have to optimize your digest for more complete proteolysis. (In silver gels, destaining might help!).

After you set up the search parameters to a type of application you need to process frequently, hit the **Save Parameter** button to store conditions.

Hit the **Start** button for a MASCOT search, or **Copy mass list** to paste the mass list into the clipboard.

Note ***From the clipboard you can paste them into any browser-based search engine on the web, such as PeptideSearch, PepSea, Profound or MS-Fit. The search results from these programs, however, cannot be imported back into BioTools, in contrast to MASCOT.***

Peptide Mass Fingerprint

URL: http://www.matrixscience.com/cgi/nph-mascot.exe?1 | Add | Del
Matrix Science home page | Edit URL

User Name: | Email:
Search title:

Taxonomy: All entries
Database: OWL | Enzyme: Trypsin
Fixed Modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)
Variable Modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)
Protein mass: kDa | Max. no. of missing cleavages: 1
Peptide tol. ±: 300 ppm

Mass values: MH+ | Mr | Monoisotopic | Average

Peaklist:
820.484656 5328.000000
906.493027 3589.000000
922.478036 1389.000000
974.448786 1840.000000
987.557706 3749.000000
1014.591461 371.000000

Results: Overview | Report top 30 hits

Copy Peaklist | Copy Masslist | Save Parameter | Start | Exit

Figure 3-50, Peptide mass fingerprint window

3.5.3.2. Mascot Sequence Query

The original input form can be found at:

http://www.matrixscience.com/cgi/search_form.pl?SEARCH=SQ

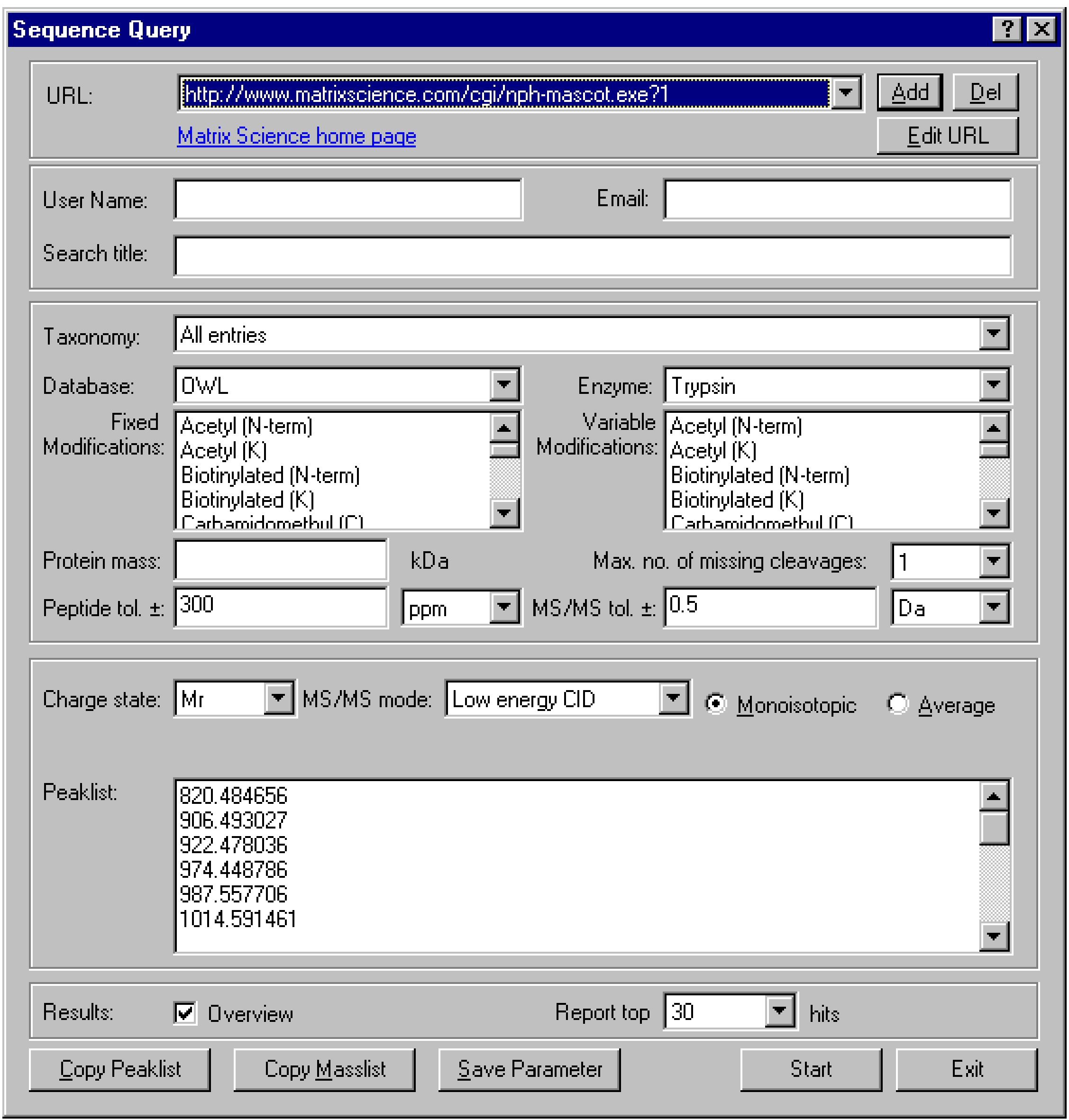


Figure 3-51, Sequence Query window

3.5.3.3. Mascot MS/MS Ion Search

The original input form can be found at:

http://www.matrixscience.com/cgi/search_form.pl?SEARCH=MIS

Typically, **fixed modifications** include the known chemistry, such as reductive disulfide cleavage and carbamidomethylation.

Unknown chemistry, such as the artifact Methionine oxidation or phosphorylation can be specified as **variable modifications**.

Important is the **peptide tolerance** (or mass error) which can be a major source of frustration due to failed identifications! So, be sure about your data quality.

The **MS/MS tolerance** is equally important, typical settings for PSD are in the range of 0.5-1.5 Da – still enabling successful searches.

Typically, the precursor **protein mass** does not need to be specified.

The number of **missed cleavages** (or partials) accounts for tolerated internal cleavage sites in matching peptides. This number should be set to 0 or 1, since higher values reduce the specificity of the search as extensive use of **variable modification** does. If higher values seem to be required on a routine basis, you have to optimize your digest for more complete proteolysis. (In silver gels, destaining might help!).

Charge state for the assumed peptide peaks in the spectrum: 1+ for PSD.

M/z of the precursor ion, of which the fragment ions are used for the search. This parent mass is automatically transferred into this field in case of MALDI-TOF spectra. All others need to manually specify m/z value and proper charge state.

MS/MS mode setting accounts for your instrument method: PSD for MALDI-TOF and low energy CID for esquireSeries are typical settings.

After you set up the search parameters to a type of application you need to process frequently, hit the **Save Parameter** button to store conditions.

Hit the **Start** button for a MASCOT search, or **Copy mass list** to paste the mass list into the clipboard.

MS/MS Ions Search

URL: http://www.matrixscience.com/cgi/nph-mascot.exe?1
Add
Del
Matrix Science home page
Edit URL

User Name:
Email:
Search title:

Taxonomy: All entries
Database: OWL
Enzyme: Trypsin
Fixed Modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)
Variable Modifications: Acetyl (N-term), Acetyl (K), Biotinylated (N-term), Biotinylated (K), Carbamidomethyl (C)
Protein mass: kDa
Max. no. of missing cleavages: 1
Peptide tol. ±: 300 ppm
MS/MS tol. ±: 0.5 Da

Charge state: 1+
MS/MS mode: Low energy CID
Monoisotopic
Average
m/z: 2093.090100
Peaklist:
56.000000 120.148346
60.100000 293.979997
70.100000 1694.858589
84.100000 1607.942764
87.100000 242.853041
101.100000 115.035651

Results: Overview
Report top 30 hits

Copy Peaklist
Copy Masslist
Save Parameter
Start
Exit

Figure 3-52, MS/MS ions search window

3.5.3.4. Mascot Batch Searches

In BioTools there is a Mascot Batch Mode dialog where you can to set up a task, i. e., a batch of spectra combined with search methods. The system performs a batch database search and judges the results of all searches.

The Batch window contains five tabs.

The Status tab shows the current status of a task with all search results – it is a real-time search monitor. From here database searches can be started for all spectra of a task or for selected spectra.

The Scout MTP tab shows a scheme of the SCOUT target with color-coded search results; it is accessible after the batch process is completed.

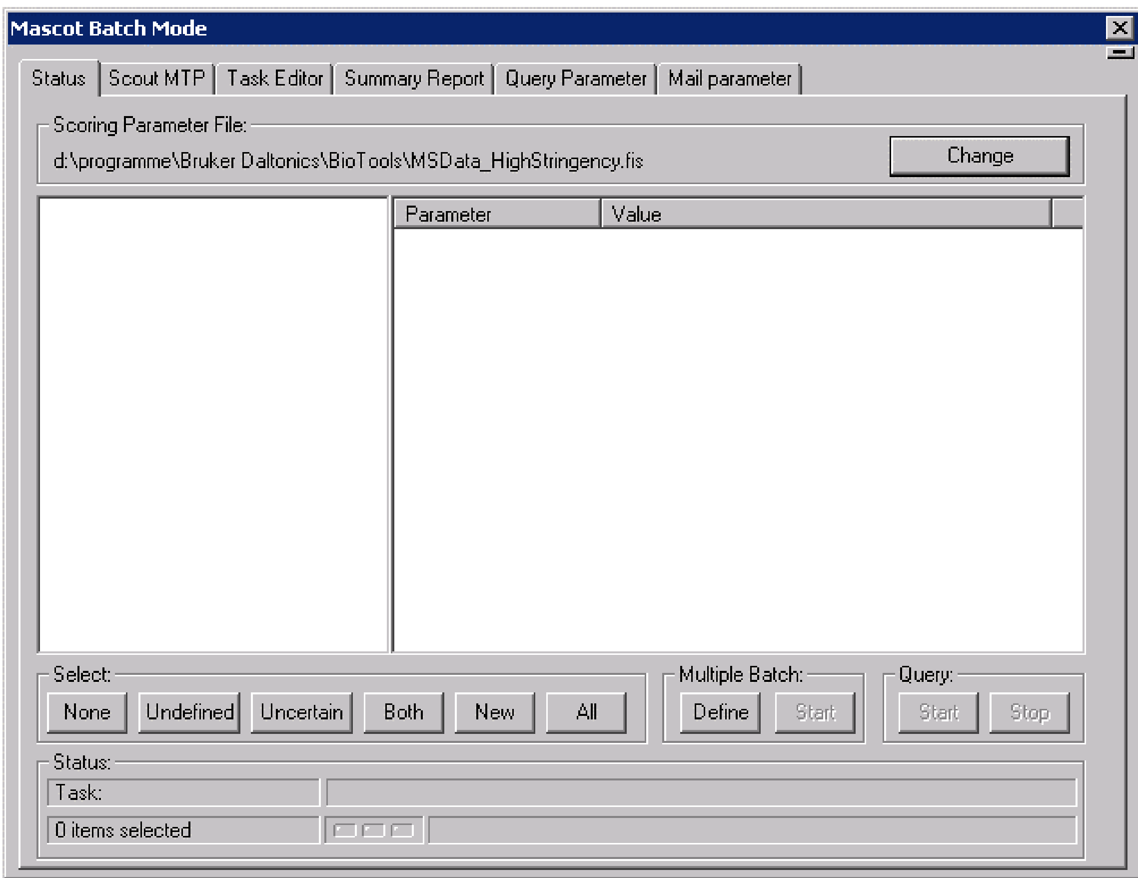


Figure 3-53, Mascot Batch Mode Window

The Task Editor tab is used to set up a new task by adding spectra to the list or by importing a spreadsheet "from AutoX". This is also the place to combine every spectrum with a BioTools search method.

The export functions of BioTools can be set in the Summary Report tab. From here it is possible to print out the result of a batch search as a table or to export the results for further use by Excel.

The Query Parameter tab allows access to BioTools search methods. Existing methods can be edited and new methods can be defined.

The results can be sent to an email recipient, chosen in the dropdown list. This mail recipients list can be edited on the Mail Parameter tab via the button Edit Recipients.

3.5.3.4.1. Tab Status

The status window displays information about the current state of a Mascot batch search. The treeview on the left shows all defined entries of the task. The colored circles represent the result of a query (for details see chapter "Scout MTP Window"). A single mouse click on one entry shows information about the query and its results on the right side of the window.

A double click on one entry on the left opens the Mascot result in the browser window and loads the corresponding spectrum (if available).

Scoring Parameter File:

Change BioTools judges the results of a query with different rule sets for different mass spectra in a "Fuzzy Logic" algorithm . Currently available are rules for MS fingerprints and esquire LC-MS/MS runs. The different rule sets or "fuzzy engines" are selectable from a series of **fis** files, which vary by their judgment stringency. After a batch search is finished you can select different fis files to find out which judges the data in a way you can agree mostly (the judgment is updated after each selection).

Select:

None All entries from the task list will be de-selected.

Undefined Selects all entries from the task list with query state "Undefined".

Uncertain Selects all entries from the task list with query state "Uncertain".

Both Selects all entries from the task list with states "Undefined" or "Uncertain".

New Selects all new entries from the task list.

All Selects all entries from the task list.

Multiple Batch:

Define With a Multiple Batch you can add several task files to fulfill them one after the other automatically.

Start After completion the Multiple Batch run is started.

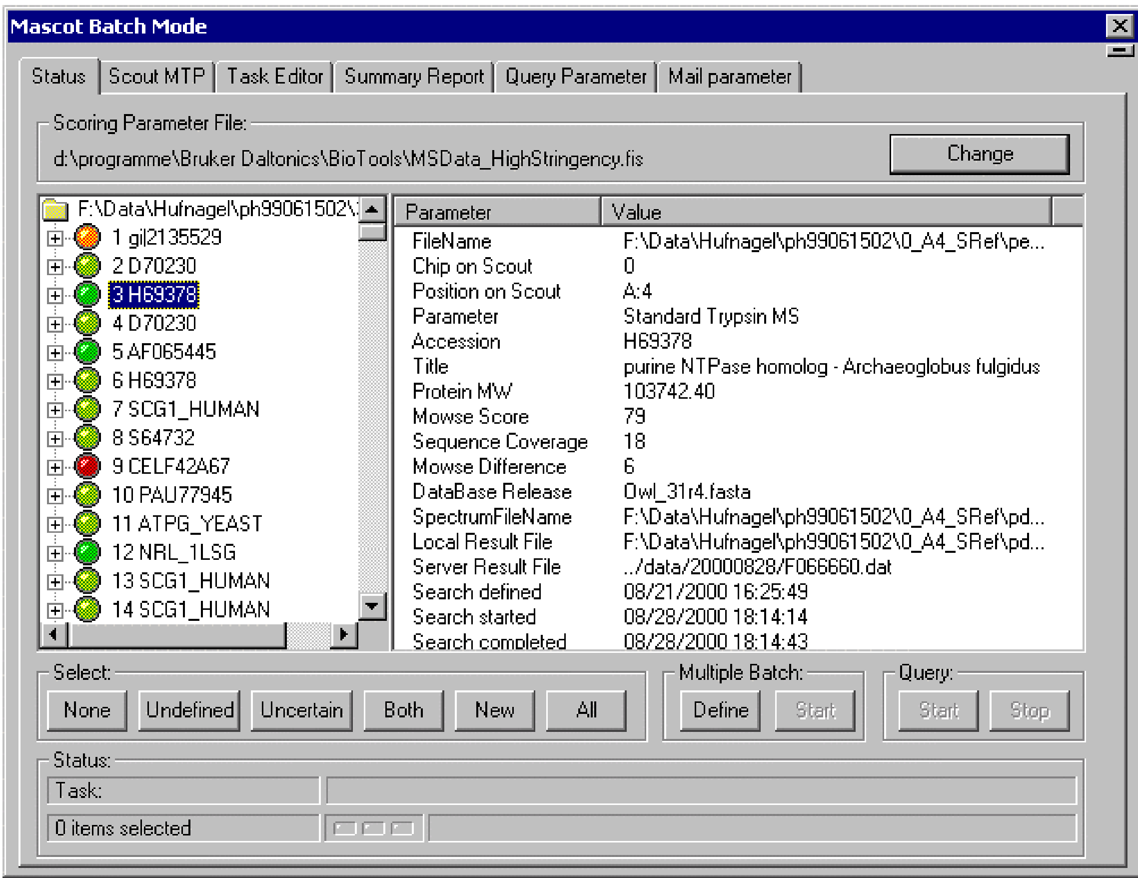


Figure 3-54, Status tab

Query:

Start Starts the query and disables all other buttons. You cannot leave or close this window while a query is running. Use the "Stop" button to interrupt a query. However, using a double click on the colored circles in the tree view on the left side you can access the Mascot results page and open the spectrum for detailed investigation, while the batch process is going on. From the results window you can import the sequence information and peak data with the *Get Hit(s)* button.

Stop After completion of the current search the batch run will stop.

Multiple Batch

Use the Multiple Batch dialog to add more task files to the batch. This is useful for easy handling of monotone and repeating steps.

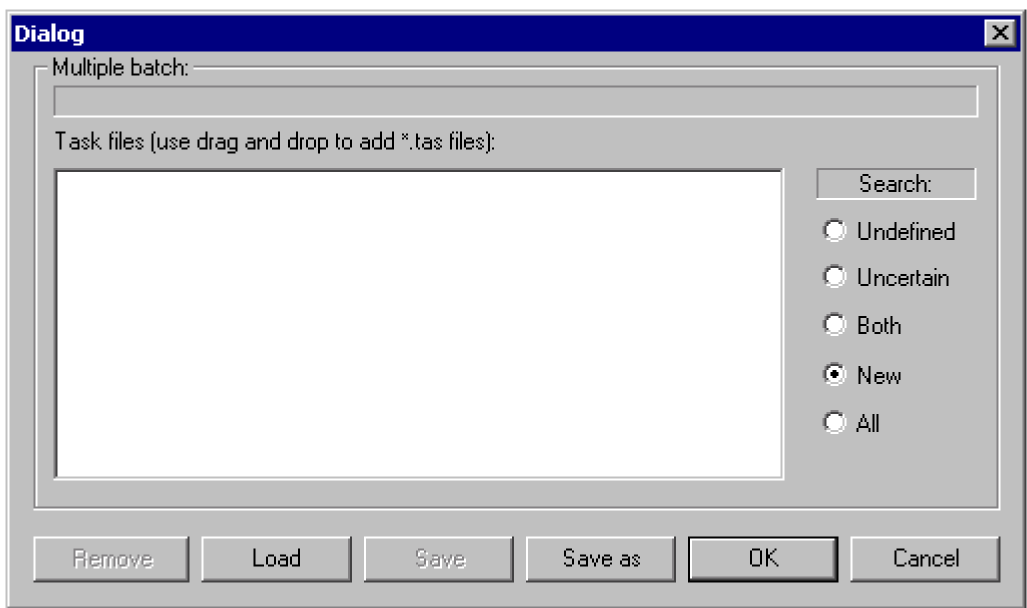


Figure 3-55, Multiple Batch

3.5.3.4.2. Tab Scout MTP

This window displays information about the state of a Mascot batch search based on data of an AutoX run (**FLEX-systems only!).** The positions shown represent the positions on the microtiter plate (MTP) shaped MALDI target (SCOUT MTP ion source only).

The colored circles represent the result of a query. Green means "identified". A pale green indicates that the result is not unique. This happens if a Mowse score (see Mascot manual for details) is very high, but the difference between the best hit and the following is very small. A yellow circle indicates an uncertain, but unique result. Orange circles represent multiple uncertain results. Red circles indicate undefined results. All uncertain and undefined results require further investigation.

Moving the mouse above the colored circles will update the information about Mowse score (calculated by Mascot), sequence coverage and score difference. A click on a spot opens the result window and loads the corresponding spectrum (if available).

The mouse position and the information of the spot will be displayed in the wintips which pops up when the mouse stays over a spot.

Scoring Parameter File:

Change

BioTools judges the results of a query with different rule sets for different mass spectra in a "Fuzzy Logic" algorithm . Currently available are rules for MS fingerprints and esquire LC-MS/MS runs. The different rule sets or "fuzzy engines" are selectable from a series of **fis** files, which vary by their judgement stringency. After a batch search is finished you can select different fis files to find out which judges the data in a way you can agree mostly (the judgement is updated after each selection).

Target:

1 - 24

A - P

Opens the result window (Figure 3-57) and loads the corresponding spectrum (if available). The position can also be changed by click on the respective spectrum probe with the left mouse button.

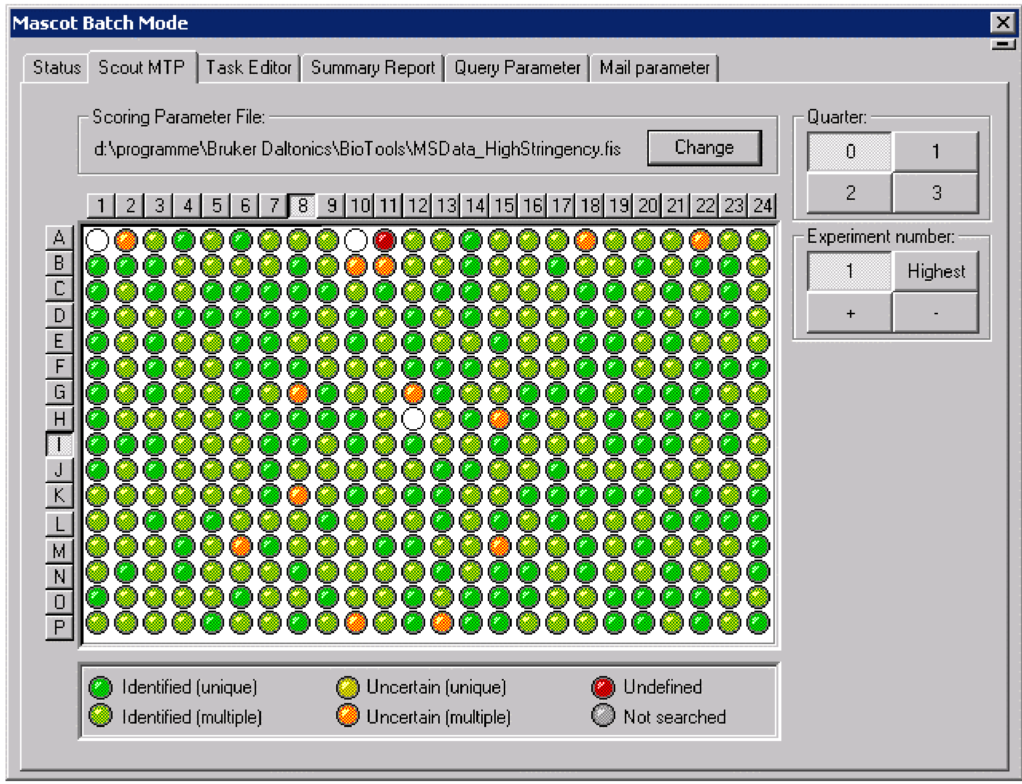


Figure 3-56, Scout MTP tab

Quarter: 0 1 2 3

Display the selected quadrant of a 1536 sample SCOUT MTP target.

0 corresponds to the 384 sample geometry.

Experiment Number: 1 highest + -

Display the current experiment number.

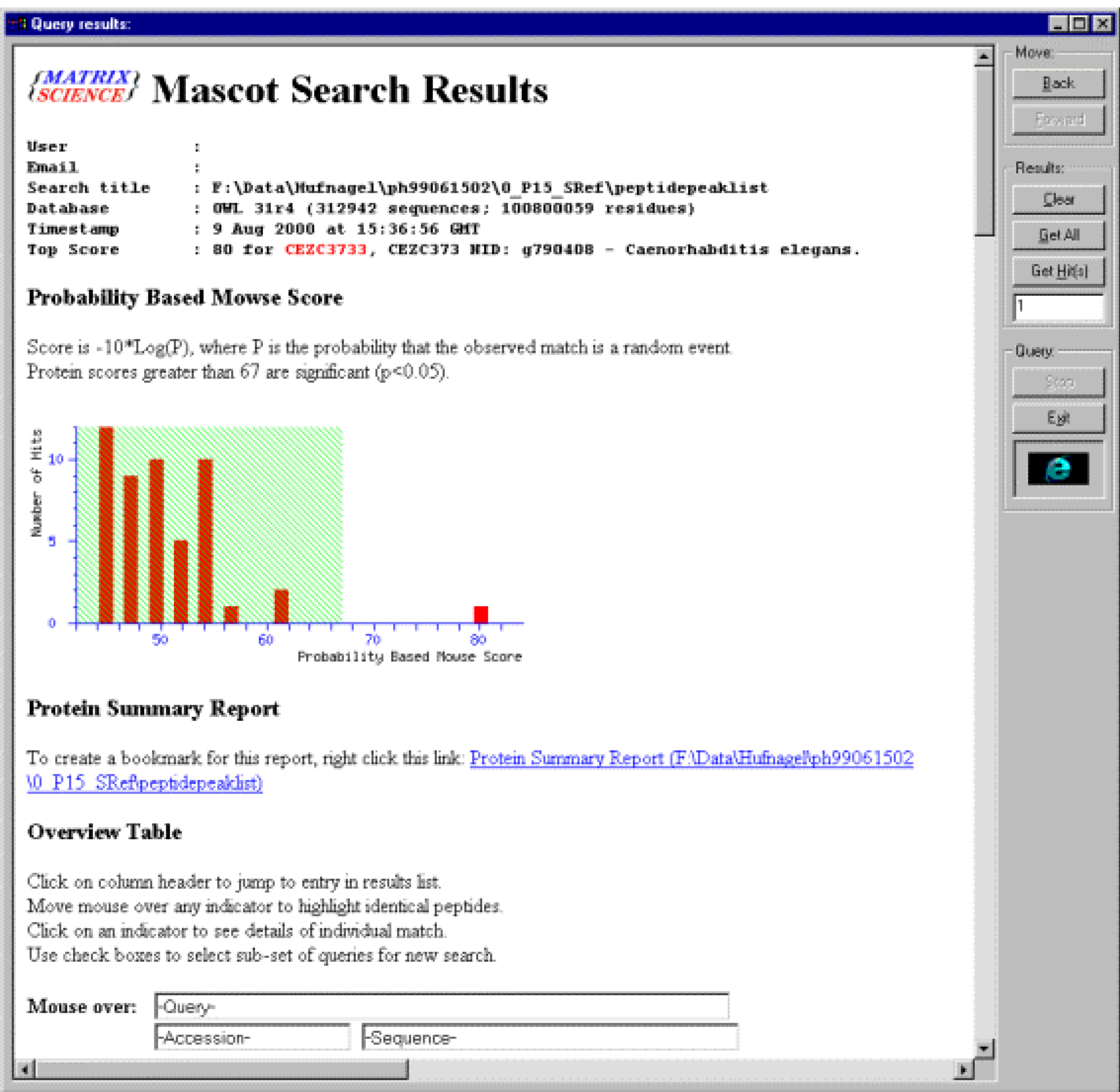


Figure 3-57, Mascot search results

3.5.3.4.3. Tab Task Editor

Use this window to load existing tasks or to create new tasks.

Task:

New — Clear current task list.

Load — Load an existing task (*.tas file).

Save — Save current task.

Save As — Save current task under a different name.

Parameter — Shows the parameters of the selected spectrum.

Apply Method — First select the input data in the left window, then select a search method from the right window and finally click the button Apply Method. This defines what search method is going to be used for each selected entry.

Apply recipient — First select the input data in the left window, then select a mail recipient from the right window and finally click the button Apply recipient. Applies the email recipients to the selceted spectra.

Edit recipients — Opens a new Edit data dialog box to make changes to the desired data, here email addresses.

Add/remove Files:

From AutoX — Opens a new dialog Data from AutoX.

Remove — Mark unwanted entries from your input data to remove them.

Add — Add data to task list. This button shows a file open dialog. Supported data types are "*.mgf" (Mascot generic files – see Mascot manual for details) from DataAnalysis 2.0 and "peptidepeaklist" from AutoX/XTOF macros.

Select Files of Status:

Undefined — Selects all entries from the task list with query state "Undefined".

Uncertain — Selects all entries from the task list with query state "Uncertain".

New — Selects all new entries from the task list.

All — Selects all entries from the task list.

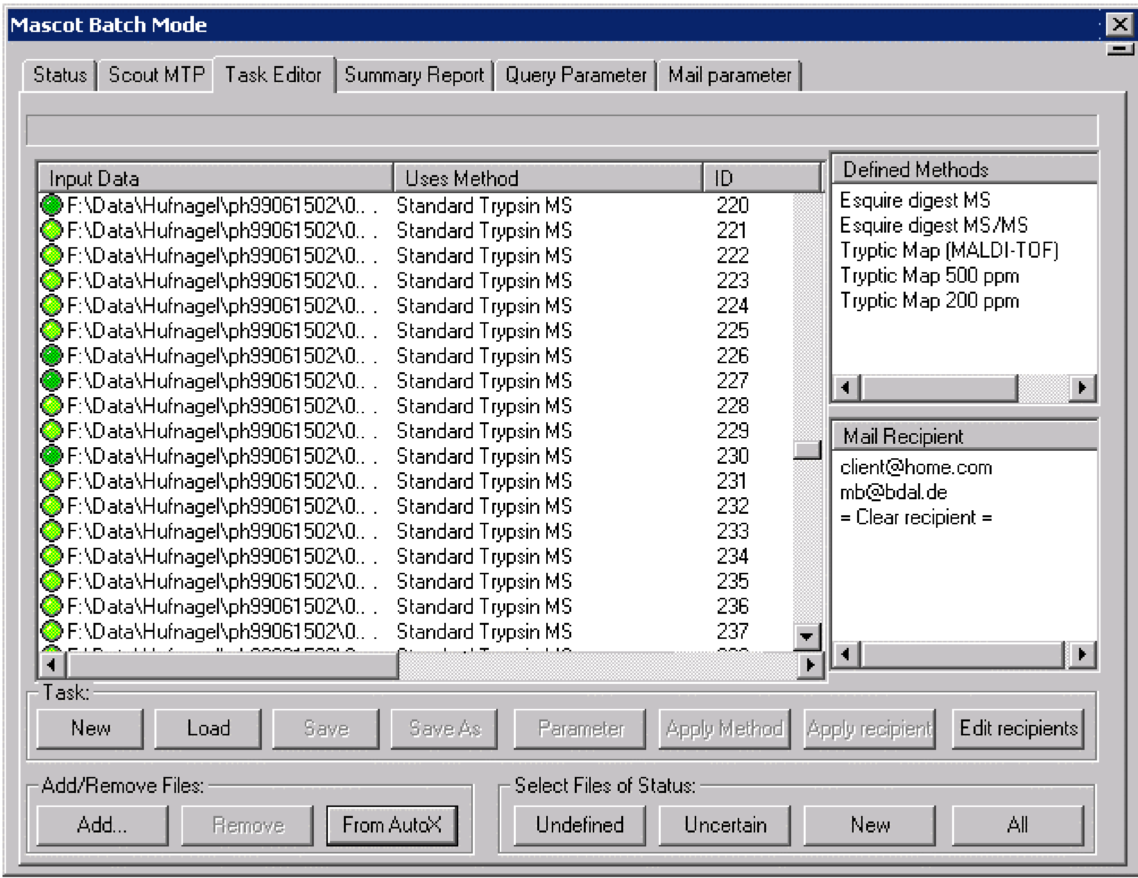


Figure 3-58, Task Editor tab

There is another way to add data to the task list: open the Microsoft Explorer and use "*Tools – Find – Files or Folders…*" and enter the parameter for "Named:" and "Look in:" as shown below. Next select the desired files and drag them into the "Input Data" window of BioTools.

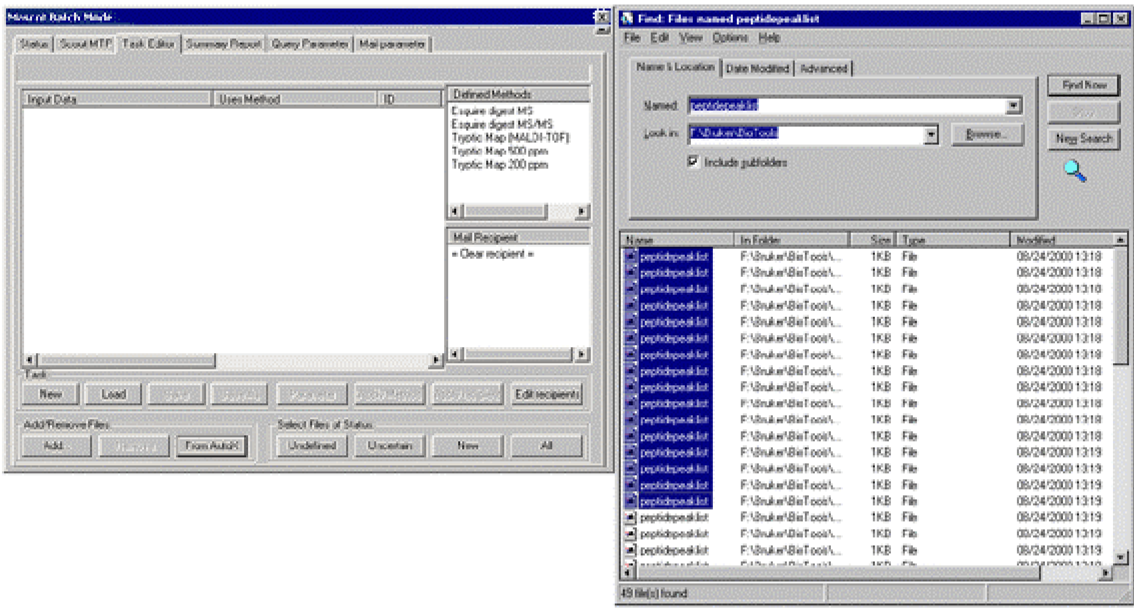


Figure 3-59, Drag and Drop

Data from AutoX

Imports the spectra information directly from the EXCEL spreadsheet used for definition of AutoX runs. The *.txt filename created from the Data Spreadsheet has to be specified. The file holds information about all positions used on the MTP. Use "Browse" to find the file. It is in the same directory in which the root directory of the AutoX run is stored "Process Spreadsheet" reads the spreadsheet, creates the information for a new task and displays a message about the number of files found and the number of missing files. To see the names of missing files use "View Missing" Typically, the calibration spectra are "missing". To see the contents of the file in case of problems, use "View Spreadsheet".

For some reasons (e.g., if the spreadsheet file is not stored on its standard location) it is necessary to enter some more information about the data used. Use the button More to expand the dialog. Use "**Data Directory**" to specify the directory where your data are stored. Use "**1SRef**" etc. to specify the spectrum type and the location of the peak lists (**1Sref** has to be used for XTOF 5.1 and higher, **SRef** for the previous versions). In the spreadsheet, the information of spectrum directory and filename is stored and normally used to locate the corresponding files. If you changed the entry for the "**Data Directory**" it may become necessary to ignore the information stored in the spreadsheet file.

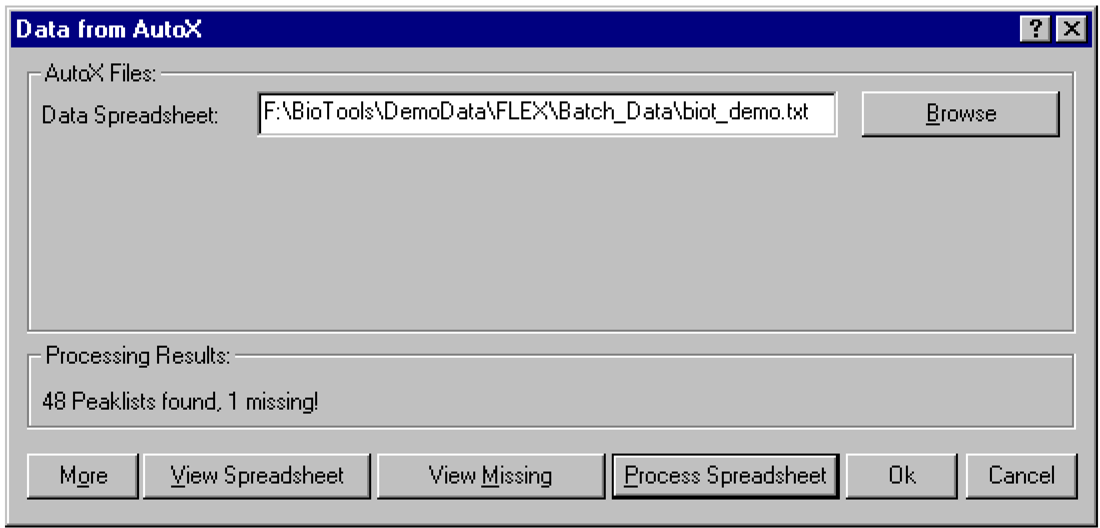


Figure 3-60, Data from AutoX, normal view

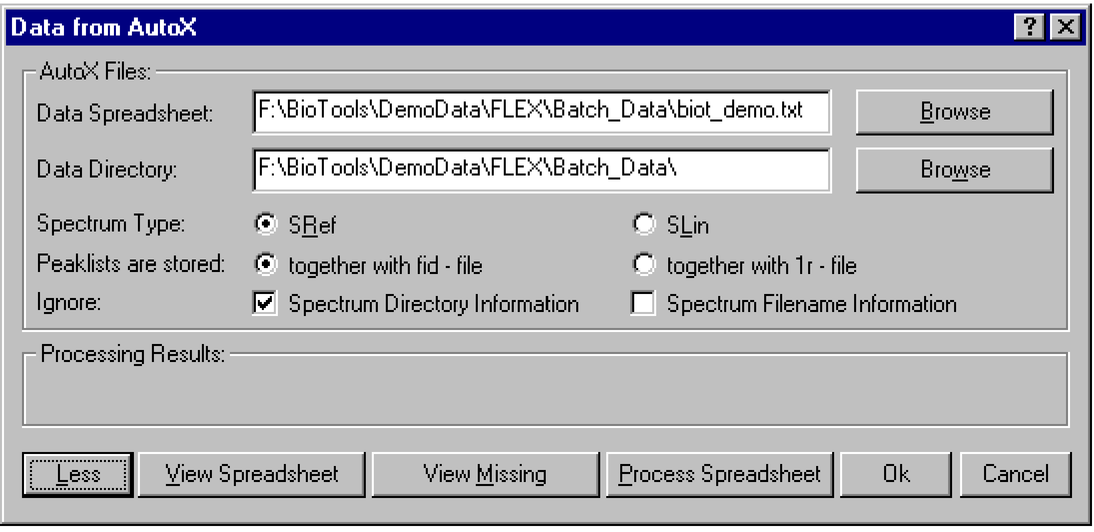
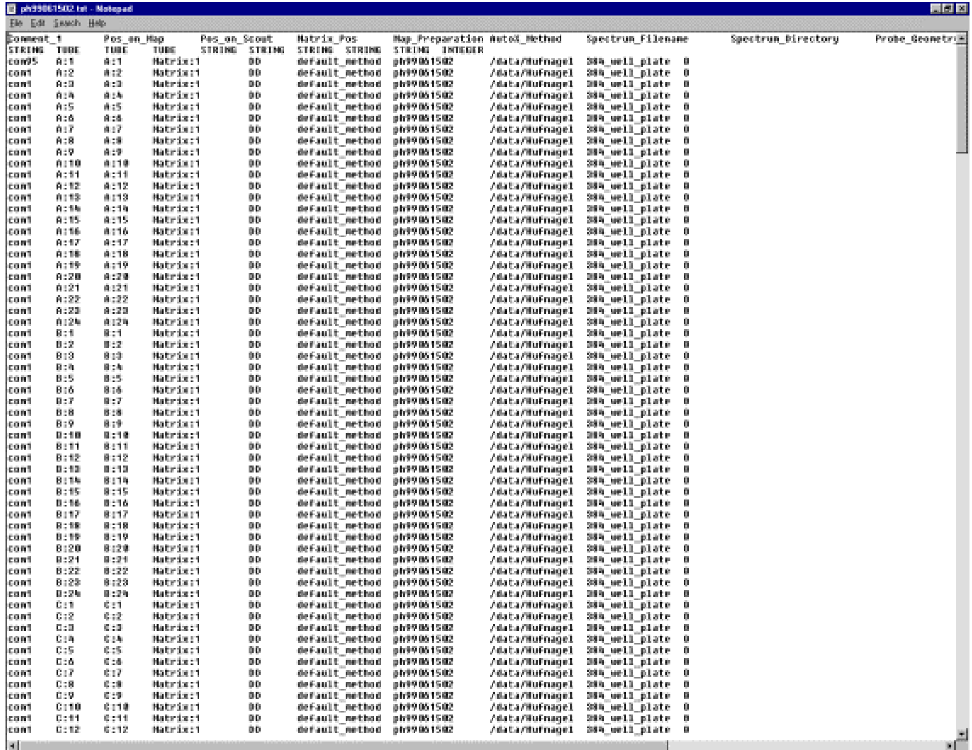


Figure 3-61, Data from AutoX, extended view (by selecting " More ")



```
Comment_1         Pos_on_Map        Pos_on_Scout      Matrix_Pos        Map_Preparation AutoX_Method      Spectrum_Filename         Spectrum_Directory        Probe_Geometr
STRING   TUBE     TUBE     TUBE     STRING   STRING   STRING   STRING   STRING   INTEGER
con95    A:1      A:1      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:2      A:2      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:3      A:3      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:4      A:4      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:5      A:5      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:6      A:6      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:7      A:7      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:8      A:8      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:9      A:9      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:10     A:10     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:11     A:11     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:12     A:12     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:13     A:13     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:14     A:14     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:15     A:15     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:16     A:16     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:17     A:17     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:18     A:18     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:19     A:19     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:20     A:20     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:21     A:21     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:22     A:22     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:23     A:23     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     A:24     A:24     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:1      B:1      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:2      B:2      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:3      B:3      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:4      B:4      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:5      B:5      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:6      B:6      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:7      B:7      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:8      B:8      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:9      B:9      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:10     B:10     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:11     B:11     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:12     B:12     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:13     B:13     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:14     B:14     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:15     B:15     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:16     B:16     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:17     B:17     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:18     B:18     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:19     B:19     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:20     B:20     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:21     B:21     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:22     B:22     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:23     B:23     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     B:24     B:24     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:1      C:1      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:2      C:2      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:3      C:3      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:4      C:4      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:5      C:5      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:6      C:6      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:7      C:7      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:8      C:8      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:9      C:9      Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:10     C:10     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:11     C:11     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
con1     C:12     C:12     Matrix:1          DD       default_method    ph99061502        /data/Hufnagel    384_well_plate    0
```

Figure 3-62, Spreadsheet

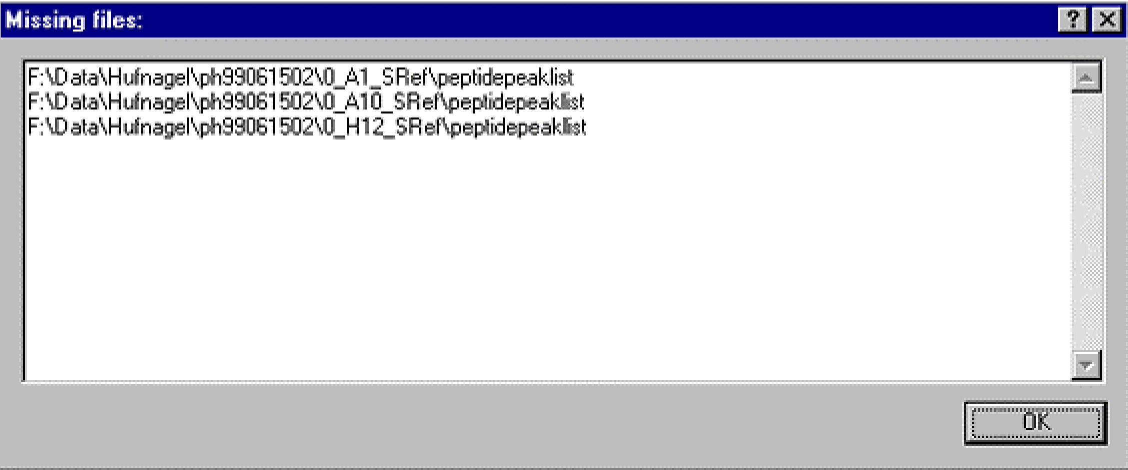


Figure 3-63, Information about missing files

Edit Data

With the Edit data dialog the method names, the email addresses or URL for web-search can be edited. The Edit data dialog is available from several Edit… buttons in the tabs of the Mascot batch mode dialog.

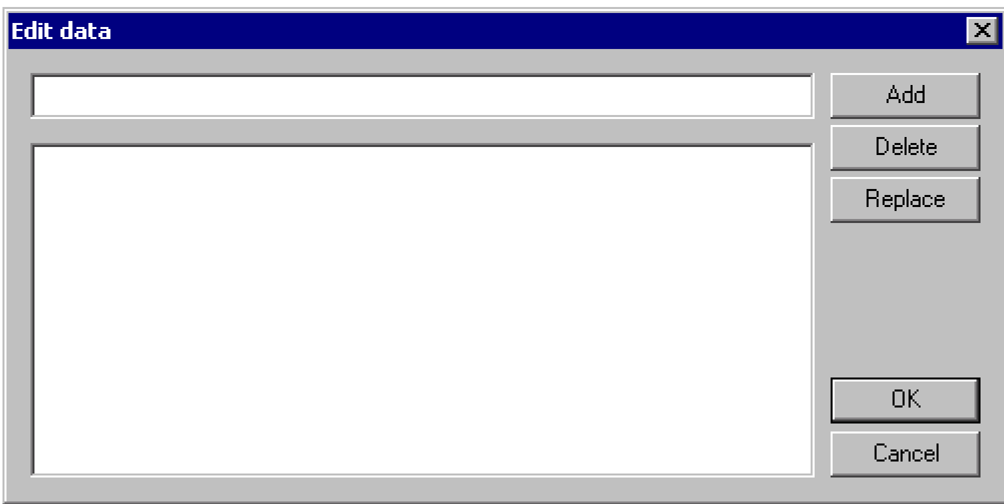


Figure 3-64, Edit data dialog box

3.5.3.4.4. Tab Summary Report

Use the report dialog to specify the parameter to be printed. The left window shows all the entries that will not be printed, the right one shows all parameter selected for printing.

Add — Add an entry to your report. Select the parameter from the "Don't Report" list and click the "Add" button.

Edit — Changes the output parameter for a selected entry via the Edit Parameter Dialog.

Up — Moves the selected entry one position upwards and changes the print order.

Down — Moves the selected entry one position downwards and changes the print order.

Remove — Removes an entry from your report. Select the parameter from the "Report" list and click the "Remove" button.

Result Layout shows the arrangement of the columns in the printout. Do not exceed 45 characters for one line in print portrait mode or 75 characters in print landscape mode, depending on printer and style.

Save Selection Saves the selection.

Export Exports the results to a text file (*.txt) **or** to a batch report file (*.btr), which can then be imported into the AutoX Excel spreadsheet.

Print Prints the results in an EXCEL-type table according to the specified field definitions.

Send Sends the results to an email recipient, chosen in the dropdown list. This mail recipients list can be edited on the Mail Parameter tab via the button Edit Recipients.

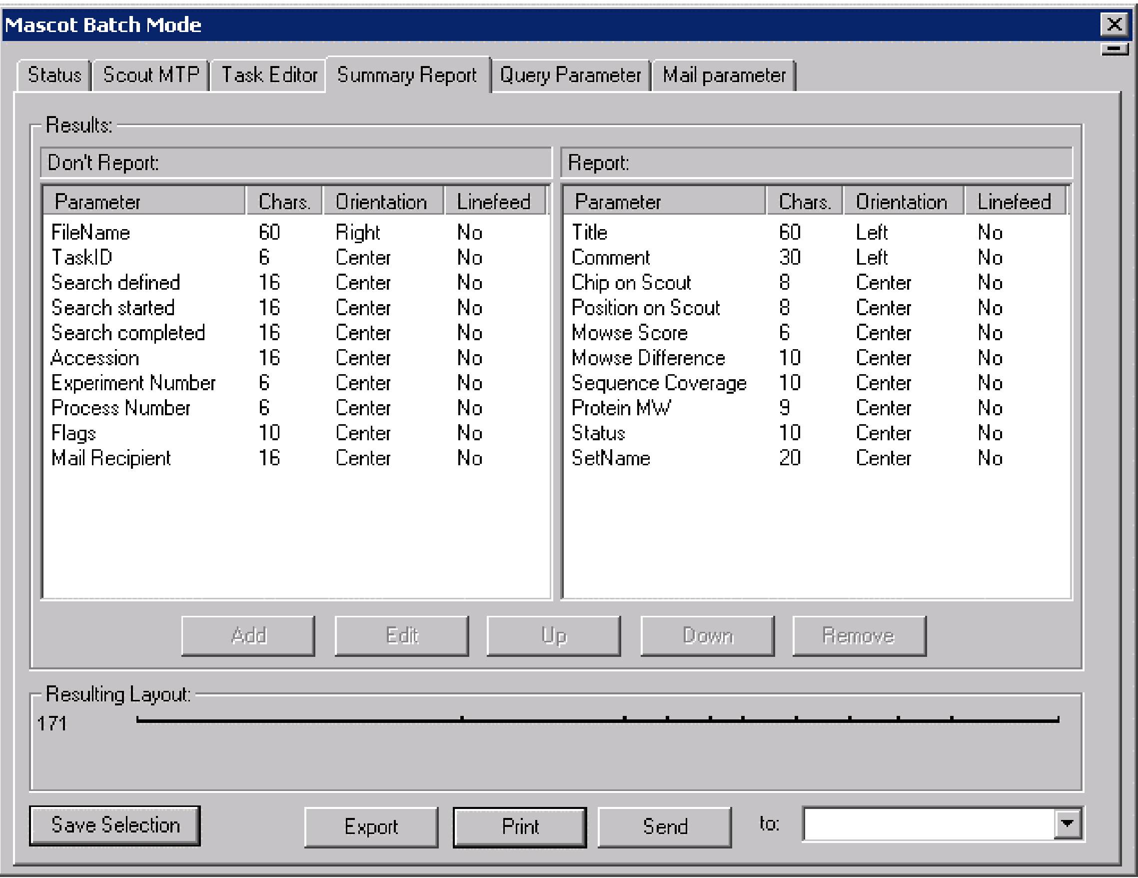


Figure 3-65, Summary Report tab

Edit Parameter

Changes the output parameter for a selected entry (number of characters, orientation and line feed, see Figure 3-66).

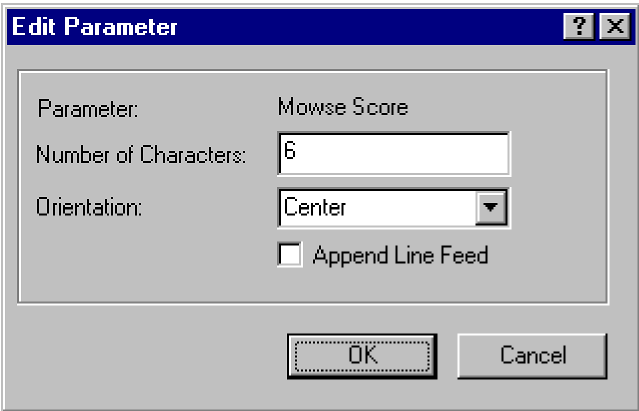


Figure 3-66, Edit parameter dialog

3.5.3.4.5. Tab Query Parameter

This window is used to specify all parameters necessary for a query.

Method:

Name: Shows the name and type of the selected Mascot search method. After installation BioTools uses some default methods. Before you run your first batch please check if the selected method is appropriate for your application.
The name of the method can be choosen from the dropdown list.

Type: The Type of the method can be choosen from **MS Fingerprint** or **MS/MS** of the dropdown list.

Edit Use this entry to change the list of methods.
The server list can be edited via the button Edit.

Note ***BioTools can execute queries initiated by Bruker DataAnalysis automatically without any user interaction. The Default method will be used in this case, make sure it is set properly before you start the esquireSeries acquisition.***

Query on www.own-server.com:

URL: The URL of your Mascot server including the program name to perform the query. If the name of your Mascot server computer is "MyServer" and you used the defaults during Mascot installation the resulting URL is:

http://myserver/mascot/cgi/nph-mascot.exe?1

Edit Use this entry to change the list of URLs for Mascot server. The server list can be edited via the button Edit. A changed method must be saved with the Apply Button at the bottom

The original input form for MS Fingerprint searches can be found at:

http://www.matrixscience.com/cgi/search_form.pl?SEARCH=PMF

The original input form for MS/MS searches can be found at:

http://www.matrixscience.com/cgi/search_form.pl?SEARCH=MIS

Defaults Use this Default button to define the current method to the **default method for MS Fingerprint** and **default method for MS/MS** searches.

Apply Saves the modified method.

Update Task Saves the modified method and updates it in the current task list.

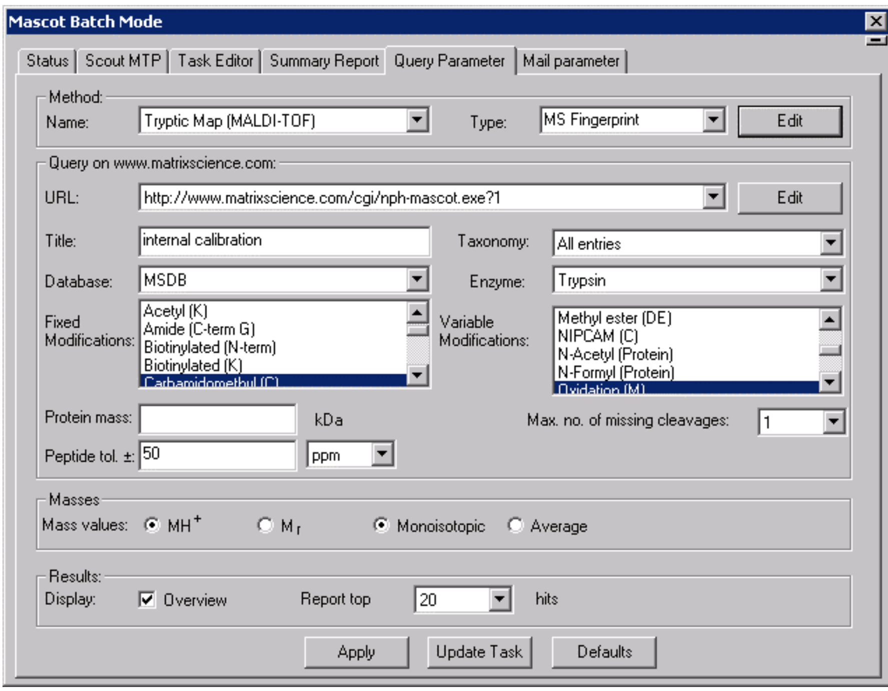


Figure 3-67, Query Parameter tab

Batch defaults

Use this button to define the current method to the **default method for MS Fingerprint** and **default method for MS/MS** searches.

The methods must be named in the Edit dialog of the method, the method and its type must be set and the new (changed) method must be applied via the Apply button in the Query Parameter tab.

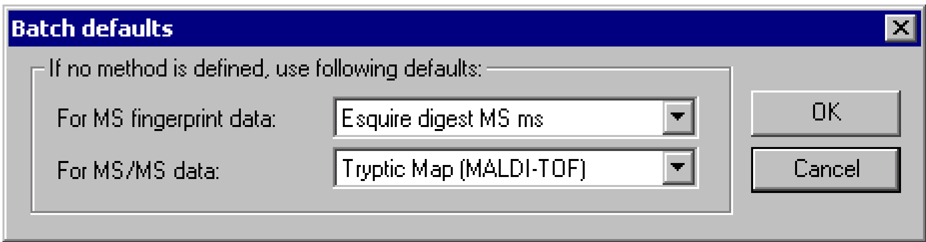


Figure 3-68, Batch defaults dialog

3.5.3.4.6. Tab Mail parameter

This window is used to adjust the mail parameters necessary for sending data via email.

Send Results — Disabling this checkbox will **not** send the results via email

Send one mail for each spectrum — For each spectrum one mail will be sent with the detailed **Mascot Search Results**

Send summary report — All results will be sent in a summary report

In the field **Mail Server** enter the respective address (e. g. postman) and in **From** and **To** the desired full email address.

If needed enable **Server requires authentication** and enter **User** name and **Password**.

In the field **Subject** enter the title of the email and in **Body** text of the email, what has to be done with the received data or other useful hints for the recipient.

Select the checkboxes **Include filename** and **Include title** to submit the filename and/or title of the data file.

Edit recipient Use this button to change the list of URLs for the mail recipient.

Send Testmail Use this button to send a mail for testing of a proper working internet connection and existing email recipient.

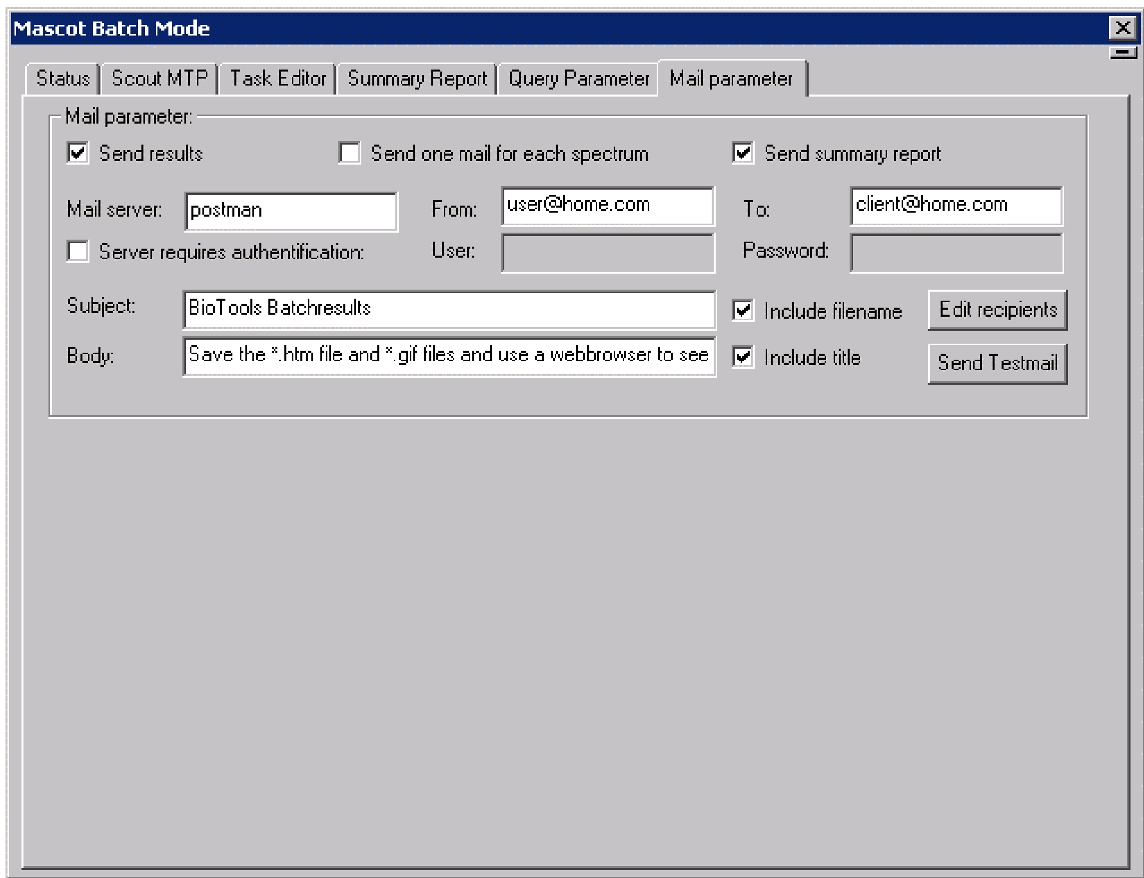


Figure 3-69, Mail Parameter tab

3.5.4. Search for Masses (SequenceEditor)

When you obtain the molecular weight of a peptide, you can find out where in a given protein sequence the peptide came from. Such a search can be performed on the whole peptide set of a digest, or only of selected peptides, e.g., the ones which were previously not been identified in a database search. For such analyses the *Search - Mass Search* command can be used in the SequenceEditor, which is automatically started and the relevant peptide masses are transferred together with all information from the Mascot Search dialog. Please Refer to the SequenceEditor tutorial **Protein Digest** for further details.

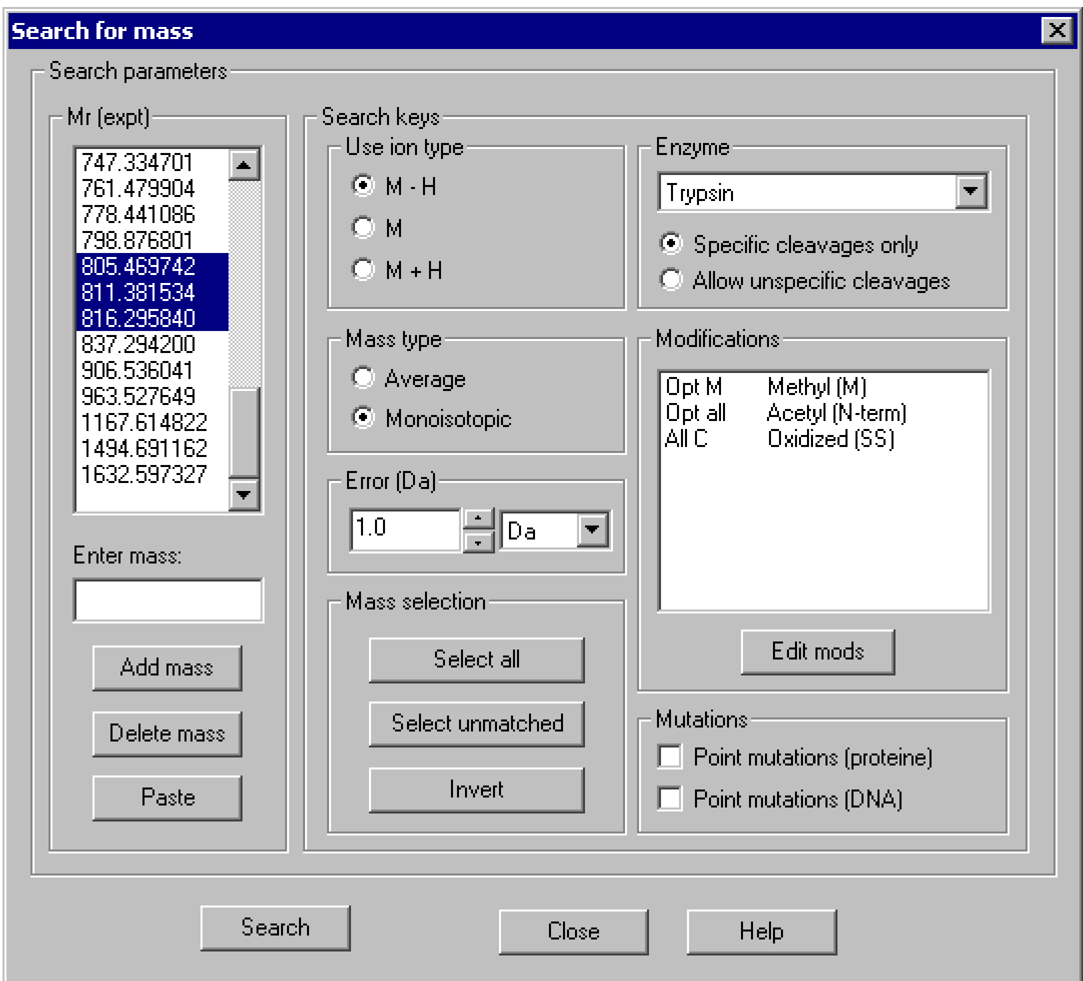


Figure 3-70, Peptide Mass query in sequence window

3.5.5. Digest (SequenceEditor)

Call this command to perform an enzymatic digest or to define new digest agents. This will change to SequenceEditor and to the Perform digest dialog.

Digest agents fragment a particular protein (or DNA or RNA) into smaller peptides after recognition of a particular amino acid residue or a more complex sequence motif. The task in these analyses is to match the mass spectrum with all its peaks to the protein sequence + modifications to account for as much structural information as possible. Artifacts make this task sometimes difficult: unspecific or irregular digests, unknown modifications, contaminations, autoproteplysis of the cutting enzyme, etc.

Calculations results can be used locally, like for the prediction of enzymatic peptide fragment masses or in conjunction with particular datasets: MALDI or LC-MS fingerprints, etc.

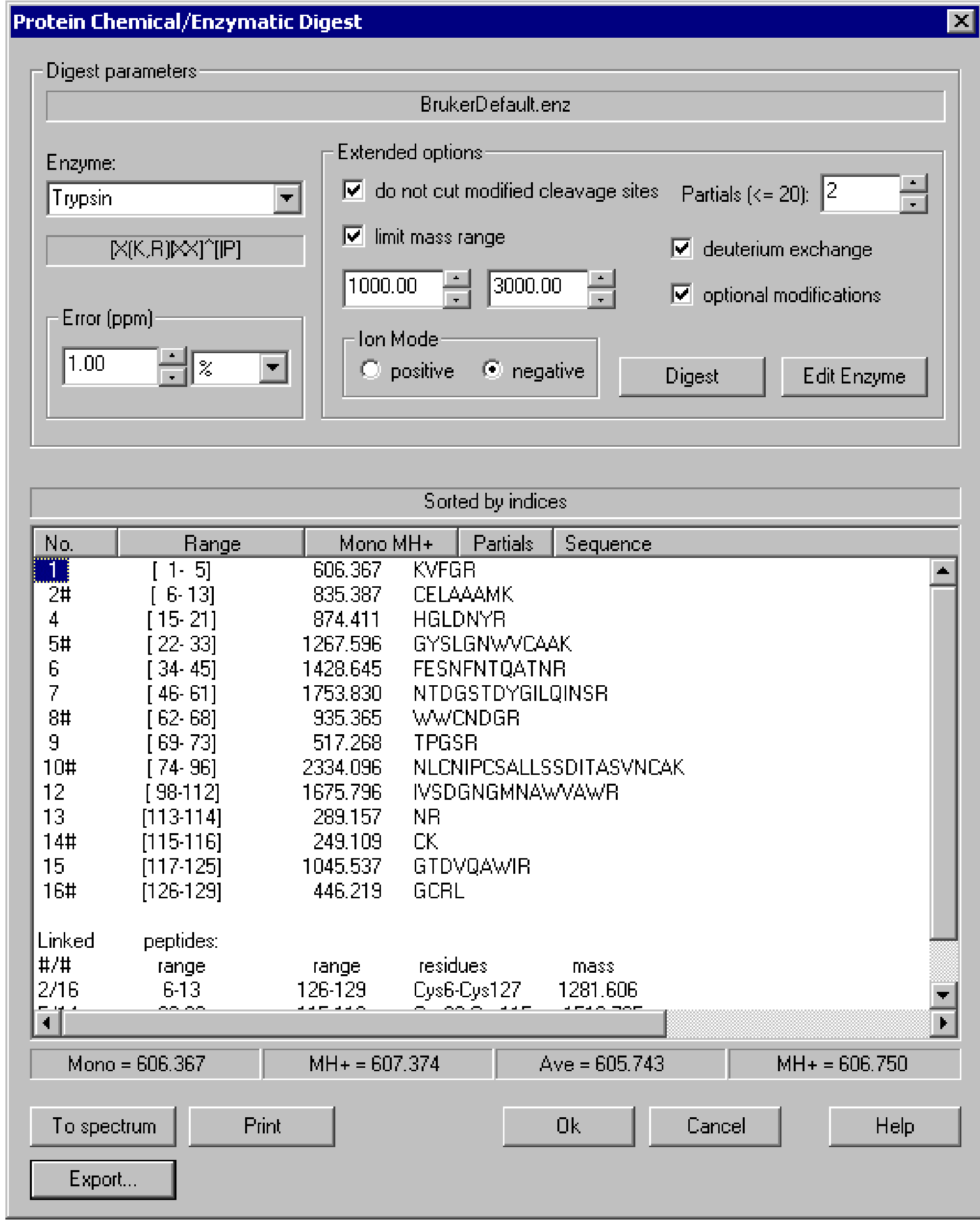


Figure 3-71, Perform Digest dialog in SequenceEditor

3.6 Tools Menu

With the *Tools Menu* the options (display, general, print profiles) can be chosen and edited. Also the parameter of the spectrum can be reviewed and other programs and commands can be started.

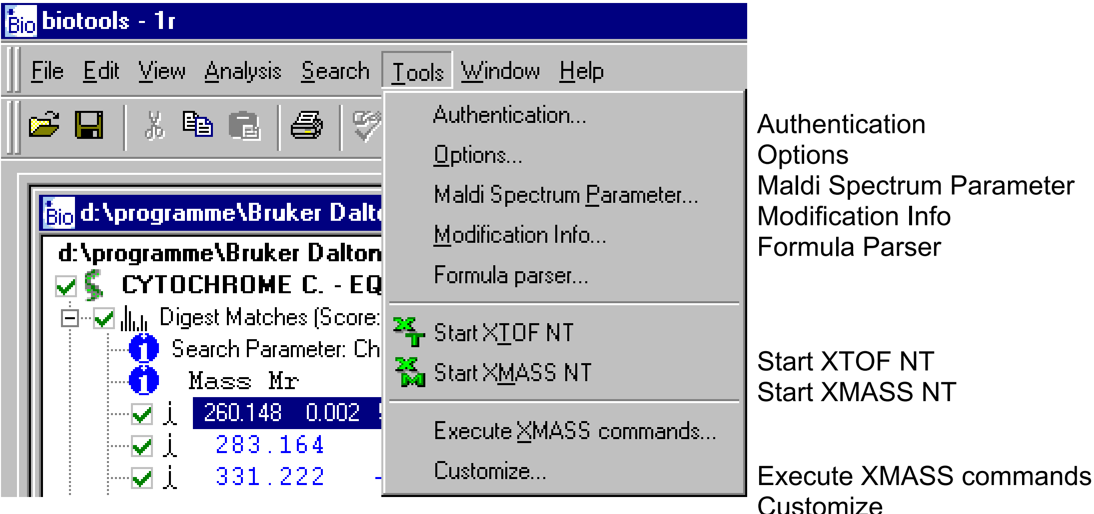


Figure 3-72, Tools menu

3.6.1. Authentication

On some networks it may be necessary to authenticate yourself using the remote host computer and/or using the local proxy server. If e.g. Mascot is running on a remote computer in your network and there is no anonymous access on this web server you will need an account on this server. Enter the username and password and the port number if the number is different from 8080. If there is a proxy server which also requires authentication enter user name and password also.

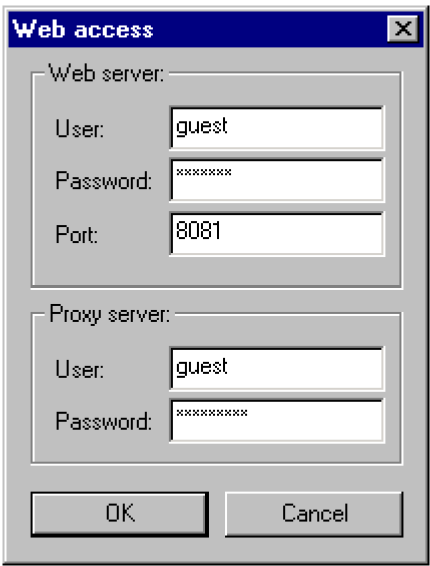


Figure 3-73, Dialog for web access and authentication to web and proxy server

3.6.2. Options

Here you can create your own print profiles (with special parameters, which shall be printed or not). Also general and display settings can be defined.

General:

Reload last File	At the startup of BioTools the last spectrum will be loaded.
Save Settings on exit	The settings (such as display and print option) were saved and reloaded by the next starting of the BioTools program.
Write Debug File	Writes a debug file (MessageLog.txt) in the BioTools directory. This file may be helpful in case of problems during Mascot queries. This text file can be read with any text editor.

Display:

Workbookmode	Changes the view between normal mode and workbookmode in which each data file is easily accessible.
Flickerfree	The display in all windows will be actualized without interruption (flicker) when a new file or new sequence is chosen (set default).
Number of Digits	Number of digits for the mass values in the fragments window and the spectrum.

Print:

Use Profile for:	List of available profile names, which can be edited
Parameter Summary	All parameters chosen in the Show parameter window under Summary are printed in the title of the first print page
Spectrum	Prints the spectrum
Processing Parameter	All parameters chosen in the Show parameter window under Processing are printed
Acquisition Parameter	All parameters chosen in the Show parameter window under Acquisition are printed
Display Parameter	All parameters chosen in the Show parameter window under Display are printed
Peaklist	The peaklist with number, mass and intensity of peaks will be printed
Calculated Masses	The MS/MS fragment ion masses are printed.
Mass Difference [Da]	The mass errors (in Da) displayed in the fragments window are printed.
Mass Difference [ppm]	The mass errors (in ppm) displayed in the fragments window are printed.
Matched Sequences	The matched sequences of the Treeview window are printed.

To create a new print profile, click on Add Profile and enter a new profile name.

To delete a print profile click on Delete Profile and confirm with yes.

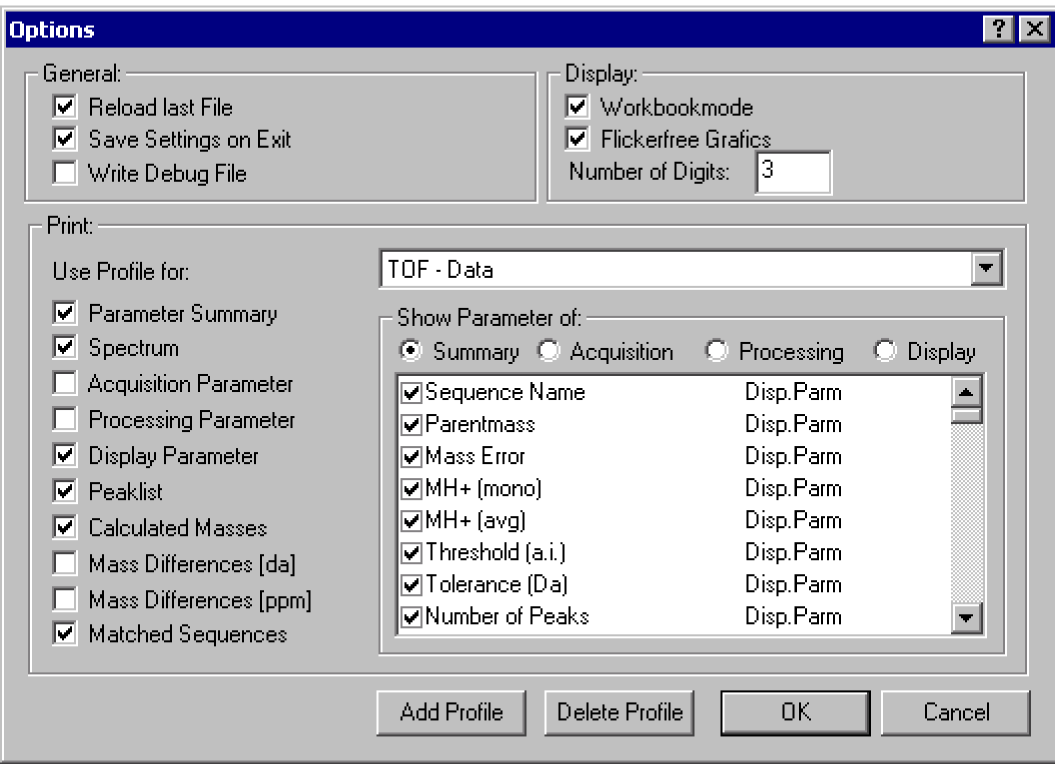


Figure 3-74, Options window

Add Profile

To create a new print profile, click on Add Profile in Figure 3-74. The new profile dialog (Figure 3-75) appears and the name can be entered.

The default print profiles (Esquire – Data, FAST – Data, ICR – Data, TOF - Data) can not be overwritten. Changes can be saved under a new name.

Figure 3-75, Add new profile window

Delete Profile

To delete a print profile click on Delete Profile in Figure 3-74. Figure 3-76 appears to confirm deleting this profile.

The default print profiles (Esquire – Data, FAST – Data, ICR – Data, TOF - Data) can not be overwritten. Changes can be saved under a new name.

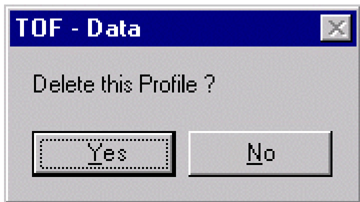
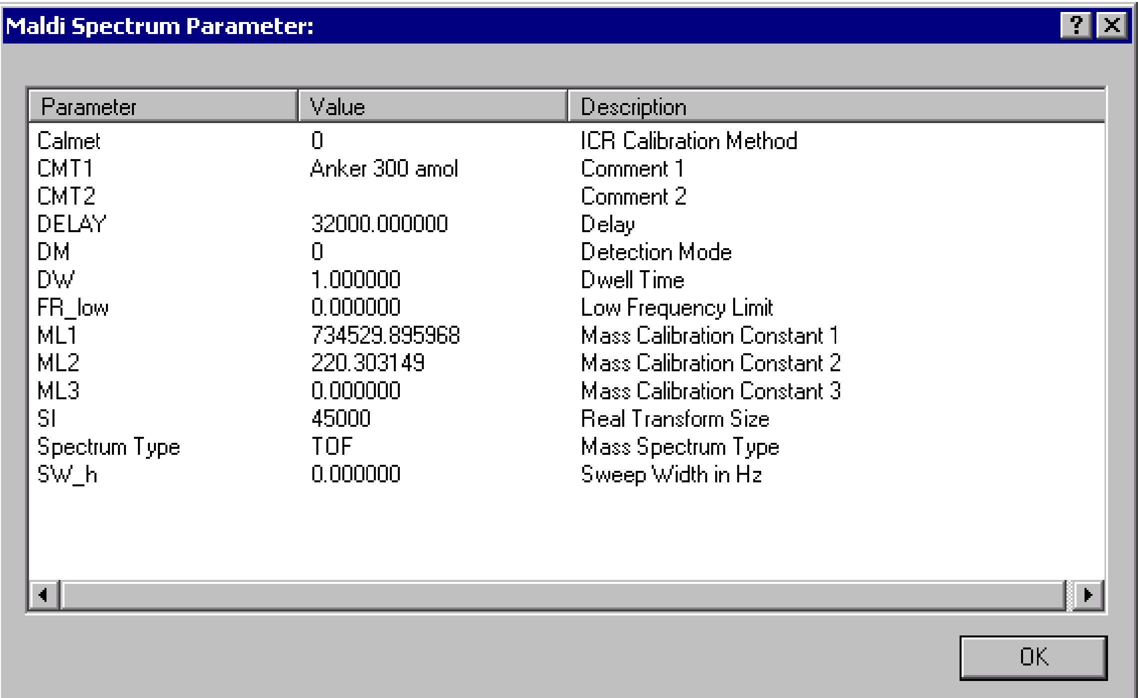


Figure 3-76, Delete profile window

3.6.3. Maldi Spectrum Parameter

With this option the Maldi spectrum parameters can be reviewed.



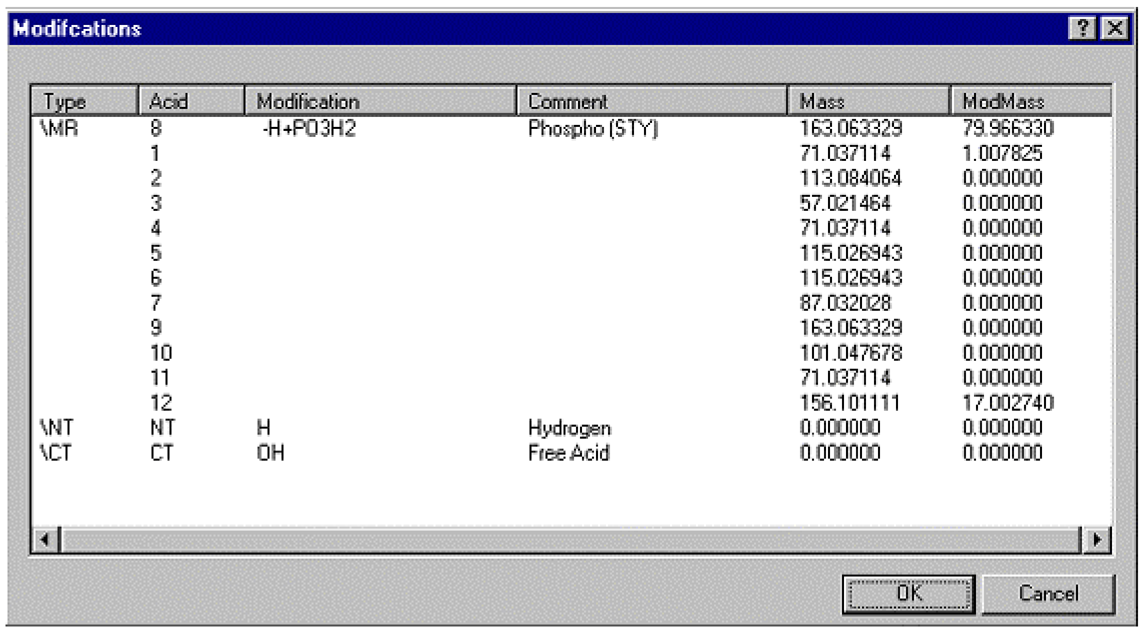
Parameter	Value	Description
Calmet	0	ICR Calibration Method
CMT1	Anker 300 amol	Comment 1
CMT2		Comment 2
DELAY	32000.000000	Delay
DM	0	Detection Mode
DW	1.000000	Dwell Time
FR_low	0.000000	Low Frequency Limit
ML1	734529.895968	Mass Calibration Constant 1
ML2	220.303149	Mass Calibration Constant 2
ML3	0.000000	Mass Calibration Constant 3
SI	45000	Real Transform Size
Spectrum Type	TOF	Mass Spectrum Type
SW_h	0.000000	Sweep Width in Hz

Figure 3-77, Maldi Spectrum Parameter window

3.6.4. Modification Info

This window shows information about the currently defined modification within the amino acid sequence.

Type	Modification type: NT: N-terminal; CT: C-terminal;
Acid	Residue Position
Modification	Elemental composition change relative to the unmodified residue
Comment	Description according to the definition in the SequenceEditor
Mass	Gross mass of the modified residue
ModMass	Mass of the modification (N-terminal H and C-terminal OH are added to the terminal residues in case of unmodified peptide)



Type	Acid	Modification	Comment	Mass	ModMass
\MR	8	-H+PO3H2	Phospho (STY)	163.063329	79.966330
	1			71.037114	1.007825
	2			113.084064	0.000000
	3			57.021464	0.000000
	4			71.037114	0.000000
	5			115.026943	0.000000
	6			115.026943	0.000000
	7			87.032028	0.000000
	9			163.063329	0.000000
	10			101.047678	0.000000
	11			71.037114	0.000000
	12			156.101111	17.002740
\NT	NT	H	Hydrogen	0.000000	0.000000
\CT	CT	OH	Free Acid	0.000000	0.000000

Figure 3-78, Modifications info

3.6.5. Formula Parser

Enter chemical formula to get the mass.

The formula parser calculates the average and mono isotopic mass of a given formula. Brackets (e.g. $CH_3(CH_2)_4CH_3$) and Deuterium (D) are not supported in the current version.

If you enter a faulty formula, the counter will not increase anymore, until the formula is correct.

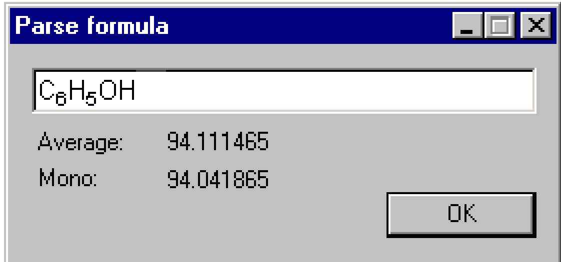


Figure 3-79, Formula Parser

3.6.6. Start XTOF NT

This menu option starts the program XTOF NT for data processing, in particular for peak (re)assignments. **Only available if MALDI-TOF, APEX or BioTOF data are used.**

3.6.7. Start XMASS NT

This menu option starts the program XMASS NT for APEX data.

3.6.8. Execute XMASS/XTOF-Commands

With this option a list of execute commands can be edited (loaded, changed and saved) and finally started to run on XMASS/XTOF under UNIX via BioTools.

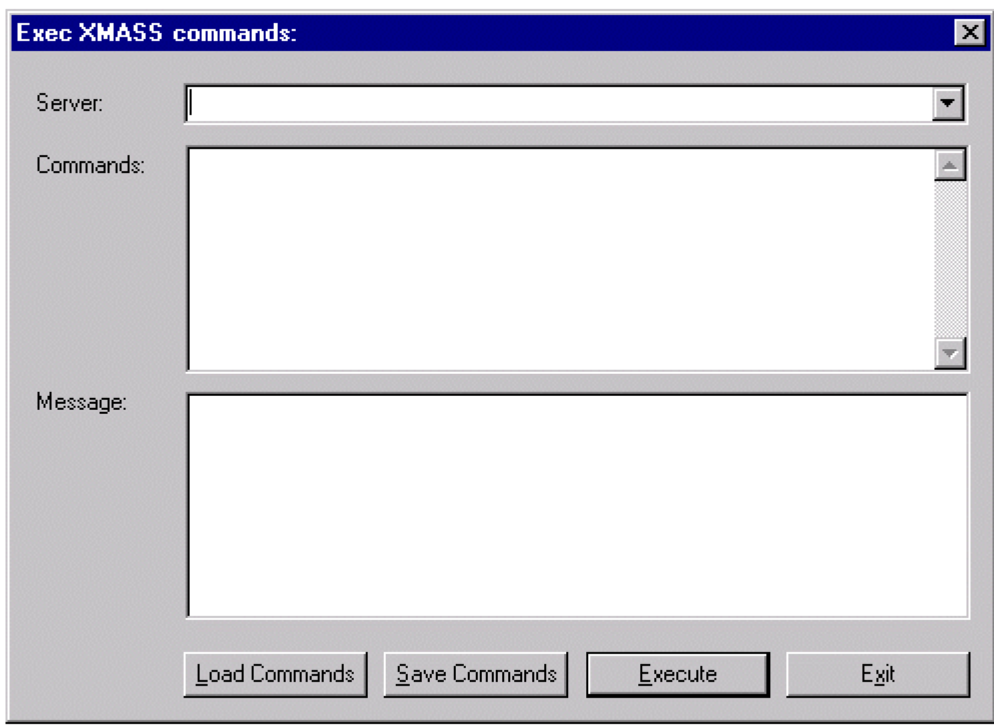


Figure 3-80, Options window

Server: Select the server to which the commands shall be sent.

Commands: Enter the desired commands by using the Load Commands button.

Message: Enter here a message as description for the commands.

Load Commands Opens a dialog box to search for *.cmd files.

Save Commands Opens a dialog box to save the list of commands in a *.cmd file.

Execute The commands will be executed for controlling if they are correctly implemented.

Exit Exits this dialog box.

3.6.9. Customize

Customize your toolbars and the menu bar.

On selecting this command, the Customize dialog box appears containing the Toolbars tab and Commands tab.

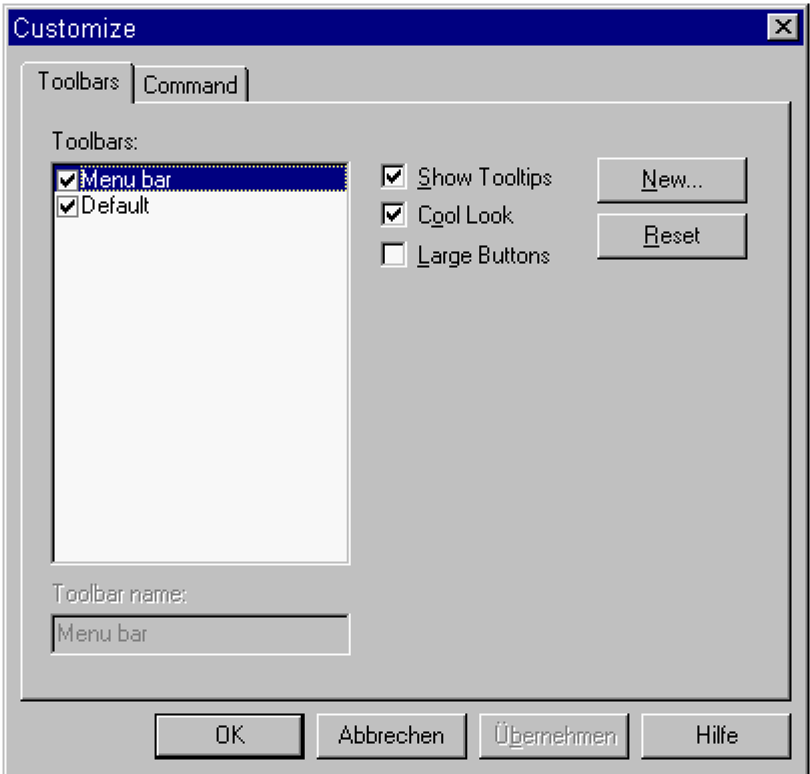


Figure 3-81, Customize dialog window

3.6.9.1. Toolbars Tab

With the toolbars tab you can define the toolbars to be displayed on the user interface, to define the toolbars display, to create new toolbars, to reset the default toolbars or menu bar, or to delete previously created toolbars.

All toolbars currently defined (default as well as customer created toolbars) are listed here. Each toolbar has a check box in front of it indicating its current display status: If the check box is checked, the respective toolbar is displayed, if it is unchecked the toolbar is hidden.

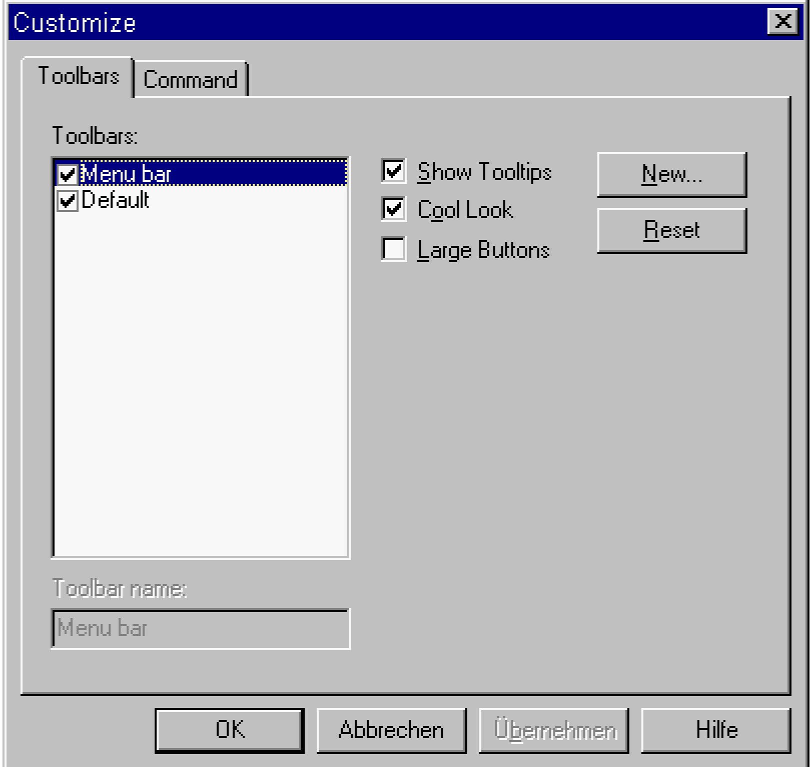


Figure 3-82, Customize toolbars

Toolbars — Lists all toolbars currently defined (default as well as customer created toolbars). Each toolbar has a check box in front of it indicating its current display status: If the check box is checked, the respective toolbar is displayed, if it is unchecked the toolbar is hidden. Check an unchecked check box to display the respective toolbar. Uncheck a checked check box to hide the respective toolbar.

Toolbar name — Displays the name of the toolbar currently selected in 'Toolbars'.

Show Tooltips Check this option if tool tips shall be displayed on positioning the mouse pointer onto toolbar buttons.

Cool Look Check this option if toolbars and menu bar shall be displayed in cool look. In cool look the buttons are displayed in flat design and each bar contains a gripper. Otherwise the toolbars are displayed in hot look.

Large buttons Check this option if toolbar buttons shall be displayed enlarged.

New Allows you to create a new toolbar.

Reset Resets the toolbar currently selected in 'Toolbars' or the menu bar to its default settings. This button is only available for default toolbars and the menu bar.

Delete Deletes the toolbar currently selected in 'Toolbars'. This button is only available for customer toolbars.

New Toolbar Dialog Box

With this dialog box a new toolbar can be created and named.

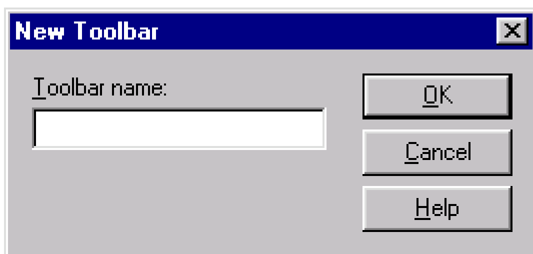


Figure 3-83, New Toolbar dialog box

Toolbar name Enter the name for the new toolbar.

OK Creates a new, undocked toolbar on the user interface.

Cancel Cancel creating a new toolbar.

3.6.9.2. Commands Tab

Use the Commands tab from the Customize dialog box to customize the toolbars and the menu bar. This tab allows you to add buttons to currently displayed toolbars and menus to the menu bar. Also you can use it to view information about the available toolbar buttons. To add a button to a toolbar, click the respective button and drag it with the left mouse button to any desired toolbar.

To view information about a particular button, click it. Its description is displayed in the 'Description' field.

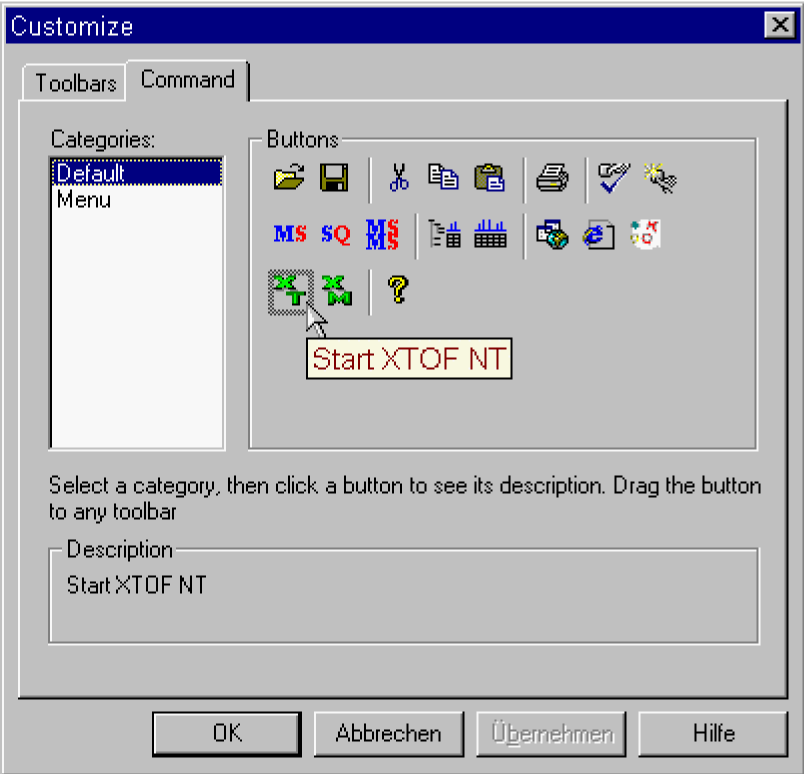


Figure 3-84, Customize options window

Categories:	Lists the default toolbars. On selecting a category, the buttons contained in this category are displayed right in the 'Buttons' field. For Menu bar the menus contained are displayed.
Buttons	Displays the buttons defined for the selected category.
Description	Displays the description for the button currently selected in the 'Buttons' field.

3.7 Window Menu

With the *Window Menu* the arrangement of all opened data files in the BioTools window can be adjusted. There is also a list of all opened data files at the bottom of this menu for easier switching between the data files.

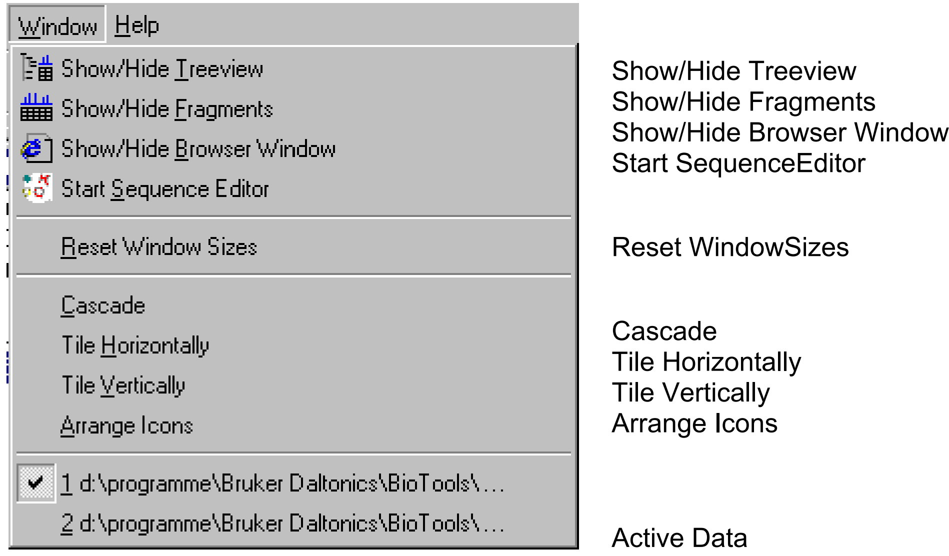


Figure 3-85,   Window menu

3.7.1. Show/Hide Treeview

With this option the left Treeview window can be shown or hidden. The treeview contains sequence information. In case of MS spectra it is sequence names followed by identified peptides together with their neutral masses and other information. In case of MS/MS data, the sequences in the treeview are used to calculate fragment ions.

3.7.2. Show/Hide Fragments

With this option the lower fragment ion window can be shown or hidden. Currently a meaningful setting is **Hide** for MS data and **Show** for MS/MS data.

The view of the Overview, Zoom, Fragments and Treeview windows can also be modified with the cursor. By moving on the line between two windows the cursor changes to a parallel cursor. Now moving the line can change the size of the windows – with held left mouse button – to the desired range. In the fragments window is shown either a Check Sequence table or the *DeNovo* sequencing parameters.

3.7.3. Show/Hide Browser Window

With this option the browser window can be shown or hidden.

3.7.4. Start SequenceEditor

Starts the Sequence Editor program to edit sequences. This entry just starts the SequenceEditor **without transmitting information** from the currently active dataset.

To **transmit the sequence to SequenceEditor** of an active MS/MS spectrum, click the Sequence button underneath the spectrum or double click on the treeview entry.

To **transmit the sequence of a protein together with digest masses**, double click onto the respective entry (Digest matches) in the treeview.

3.7.5. Reset Window Sizes

The arrangement of the window in BioTools will be set to the default settings of the manufacturer.

3.7.6. Cascade

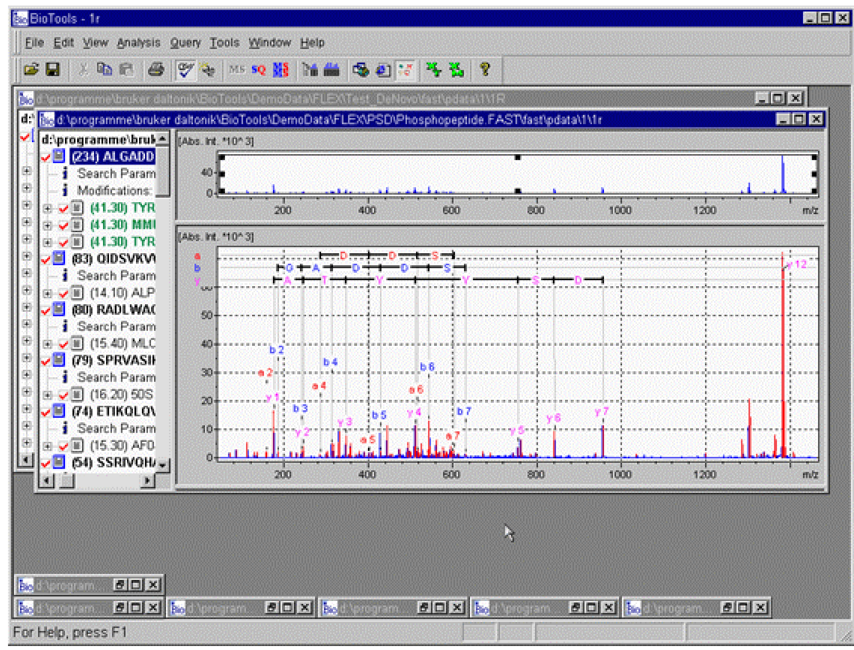


Figure 3-86, Cascaded data files in the BioTools window

With this option all opened, but not minimized data files are cascaded in the BioTools window. All minimized data files are arranged simultaneously at the bottom of the BioTools window. The active data file will be displayed on the top of all other data files.

3.7.7. Tile Horizontally

With this option all opened, but not minimized data files are arranged horizontally in the BioTools window. All minimized data files are arranged simultaneously at the bottom of the BioTools window. The active data file will be displayed on the top of all other data files.

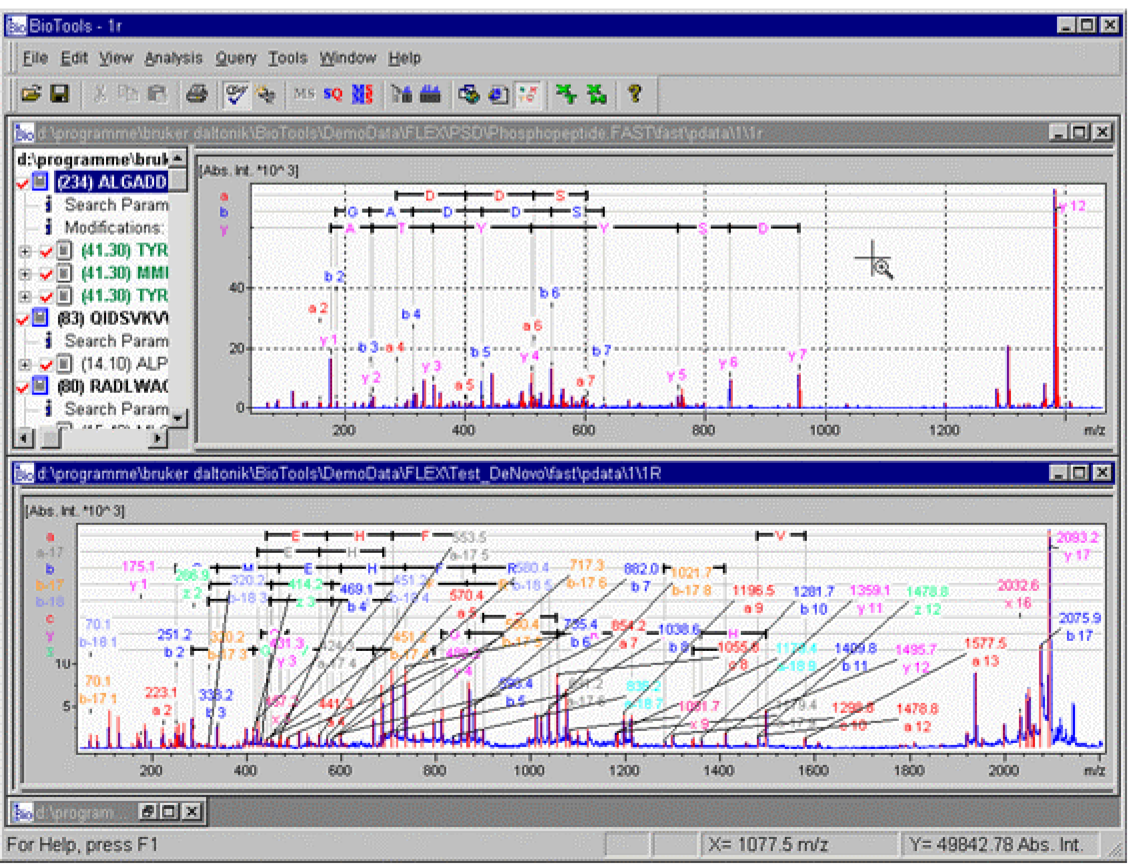


Figure 3-87, Horizontally tiled data files in the BioTools window

3.7.8. Tile Vertically

With this option all opened, but not minimized data files are arranged vertically in the BioTools window. All minimized data files are arranged simultaneously at the bottom of the BioTools window. The active data file will be displayed on the left of all other data files.

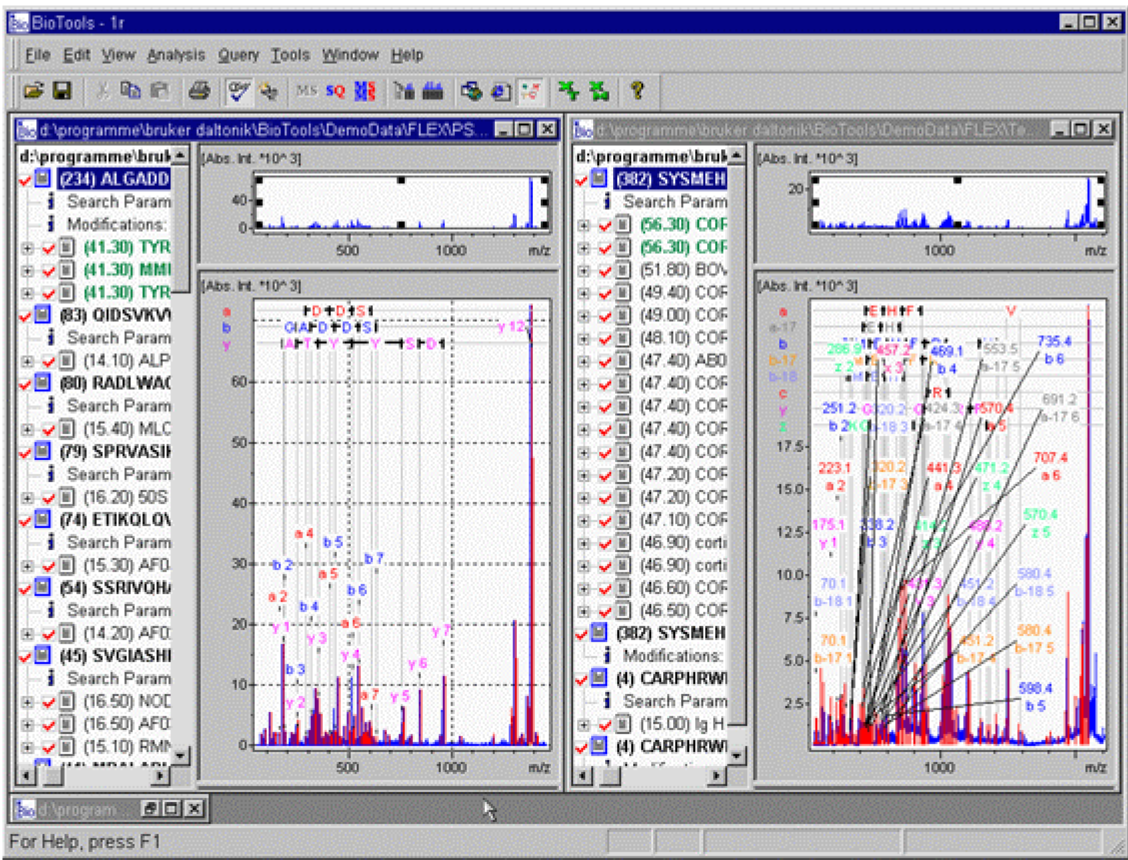


Figure 3-88, Vertically tiled data files in the BioTools window

3.7.9. Arrange Icons

With this option all minimized data files are arranged at the bottom of the BioTools window.

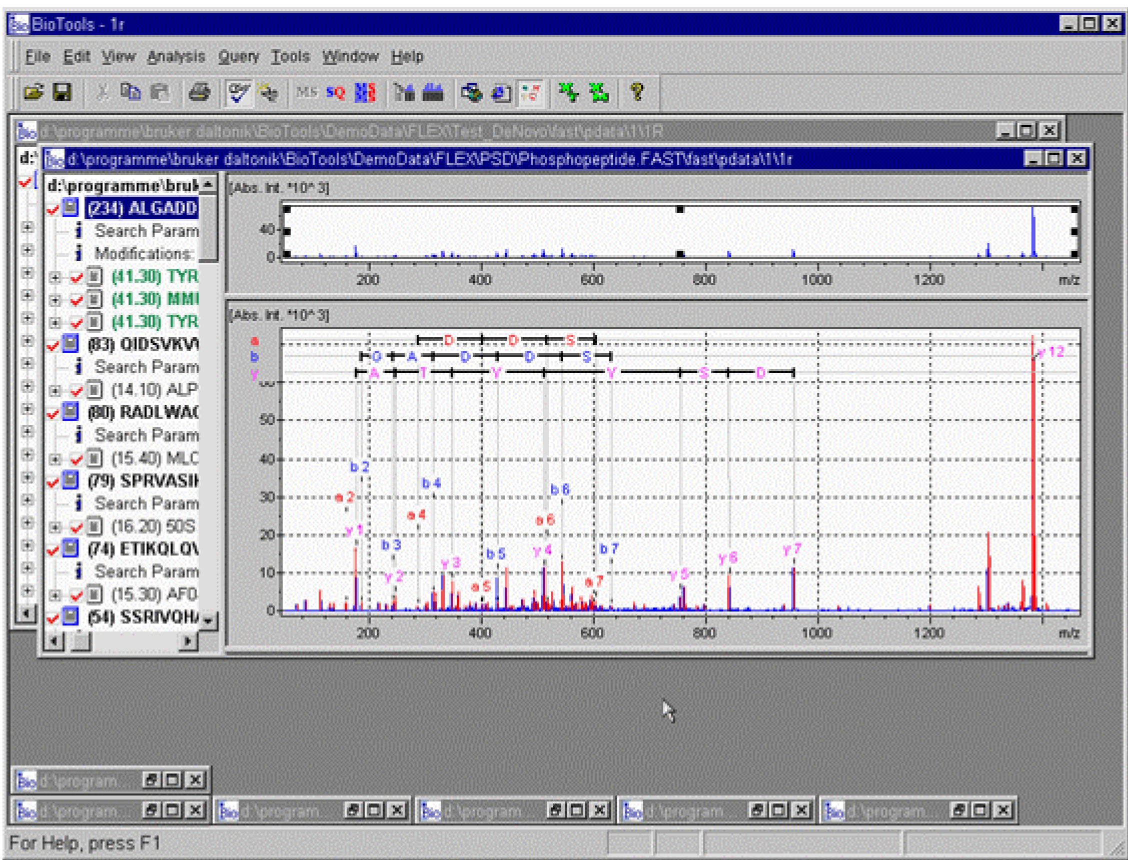


Figure 3-89, Arranged icons at the bottom of the BioTools window

3.7.10. Active Data

Gives a list of all opened data files. The active data file is signed with a mark.

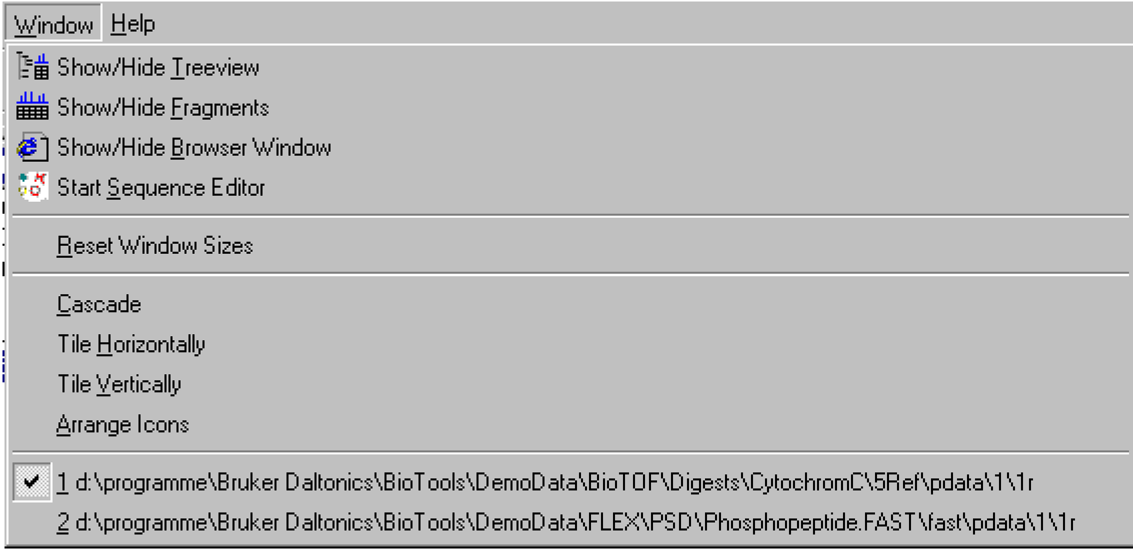


Figure 3-90, List of opened data files, active data file (1) is marked

3.8 Help Menu

With the *Help Menu* the general information and help function are available.

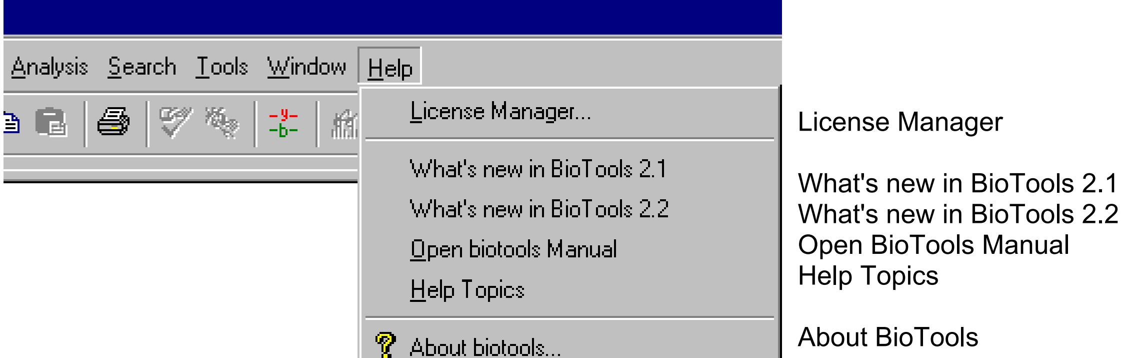


Figure 3-91, Help menu

3.8.1. License Manager

If you start BioTools the first time, you must enter a license key. The key comes together with the BioTools documentation. Use the license manager to add or remove licenses.

If an invalid license key is entered, the program can not be used.

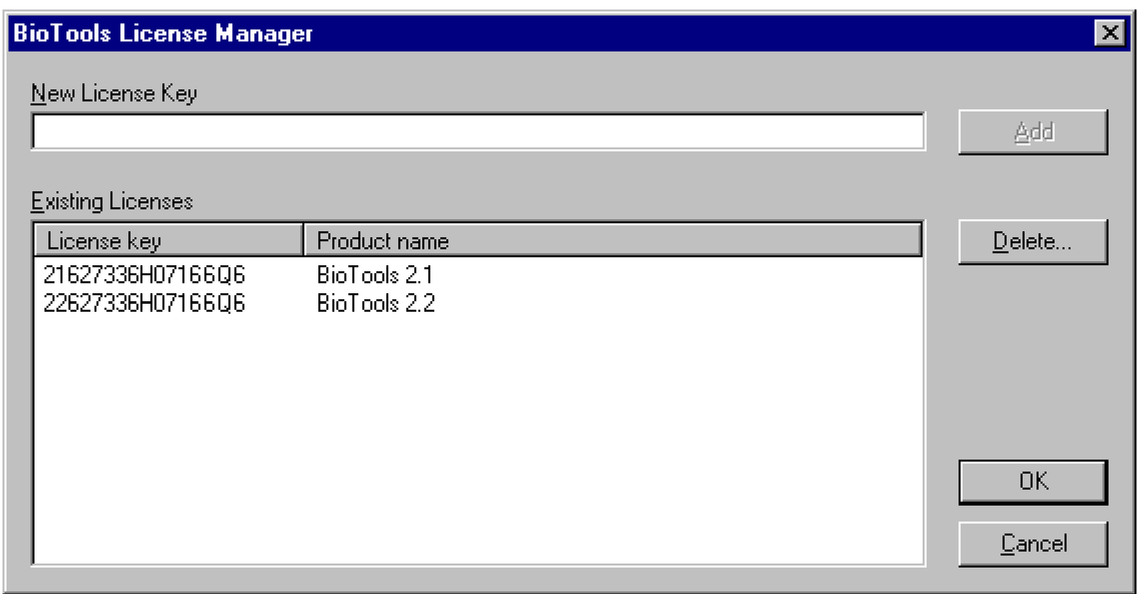


Figure 3-92, License Manager

3.8.2. What's new in BioTools 2.1

Via this option a file is available to get the newest info about the changes since the last version. This file is a powerpoint-file. If necessary the powerpoint-viewer will be proposed for downloading.

3.8.3. What's new in BioTools 2.2

Via this option a file is available to get the newest info about the changes since the last version. This file is a powerpoint-file. If necessary the powerpoint-viewer will be proposed for downloading.

3.8.4. Open BioTools Manual

This command starts the Acrobat Reader and opens the online manual. From here you can access the online tutorials for various applications.

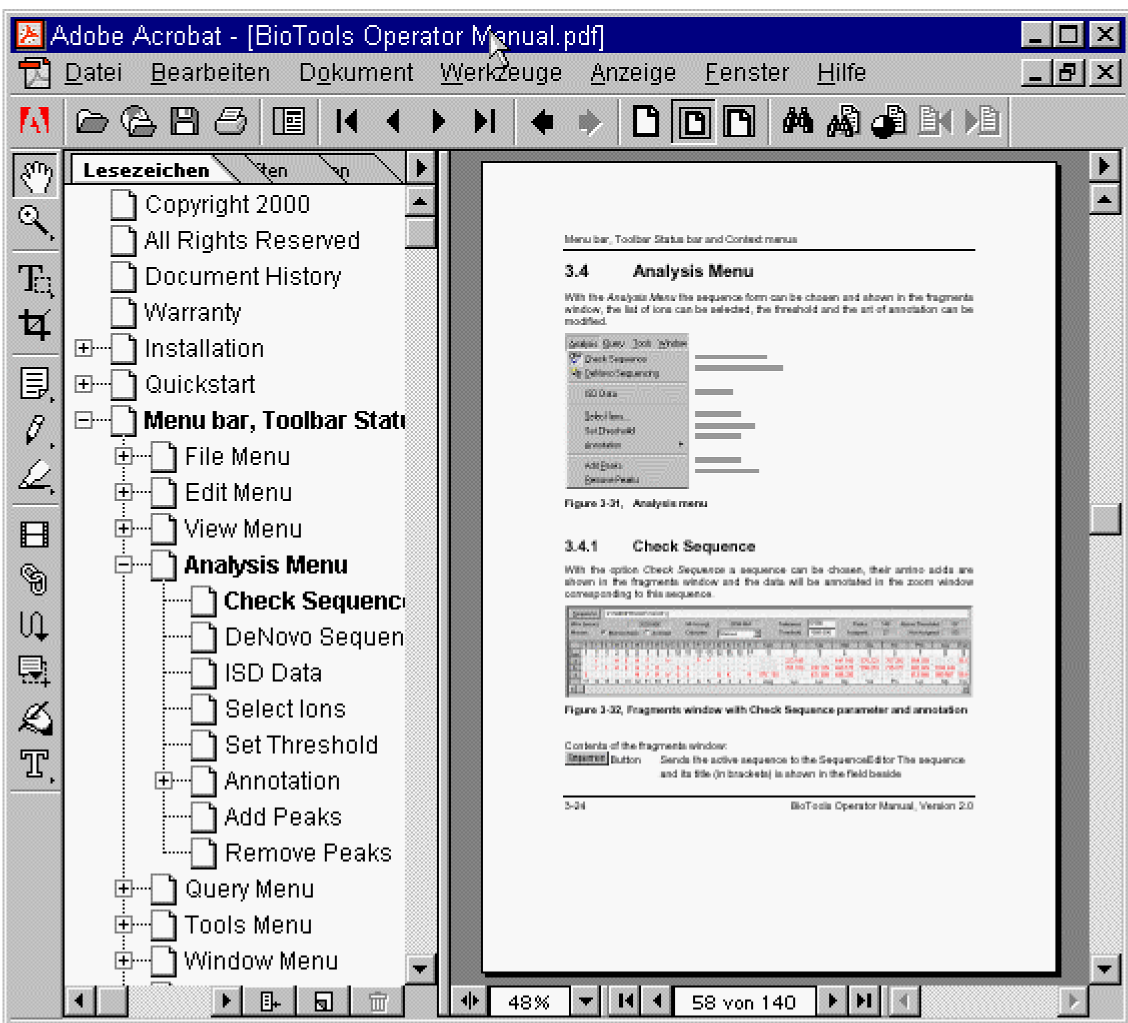


Figure 3-93, BioTools Manual

If the Acrobat Reader program is not properly installed, you will get following message:

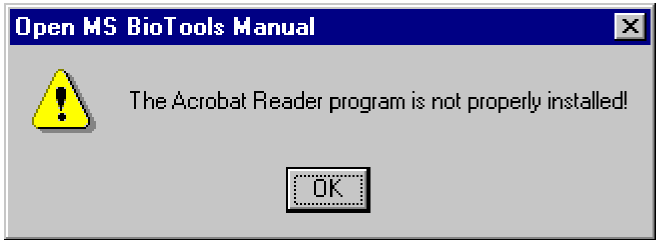


Figure 3-94, The Acrobat Reader is not installed

The Acrobat Reader program can be found on the installation disk of BioTools.

3.8.5. Help Topics

With *Help Topics* the complete online help will be opened with its contents view.

Important note for the use of the User Manual:

The fastest way to use the online help is: move the mouse to the object of interest: a button, menu entry, etc., and press the **F1 key** to obtain the specific help text.

3.8.6. About BioTools

With *About BioTools* the About BioTools window with license information will be opened.

Figure 3-95, About BioTools window

3.9 Menu Bar

The menu bar can be moved with the mouse. Clicking on the background of the menu bar with the left mouse button and move the menu bar with held mouse button to the desired position.

A Appendix: Amino acid Residues and Fragmentations

A.1 Amino Acid Residues

A.1.1. Single letter code

Most of the analysis programs use the single letter code - learn it by heart from: http://alpha2.bmc.uu.se/~kenth/bioinfo/singleletter.html

Name	Three Letter Code	Single Letter Code	Mnemonic
Alanine	Ala	A	(Alanine)
Cysteine	Cys	C	(Cysteine)
Aspartic Acid	Asp	D	(aciD)
Glutamic Acid	Glu	E	(E comes after D)
Phenylalanine	Phe	F	(Ph=F)
Glycine	Gly	G	(Glycine)
Histidine	His	H	(Histidine)
Isoleucine	Ile	I	(Isoleucine)
Lysine	Lys	K	(L follows K)
Leucine	Leu	L	(Leucine)
Methionine	Met	M	(Methionine)
Asparagine	Asn	N	(AsparagiNe)
Proline	Pro	P	(Proline)
Glutamine	Gln	Q	(Qlutamine)
Arginine	Arg	R	(aRginine)
Serine	Ser	S	(Serine)
Threonine	Thr	T	(Threonine)
Valine	Val	V	(Valine)
Tryptophan	Trp	W	(Double ring - W)
Tyrosine	Tyr	Y	(tYrosine)

and:

http://www-lehre.img.bio.uni-goettingen.de/edv/Bio_Inf/MolBiol/intro1.htm

Single Letter Code	Three Letter Code	Mnemonic
A	Ala	**A**lanine
C	Cys	**C**ysteine
D	Asp	aspar**D**ic acid
E	Glu	glu**E**tamic acid
F	Phe	**F**enylalanine
G	Gly	**G**lycine
H	His	**H**istidine
I	Ile	**I**soleucine
K	Lys	before **L**
L	Leu	**L**eucine
M	Met	**M**ethionine
N	Asn	Asparagi**N**e
P	Pro	**P**roline
Q	Gln	**Q**-tamine
R	Arg	a**R**ginine
S	Ser	**S**erine
T	Thr	**T**hreonine
V	Val	**V**aline
W	Trp	t**W**o rings
Y	Tyr	t**Y**rosine

A.1.2. Genetic Code

1.		2.			3.
	T	**C**	**A**	**G**	
T	TTT Phe [F]	TCT Ser [S]	TAT Tyr [Y]	TGT Cys [C]	**T**
	TTC Phe [F]	TCC Ser [S]	TAC Tyr [Y]	TGC Cys [C]	**C**
	TTA Leu [L]	TCA Ser [S]	TAA Ter [end]	TGA Ter [end]	**A**
	TTG Leu [L]	TCG Ser [S]	TAG Ter [end]	TGG Trp [W]	**G**
C	CTT Leu [L]	CCT Pro [P]	CAT His [H]	CGT Arg [R]	**T**
	CTC Leu [L]	CCC Pro [P]	CAC His [H]	CGC Arg [R]	**C**
	CTA Leu [L]	CCA Pro [P]	CAA Gln [Q]	CGA Arg [R]	**A**
	CTG Leu [L]	CCG Pro [P]	CAG Gln [Q]	CGG Arg [R]	**G**
A	ATT Ile [I]	ACT Thr [T]	AAT Asn [N]	AGT Ser [S]	**T**
	ATC Ile [I]	ACC Thr [T]	AAC Asn [N]	AGC Ser [S]	**C**
	ATA Ile [I]	ACA Thr [T]	AAA Lys [K]	AGA Arg [R]	**A**
	ATG Met [M]	ACG Thr [T]	AAG Lys [K]	AGG Arg [R]	**G**
G	GTT Val [V]	GCT Ala [A]	GAT Asp [D]	GGT Gly [G]	**T**
	GTC Val [V]	GCC Ala [A]	GAC Asp [D]	GGC Gly [G]	**C**
	GTA Val [V]	GCA Ala [A]	GAA Glu [E]	GGA Gly [G]	**A**
	GTG Val [V]	GCG Ala [A]	GAG Glu [E]	GGG Gly [G]	**G**

A.1.3. Formulas and Molecular Weights

Table A-1, Amino acid residue masses (mono-isotopic and average) together with 3-and 1-letter code and elemental composition

Name	Symbol	S.	C	H	N	O	S	Monoisotopic Mass	Averaged Mass
Alanine	Ala	A	3	5	1	1	0	71,03712	71,079
Cysteine	Cys	C	3	5	1	1	1	103,00919	103,145
Aspartic acid	Asp	D	4	5	1	3	0	115,02695	115,089
Glutamic acid	Glu	E	5	7	1	3	0	129,0426	129,116
Phenylanaline	Phe	F	9	9	1	1	0	147,06842	147,177
Glycine	Gly	G	2	3	1	1	0	57,02146	57,052
Histidine	His	H	6	7	3	1	0	137,05891	137,141
Isoleucine	Ile	I	6	11	1	1	0	113,08407	113,159
Lysine	Lys	K	6	12	2	1	0	128,09497	128,174
Leucine	Leu	L	6	11	1	1	0	113,08407	113,159
Methionine	Met	M	5	9	1	1	1	131,04049	131,199
Asparagine	Asn	N	4	6	2	2	0	114,04293	114,104
Proline	Pro	P	5	7	1	1	0	97,05277	97,117
Glutamine	Gln	Q	5	8	2	2	0	128,05858	128,131
Arginine	Arg	R	6	12	4	1	0	156,10112	156,188
Serine	Ser	S	3	5	1	2	0	87,03203	87,078
Threonine	Thr	T	4	7	1	2	0	101,04768	101,105
Valine	Val	V	5	9	1	1	0	99,06842	99,133
Tryptophan	Trp	W	11	10	2	1	0	186,07932	186,213
Tyrosine	Tyr	Y	9	9	1	2	0	163,06333	163,176

A.1.4. Chemical Structures

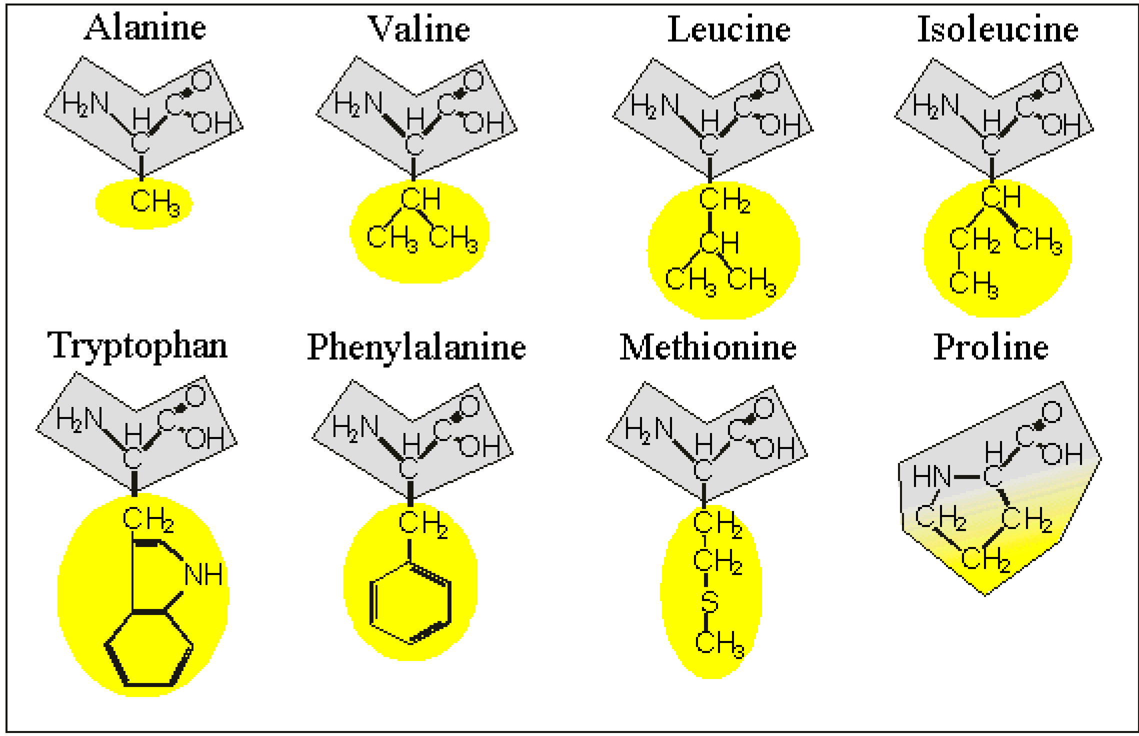


Figure A-1, Neutral hydrophobic Amino acid

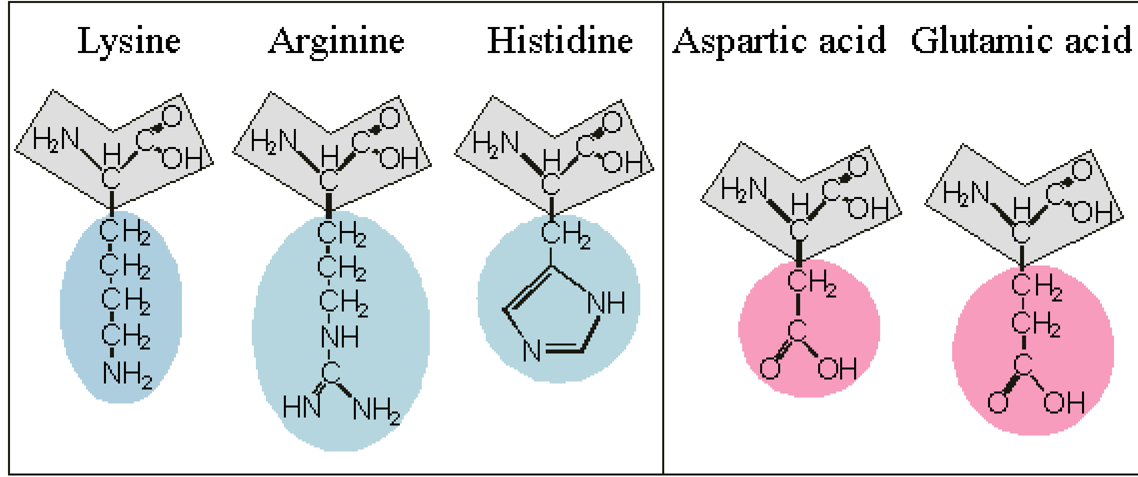


Figure A-2, Basic and acid Amino acids

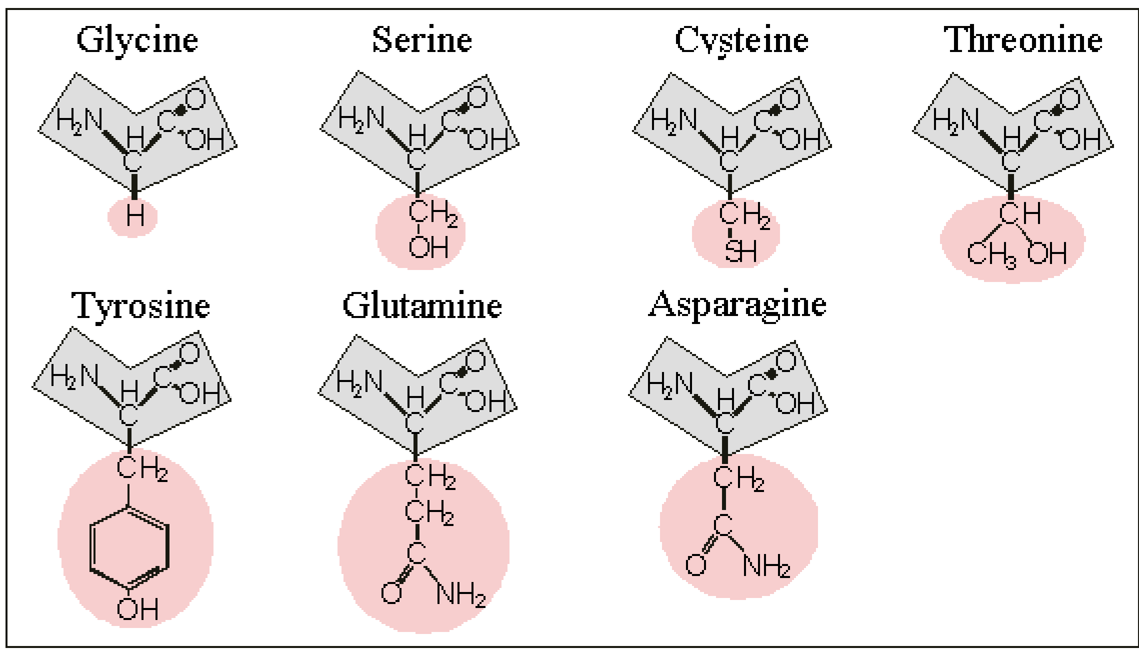


Figure A-3, Neutral polar Amino acids

A.2 Peptide Fragmentation

The types of fragment ions observed in an MS/MS spectrum depend on many factors including primary sequence, the amount of internal energy, how the energy was introduced, charge state, etc. The nomenclature used for fragment ions is the Biemann nomenclature (R. S. Johnson, S. A. Martin & K. Biemann (1988), Int. J. Mass Spec. Ion Procs. 86, 137-154).

Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either *a*, *b* or *c*. If the charge is retained on the C terminal, the ion type is either *x*, *y* or *z*. An index indicates the number of residues in the fragment. In addition to the proton(s) carrying the charge, c ions and y ions abstract an additional proton from the precursor peptide. Note that these structures include a single charge carrying proton. In electrospray ionization, tryptic peptides generally carry two or more charges, so those fragment ions may carry more than one proton.

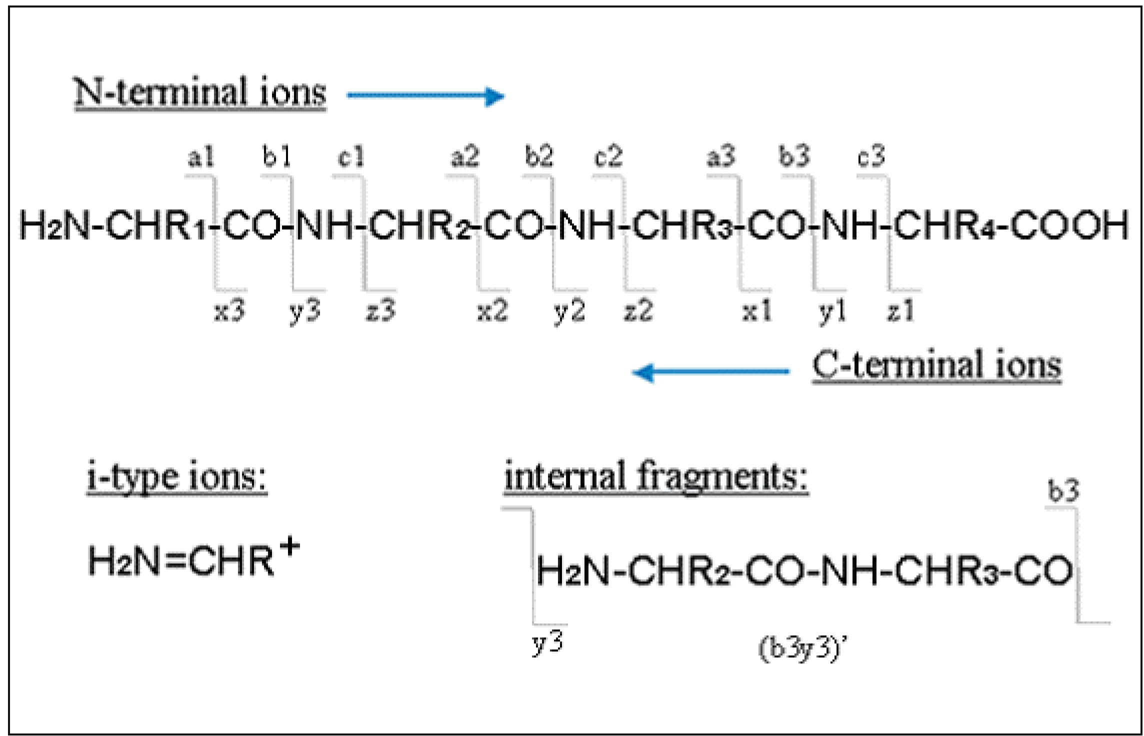


Figure A-4, Peptide Fragmentation

Typical fragment ions observed are:

- Low energy CID: b and y
- PSD: a, b, y and i, including neutral losses of NH_3 from a and b
- ISD: c and y
- ECD-FTICR: c and z

Fragmentation of the backbone at two sites gives rise to internal fragments. Usually, these are formed by a combination of *b*-type and *y*-type cleavage to produce the illustrated structure, amino-acylium ion. Sometimes, internal ions can be formed by a combination of *a*-type and *y*-type digest, an amino-immonium ion.

An internal fragment with just a single side chain formed by a combination of *a* type and *y* type digest is called an immonium ion. The immonium ions can be used for *DeNovo* sequencing. The values from the following table are used to find these ions.

Table 4-2, Immonium and related ion masses

Residue	3-letter	1-letter	Immonium	Related ions
Alanine	Ala	A	44	
Cysteine	Cys	C	76	
Aspartic acid	Asp	D	88	
Glutamic acid	Glu	E	102	
Phenylalanine	Phe	F	120	148
Glycine	Gly	G	30	
Histidine	His	H	110	82 155
Isoleucine	Ile	I	86	44 72
Lysine	Lys	K	101	84
Leucine	Leu	L	86	44 72
Methionine	Met	M	104	60
Asparagine	Asn	N	87	
Proline	Pro	P	70	98
Glutamine	Gln	Q	76	
Arginine	Arg	R	112	100 87 70 60
Serine	Ser	S	60	
Threonine	Thr	T	74	
Valine	Val	V	72	
Tryptophan	Trp	W	159	
Tyrosine	Tyr	Y	136	

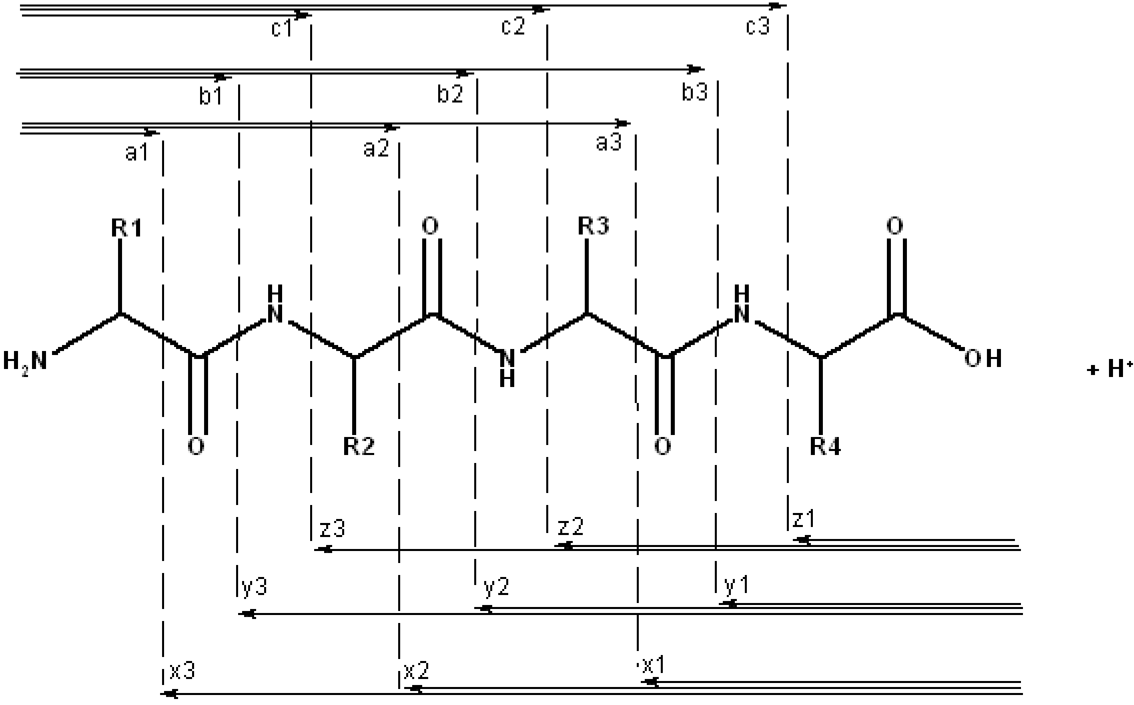


Figure A-5, Fragmentation pattern for peptides

a2

x2

b2

y2

c2

z2

Figure A-6, Structures of the fragments

A.3 Menu and Shortcut list

Toolbar button	*Menu option*	Shortcut F-key	Description
	File *– Find*		Opens a file finder
	File – Open Spectrum	Ctrl + O	Opens a file manager
	File – Multiple open 1r		**All** existing data files in a folder are opened.
	File – Close		Closes the active window
	File – Combine multiple LIFT spectra		Multiple spectra can be combined for use in one data file.
	File – Save		Saves the active data file
	File – Print		Opens the print dialog
	File – Print Preview		Starts the print preview
	File – Print Setup	Ctrl + P	Opens the printer setup dialog
			Prints the active data file immediately in accordance to the Print Setup (page one landscape, other portrait)
	File – Send		Opens an email program to send the active spectrum to an email address as an attached file.
	File – Last used data files		The last used data files are listed here
	File – Exit	ALT + F4	Terminates the program
	Edit *– Undo*	Ctrl + Z	The previous action is undone
	Edit – Cut	Ctrl + X	Deletes and copies into clipboard
	Edit – Copy	Ctrl + C	Copies from clipboard to cursor position
	Edit – Paste	Ctrl + V	Pastes from clipboard to cursor position
	Edit – Sequence…		Starts the SequenceEditor
	View *– Toolbar*		Show or hide the toolbar
	View – Status Bar		Show or hide the status bar
	View – Query Results		Reloads results of a former query
	View – View Fingerprint Results		Changes fragment window to fingerprint view
	View – View MS/MS Results		Changes fragment window to sequence view
	View – Matched Peaks		Displays masses and sequence positions for peaks matching a protein sequence (**black** in general, **blue for peptides containing optional modifications**)

Toolbar button	*Menu option*	Shortcut F-key	Description
	View – Unmatched Peaks		Displays masses in **red** for peaks not matching a protein sequence
	View – Matched and Unmatched Peaks		Displays all peaks
	View – Picked Peaks		Shows the peaks in the spectrum window
	View – Coordinates		Show or hide mass and intensity information within the status bar
	View – Grid		Show or hide a grid within the spectrum window
	View – Scaling		Opens the expand manual dialog or resets to whole spectrum
	View – Zooming		Enables or disables zoom mode
	View – Undo zooming	-	The previous zoom action is undone
	View – Redo zooming	-	The previous zoom action is redone
	View – Distance		Enables or disables the line distance mode
	View – Data Cursor		Enables or disables display of mass and intensity of current peak
	View –Colors		Opens a dialog to change colors
	View – Display Mode		Offers different display modes
	Analysis *– Check Sequence*		Changes the Fragments window to Check Sequence
	Analysis – DeNovo Sequencing		Changes the Fragments window to DeNovo Sequencing
	Analysis – Full DeNovo Sequencing		Opens the DeNovo Sequencing procedure performs a stepwise build-up of the amino acid sequence.
	Analysis – ISD Data		Opens a dialog to select the ions to be calculated
	Analysis – Set Threshold		Defines a threshold (only peaks above the threshold will be taken into account)
-y- -b-	*Analysis – Annotation*		Defines which calculated ions shall be displayed and performs or resets an annotation.
	Analysis – Picked Peaks		Shows the peaks in the spectrum window
	Analysis – Add Peaks		Switches to peak picking mode (max. data point assignment only!).
	Analysis – Remove Peaks		Switches to peak deletion mode.
	Search *– EMBL*		Opens the internet search using EMBL protein identification by peptide masses

Toolbar button	*Menu option*	Shortcut F-key	Description
	Search – Prowl Peptide Mapping		Opens the internet search using Prowl protein identification by peptide mapping
	Search – Mascot Peptide Mass Fingerprint		Opens the Mascot Peptide Mass Fingerprint Query dialog
	Search – Mascot Sequence Query		Opens the Mascot Sequence Query dialog
	Search – Mascot MS/MS Ion Search		Opens the Mascot MS/MS Search dialog
	Search – Mascot Batch	-	Opens the Mascot Batch Mode window
	Search – Search Mass Search SequenceEditor	-	Opens the Search for mass dialog box in SequenceEditor (*Search – Mass Search)*
	Search –Digest SequenceEditor	-	Opens the Perform Digest dialog box in the SequenceEditor (*Search – Perform Digest*)
	Tools *– Authentication*		Opens a dialog to define web access
	Tools – Options		Opens a dialog to define general, display and print options
	Tools – Maldi Spectrum Parameter		Displays basic information about loaded spectrum
	Tools – Modification info		This window shows information about the currently defined modification within the amino acid sequence.
	Tools – Formula parser		This formula parser calculates the mass weight of a chemical formula.
	Tools – Start XTOF NT		Starts the XTOF NT program
	Tools – Start XMASS NT		Starts the XMASS NT program
	Tools – Execute XMASS commands		Executes commands on XMASS running on a remote computer
	Tools – Customize		Customize the toolbar(s) and the menu bar
	Window *– Show/Hide Treeview*		Show or hide the Treeview window
	Window – Show/Hide Fragments		Show or hide the Fragments window
	Window – Show/Hide Browser Window	-	Starts the Internet Browser to get Query results
	Window – Start SequenceEditor		Starts the SequenceEditor
	Window – Reset Window Sizes		The arrangement of the window in BioTools will be set to the default settings

Toolbar button	*Menu option*	Shortcut F-key	Description
	Window – Cascade		Shows all windows cascaded
	Window – Tile Horizontally		Shows all windows tiled horizontally
	Window – Tile Vertically		Shows all windows tiled vertically
	Window – Arrange icons		Arranges all icons on the bottom
	Window – Active data files		All active data files are listed here
	***Help** – License Manager*		Opens the License Manager dialog to verify or set the license keys
	Help – What's new in BioTools 2.1		Opens a presentation with new features in BioTools 2.1
	Help – What's new in BioTools 2.2		Opens a presentation with new features in BioTools 2.2
	Help – Open BioTools Manual		Starts the Acrobat Reader (if installed) and opens the BioTools manual
	Help – Help Topics		Opens the online help
	Help – About BioTools		Opens About BioTools window

A.4 Toolbar Reference list for BioTools

To hide or display the Toolbar, choose menu *View - Toolbar*.

Toolbar button	*Menu option*	Shortcut / Function key	Description
	File – Open Spectrum	Ctrl + O	Opens a file manager
	File – Save	Ctrl + S	Saves the state of the active data file
	Edit – Cut	Ctrl + X	Deletes and copies into clipboard
	Edit – Copy	Ctrl + C	Copies from clipboard to cursor position
	Edit – Paste	Ctrl + V	Pastes from clipboard to cursor position
	-	-	Prints the active data file immediately in accordance to the Print Setup

Toolbar button	*Menu option*	Shortcut / Function key	Description
	Analysis – Check Sequence	-	Changes the fragment window to check sequence mode
	Analysis – DeNovo Sequencing	-	Changes the fragment window to *DeNovo* sequencing mode
	Analysis – Annotation Parameter	-	Opens the annotation options dialog box
	Search – Search Mass Search SequenceEditor	-	Opens the Search for mass dialog box in SequenceEditor (*Search – Mass Search)*
	Search –Digest SequenceEditor	-	Opens the Perform Digest dialog box in the SequenceEditor (*Search – Perform Digest*)
	Search – Mascot Peptide Mass Fingerprint	-	Opens the internet search via Peptide Mass Fingerprint
	Search – Mascot Sequence Query	-	Opens the internet search via Sequence Query
	Search – Mascot MS/MS Ion Search	-	Opens the internet search via MS/MS Ion Search
	Window – Show/Hide Treeview	-	Show or hide the treeview window
	Window – Show/Hide Fragments	-	Show or hide the fragment window
	View – Undo zooming	-	The previous zoom action is undone
	View – Redo zooming	-	The previous zoom action is redone
	Search – Mascot Batch	-	Opens the Mascot Batch Mode window
	Window – Show/Hide Browser Window	-	Starts the Internet Browser to get Query results
	Edit – Sequence…	-	Loads a sequence into the Sequence Editor and starts this program
	Tools – Start XTOF NT	-	Starts the XTOF NT program
	Tools – Start XMASS NT	-	Starts the XMASS NT program
	Help – About BioTools	-	Opens About BioTools window

A.5 Toolbar Reference list for SequenceEditor

To hide or display the Toolbar, choose menu *View - Toolbar*.

Toolbar button	*Menu option*	Shortcut/ Function key	Description
	File – New Sequence	Ctrl + N	Creates a new sequence, also from the web (Sequences from the Web).
	File – Save	Ctrl + S	Saves the active sequence with its current name. If you have not named the sequence, your SequenceEditor displays the Save As dialog box.
	Edit – Cut	Ctrl + X	Removes selected data from the sequence and stores it on the clipboard.
	Edit – Copy	Ctrl + C	Copy the selection to the clipboard.
	Edit – Paste	Ctrl + V	Inserts the contents of the clipboard at the insertion point.
Bio	-		The modified sequence or a marked range of it will be Send data to BioTools for further processing.
	-		Prints the active data file immediately
	Help – About SequenceEditor…		Opens About SequenceEditor window
	-		Activates the context sensitive help.
	File – Print	Ctrl + P	Opens the printer dialogue
	File – Exit	ALT + F4	Terminates the program

A.6 Part Numbers

# 216941	Software-Package BioTools 2.2
# 216935	License BioTools 2.2
# 217032	License RapiDeNovo Sequencing
# 216936	BioTools User Manual

I Index

A

Active Data 3-95
Add Peaks 3-46
Add Profile 3-81
Analysis Menu 3-25
Annotation 3-44, 3-45
Annotation Parameter 3-44
Arrange Icons 3-94
Authentication 3-78, 3-79

B

Background Color 3-23
Batch defaults 3-74
b-y Ion Algorithm 3-27

C

Cascade 3-92
Check Sequence 3-25
Chemical Structures A-V
Close 3-8
Color 3-23, 3-24
Combine multiple LIFT spectra 3-8
Commands tab 3-89
Context Menus 3-1
Coordinates 3-18
Copy 3-13
Cursor 3-20
  Data 3-22
  Distance 3-20, 3-21
  Move 3-20
  Zoom 3-20
Customize 3-86
Cut 3-13

D

Data Cursor 3-22
Data from AutoX 3-68
Data from Bruker DataAnalysis 2-8
Data from Bruker XMASS/XTOF 2-7
Data points 3-24
Delete Profile 3-82
DeNovo - define sequence tag 3-39
DeNovo - result analysis 3-41
DeNovo advanced settings 3-39
DeNovo Sequencing 3-30
DeNovo Sequencing Processing 2-10
DeNovo settings details 3-38
DeNovo settings smart 3-37
Digest (SequenceEditor) 3-76
Display Mode 3-24
Display of Picked Peaks 2-9
Distance Cursor 3-21

E

Edit Data 3-70
Edit Menu 3-13
Edit Parameter 3-72
EMBL 3-52
Execute XMASS Commands 3-85
Exit 3-12

F

File Menu 3-4
Find 3-5
Formula and Masses A-IV
Formula Parser 3-84
Full DeNovo Sequencing 3-35

G
Grid 3-18

H
Help Menu 3-96
Help Topics 3-99

I
Installation 1-1
ISD Data 3-42

L
Last Used Data 3-12
License Manager 1-8, 3-97
Loading processed data 2-2
Loading Spectra 2-7

M
Maldi Spectrum Parameter 3-82
Mascot 3-54
Mascot Batch Searches 3-59
Mascot Database Query Processing (MS/MS) 2-14
Mascot MS/MS Ion Search 3-57
Mascot Peptide Mass Fingerprint 3-54
Mascot Sequence Query 3-56
Matched and Unmatched Peaks 3-18
Matched Peaks 3-18
Menu A-XI, A-XIII
  Analysis 3-25
  Edit 3-13
Menu and Shortcut list A-XI
Menu Bar 3-2, 3-100
  Toolbar

    Status Bar and Context Menus 3-1
menu structure A-XI
Modification Info 3-83
Multiple Batch 3-62
Multiple open 1r 3-8

N
New Profile 3-81
New Toolbar dialog box 3-88

O
Open 3-7
Open BioTools Manual 3-98
Open Spectrum 3-7
Options 3-81

P
Paste 3-13
Peak Picking 3-46
Peptide Fragmentation A-VII
Picked Peaks 3-18
Print 3-9
Print Preview 3-10
Print Setup 3-11
Processing of MS Data 2-10
Program Setup 1-1
PROWL Peptide Mapping 3-53

Q
Query Results 3-17
Quickstart 2-1

R
Redo Zooming 3-20
reference list A-XI
Remove Peaks 3-46
Reset Window Sizes 3-91

S
Save 3-9
Scaling 3-19
Search for Masses (SequenceEditor) 3-75
Search Menu 3-49
Send 3-12
Send data 3-12
Sequence 3-13
Set Threshold 3-43
shortcut A-XI
Show/Hide Browser Window 3-91
Show/Hide Fragments 3-90
Show/Hide Treeview 3-90
single letter code A-I
SNAP 3-47
Spectrum Color 3-23
Start Sequence Editor 3-91
Start SequenceEditor 3-91
Start XMASS NT 3-84
Start XTOF NT 3-84
Starting the Installation 1-1
Starting the Program 1-8
Status Bar 3-1, 3-3, 3-16
sum 3-48
Sum Peak Finder 3-48
System Requirements 1-1

T
Tab Mail parameter 3-74
Tab Query Parameter 3-72
Tab Scout MTP 3-62
Tab Status 3-60
Tab Summary Report 3-70
Tab Task Editor 3-65
Table of changes vi
Tile Horizontally 3-92
Tile Vertically 3-93
Toolbar A-XI, A-XIV, 3-1, 3-2, 3-3, 3-15
toolbar buttons A-XI
Toolbar Reference list A-XIV
Toolbar Reference list for BioTools A-XIV
Toolbars tab 3-87
Tools Menu 3-78

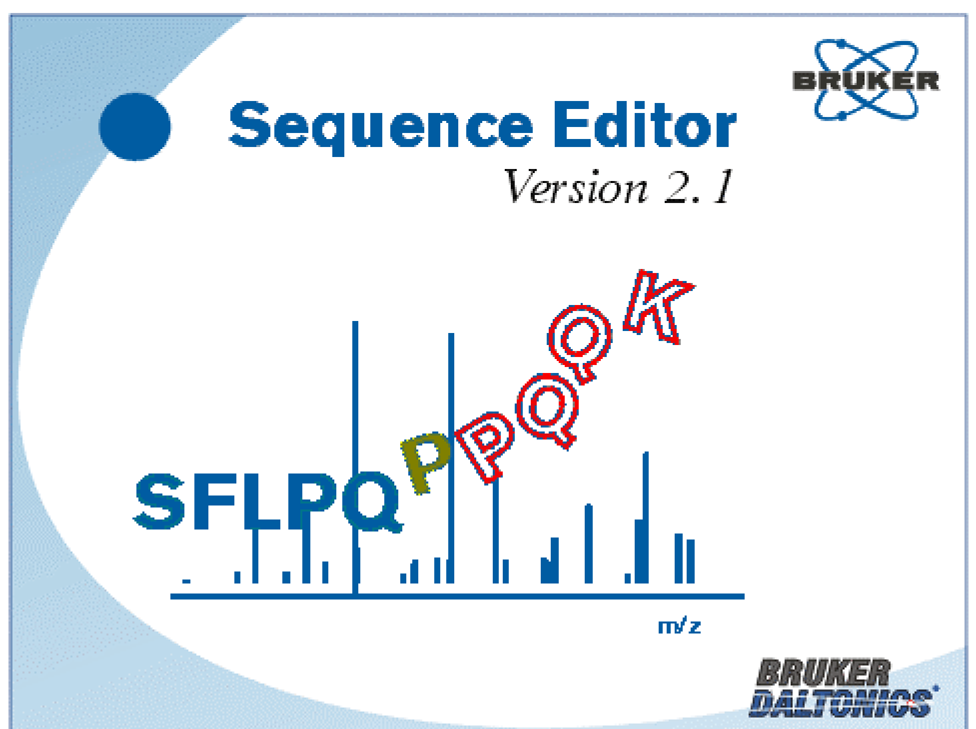
U
Undo 3-13
Undo Zooming 3-20
Unmatched Peaks 3-18

V
View Fingerprint Results 3-18
View Menu 3-14
View MS/MS Results 3-18

W
What s new in BioTools 2.1 3-97
What s new in BioTools 2.2 3-97
Window Menu 3-90

X
XTOF Unix 2-7
XTOF Windows NT 2-7

SequenceEditor Reference Manual



(August 2002)

Copyright

Copyright 2000

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

SequenceEditor Reference Manual, Version 2.2 (August 2002)

First edition: October 2000
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.

Copyright: **Bruker Daltonik GmbH**
Fahrenheitstrasse 4
28359 Bremen
Germany

Phone: +49 (4 21) 22 05-200
FAX: +49 (4 21) 22 05-103
Email: sales@bdal.de
Internet: www.bdal.de

Table of changes

Version	*Date*	*Changes*
2.0	2000-10-14	Software versions: BioTools, Version 2.0 SequenceEditor, Version 1.0
2.1	2002-01-18	Software versions: BioTools, Version 2.1 SequenceEditor, Version 2.1
2.2	2002-08-31	Software versions: BioTools, Version 2.2 SequenceEditor, Version 2.1

How to use this Manual and the Help Functionality within BioTools?

The **User Manuals** are meant to be online help documents, which can be accessed using the **F1 key** at any point of entry of the program. In addition they can be read like a book, which is not recommended whatsoever.

The two **Tutorial blocks** with focus on either the SequenceEditor or BioTools contain **procedure descriptions for specific analytical problems**. Work your way through those applications, which you want to learn about and use the F1 access to online help only where needed.

The complete information is given in the manual in the program directory and is accessible via the menu *Help – Open BioTools Manual* in BioTools software.

Contents

0 SEQUENCEEDITOR WINDOW 0-1
0.1 Menu bar 0-1
0.2 Toolbar 0-2
0.2.1 Send data to BioTools 0-3
0.2.2 Optional Modification Transfer Dialog 0-3
0.2.3 Sequences from the Web 0-4
0.3 Tree structure 0-4
0.4 Sequence Toolbar 0-5
0.4.1 Total mass 0-5
0.4.2 Mono-Ave-Button 0-5
0.4.3 Amino mass 0-5
0.4.4 Range field 0-5
0.4.5 Mass Search Button 0-6
0.4.6 Perform Enzymatic Digest Button 0-6
0.4.7 Building block 0-6
0.4.8 Cysteine state 0-6
0.4.9 N-terminal 0-7
0.4.10 C-terminal 0-7
0.4.11 Sequence title 0-7
0.5 Sequence area 0-8
0.6 Scrollbar 0-8
0.7 Inactive cys-disulfide crosslinks 0-9
0.8 Status Bar 0-10

1 FILE MENU 1-1
1.1 New sequence 1-2
1.2 Close command 1-2
1.3 Save command 1-2
1.4 Save As command 1-3
1.4.1 File name 1-3
1.4.2 Save in 1-3
1.4.3 Files list 1-3
1.4.4 Save as type 1-3
1.4.5 Default files 1-4
1.4.6 Altering the default files 1-4
1.5 Print command 1-5
1.6 Print Preview command 1-6
1.7 Print Setup command 1-7

1.8 Import sequence ... 1-8
1.9 Export FastA ... 1-8
1.10 Load LCMS results ... 1-8
1.11 1, 2, 3, 4 command ... 1-8
1.12 Exit command ... 1-8

2 EDIT MENU ... 2-1
2.1 Undo ... 2-1
2.2 Cut ... 2-1
2.3 Copy ... 2-2
2.4 Paste ... 2-2
2.5 Cysteine state ... 2-2
2.6 Crosslinks ... 2-3
2.6.1 Current Crosslink File ... 2-4
2.6.2 List of Crosslinks ... 2-4
2.6.3 Delete a Crosslink ... 2-4
2.6.4 Conflicts information ... 2-4
2.6.5 New Link Definition ... 2-4
2.7 Crosslink-Types ... 2-6
2.7.1 Current Crosslink File ... 2-7
2.7.2 Crosslink Definition ... 2-7
2.7.3 Load a crosslink file ... 2-8
2.7.4 Save a crosslink file ... 2-8
2.7.5 Save a new crosslink file ... 2-9
2.8 Modifications ... 2-10
2.8.1 Current Modification File ... 2-11
2.8.2 Modifications Profile ... 2-12
2.8.3 Kinds of Modifications ... 2-14
2.8.4 Define new modification ... 2-14
2.9 Modif.-Types ... 2-16
2.9.1 Current Modification File ... 2-17
2.9.2 Modification name ... 2-17
2.9.3 Modifications list ... 2-17
2.9.4 Chemistry ... 2-17
2.9.5 Specifity ... 2-18
2.9.6 Applicability ... 2-18
2.9.7 Add modification type ... 2-18
2.9.8 Replace modification type ... 2-18
2.9.9 Delete modification type ... 2-18
2.9.10 Default modification type ... 2-18
2.9.11 Set default ... 2-18
2.9.12 Modification File ... 2-19

3 VIEW MENU 3-1
3.1 Toolbar 3-2
3.2 Status Bar 3-3
3.3 Mono - Ave 3-4
3.4 Show residues 3-4
3.5 Show all types 3-5
3.6 Hide all types 3-5
3.7 Edit amino acids 3-5
3.7.1 One-letter code 3-6
3.7.2 Three-letter code 3-6
3.7.3 Amino acids list 3-6
3.7.4 Chemical composition 3-6
3.7.5 Amino acid mass 3-6
3.7.6 Print Amino Acids Table 3-6

4 SEQINFO MENU 4-1
4.1 Chain names 4-1
4.2 Sample info 4-1
4.3 Isotopic pattern 4-1
4.4 Data Base entry text 4-1
4.5 Charge States 4-2
4.5.1 Element composition 4-2
4.5.2 Total masses 4-2
4.5.3 Calculation Parameters 4-3
4.5.4 Calculated Masses 4-3
4.5.5 Copy button 4-3
4.5.6 Copy Charge States 4-4
4.5.7 Print button 4-5
4.6 Composition 4-6

5 SEARCH MENU 5-1
5.1 Mass search 5-1
5.1.1 Search Parameters 5-2
5.1.2 Mass Search Results 5-6
5.1.3 Export theoretical digest results dialog 5-7
5.2 Sequence 5-8

6 DIGEST MENU 6-1
6.1 Perform digest 6-1
6.1.1 Digest parameters 6-3
6.1.2 Digest Results 6-5
6.1.3 To spectrum 6-6
6.1.4 Print 6-7
6.1.5 Export theoretical digest results dialog 6-7

6.2 Edit enzymes 6-8
6.2.1 Enzyme name 6-9
6.2.2 Digest rules 6-9
6.2.3 Enzyme Definition 6-9
6.2.4 Load an enzyme file 6-11
6.2.5 Save an enzyme file 6-12
6.2.6 Save as a new enzyme file 6-12

7 OPTIONS MENU 7-1
7.1 Colors 7-1
7.2 System 7-4
7.3 Default Building Blocks 7-4

8 TOOLS MENU 8-1
8.1 Customize 8-1
8.1.1 Prompt for New Composition 8-1
8.1.2 Prompt for save 8-1
8.1.3 Prompt for exit 8-1
8.1.4 Send BioTools selection message 8-2

9 WINDOW MENU 9-1
9.1 New window 9-1
9.2 Cascade windows 9-1
9.3 Tile windows 9-1
9.4 Arrange Icons 9-2
9.5 List of active Data Files 9-2

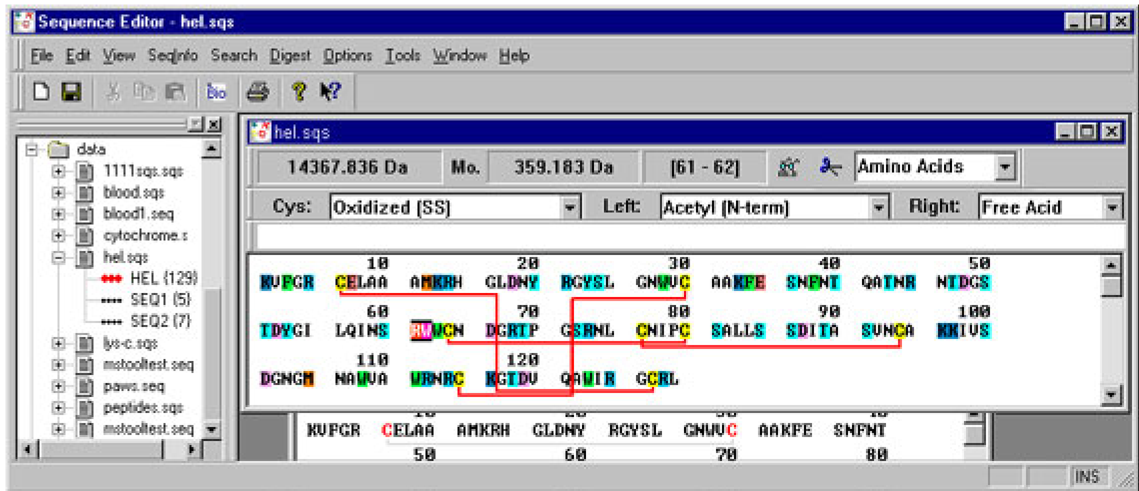
10 HELP MENU 10-1
10.1 Context sensitive Help 10-1
10.2 About SequenceEditor 10-2
10.3 Help Topics 10-2

11 SYSTEM MENU 11-1
11.1 Restore command 11-1
11.2 Move command 11-1
11.3 Size command 11-1
11.4 Minimize Window 11-2
11.5 Maximize command 11-2
11.6 Close command 11-2

I INDEX 11-1

0 SequenceEditor window

The SequenceEditor window contains the following fields and buttons:



Multiple sequences in multiple windows are supported. Displaying of other sequences is performed using the tree structure situated on the left side of the screen.

The toolbar situated at the top of the window shows the total molecular weight of the sequence, the weight of a residue the cursor is upon (or of the selected group of residues) and the range of the residue (range of the selected group). These values are switchable between monoisotopic and average modes.

The cysteine state and the N- and C-terminals are shown in the toolbar and can be dynamically changed by means of the combo boxes placed in the toolbar. The last information supported is the title of the amino sequence being inspected.

0.1 Menu bar

The menu bar is displayed at the top of the main application window and is accessible to all sequence windows currently existing. The menu bar provides mouse access to the SequenceEditor menus.

0.2 Toolbar

The toolbar is displayed across the top of the main application window, below the menu bar and is accessible to all sequence windows currently existing. The toolbar provides quick mouse access to many tools used in SequenceEditor.

To hide or display the Toolbar, choose *View menu - Toolbar*.

Toolbar button	*Menu option*	Shortcut F-key	*Description*
	File – New Sequence	Ctrl + N	Creates a new sequence, also from the web (Sequences from the Web).
	File – Save	Ctrl + S	Saves the active sequence with its current name. If you have not named the sequence, your SequenceEditor displays the Save As dialog box.
	Edit – Cut	Ctrl + X	Removes selected data from the sequence and stores it on the clipboard.
	Edit – Copy	Ctrl + C	Copy the selection to the clipboard.
	Edit – Paste	Ctrl + V	Inserts the contents of the clipboard at the insertion point.
	-		The modified sequence or a marked range of it will be Send data to BioTools for further processing.
	-		Prints the active data file immediately
	Help – About SequenceEditor…		Opens About SequenceEditor window
	-		Activates the context sensitive help.
	File – Print	Ctrl + P	Opens the printer dialogue
	File – Exit	ALT + F4	Terminates the program

0.2.1 Send data to BioTools

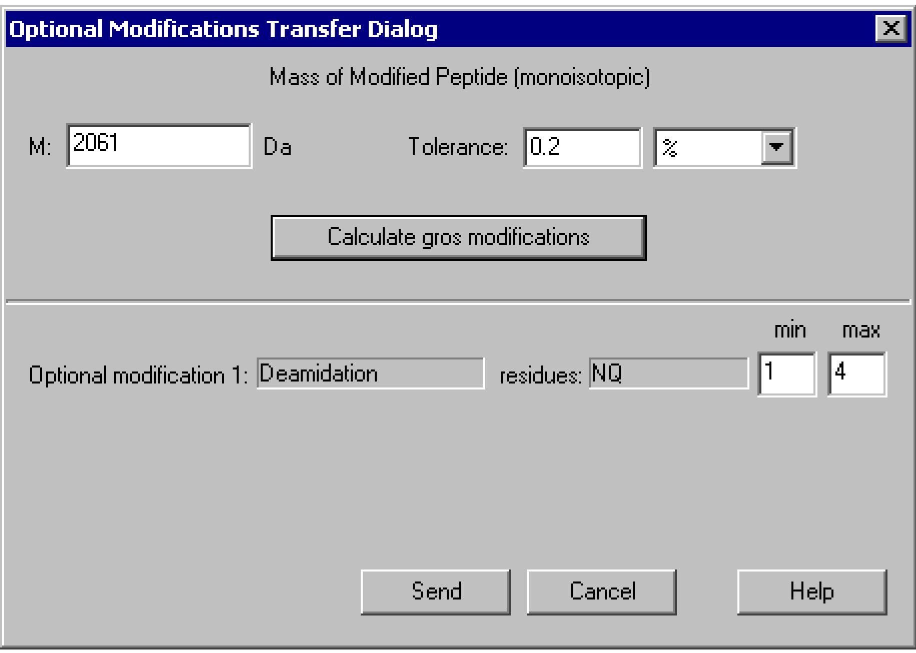
The sequence will be sent to BioTools for further processing.

If in the menu *Tools - Customize* the option **Send BioTools selection message** is selected the Send BioTools selection message will appear if only a part of the sequence is selected.

If the sequence(part) includes optional modifications the transfer dialog Optional Modification Transfer Dialog appears. There the parameters "min" and "max" of optional modifications can be set. Clicking the button Send will send the data to BioTools.

0.2.2 Optional Modification Transfer Dialog

Click the "BioTools" button to transfer the changed modification information back to BioTools.



To get a proposal how many optional modifications are in the sequence(part) enter a mass value and the Tolerance, then click the button Calculate gros modifications.

After calculating, the parameters "min" and "max" are set with possible values, other than zero.

These values are proposals, if desired, change the parameters "min" and "max" manually. Be careful setting these parameters not to get too much entries in BioTools.

Example: The sequence(part) contains 5 optional modifications, after calculating the parameters min = 1 and max = 4, set them to 1 and 2. This will result in 15 sequences send to BioTools (5 x one opt. modif. and 10 x two opt. modif.)!

Use the button Send to transfer the data to BioTools.

0.2.3 Sequences from the Web

Select the sequence text in the web page. Odd characters, numbers or letters, which do not code for the standard amino acids will be automatically excluded during pasting into the SequenceEditor.

Exceptions to the rule are these letters:

- B codes for D or N; is read as N with optional modification (Deamidation)
- Z codes for E or Q; is read as Q with optional modification (Deamidation)
- U codes for selenocystein; is read as C with modification (Selenocystein)
- X codes for unknown or other residue. It requires manual investigation of the database entry text, to properly define this residue in the SequenceEditor.

Reference: IUPAC definitions for non-standard amino acids in SequenceEditor, BioTools and Protein Search from:
http://www.chem.qmw.ac.uk/iupac/AminoAcid/A2021.html

0.3 Tree structure

In each data file can reside several amino sequences. Clicking on the respective symbol in the tree structure the contents of the file reveals. Double click on the selected sequence results in the appearance of the sequence in the display area of SequenceEditor. The simultaneous read operation of several sequences residing in the same file is possible. Normally, sequences located in files with extension ".SQS" (Bruker format) can be read in. However, the SequenceEditor supports loading and processing of GPMAW-format files (extension ".SEQ"), too. If saved, the new file becomes the extension ".SQS".

0.4 Sequence Toolbar

Along to the application toolbar accessible from each sequence window, each sequence becomes a separate sequence toolbar:

To hide or display the first line of the toolbar, choose *View menu - Toolbar*.

0.4.1 Total mass

In case of any relevant change to the amino sequence, the total mass is repeatedly calculated. The calculation considers the raw mass of each amino acid, the current cysteine state and termini and all the modifications and crosslinks defined within the sequence. The mass computation is performed both for monoisotopic and average modes. The updated total mass is shown in the application toolbar.

0.4.2 Mono-Ave-Button

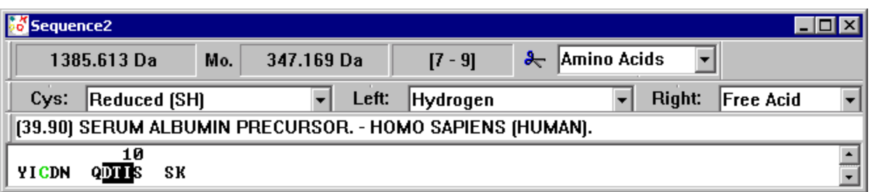
Pressing this button Mo. or using View – Mono/Ave, the mass display toggles between monoisotopic and average modes.

0.4.3 Amino mass

The mass of the amino acid or of the selected group of residues is calculated and shown in the toolbar. The calculation considers the raw mass of the amino acid, eventually the current cysteine state and termini and the modifications and crosslinks defined within the residue or the residue group. The mass computation is performed both for monoisotopic and average modes.

0.4.4 Range field

The position of the amino acid the cursor is currently upon, or the range of the selected and marked residue group is shown in the range field.



0.4.5 Mass Search Button

Use the button to start a Mass search.

0.4.6 Perform Enzymatic Digest Button

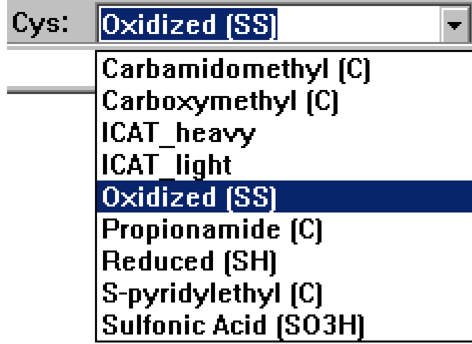
Use the button to perform an enzymatic digest.

0.4.7 Building block

In this drop-down list you can choose whether all amino acids or only these one of the DNA (T G C A) should be enabled. How to set the default is described in the menu *Options – Default building blocks* (Default Building Blocks).

0.4.8 Cysteine state

The combo box shown beneath is located in the toolbar of the SequenceEditor.



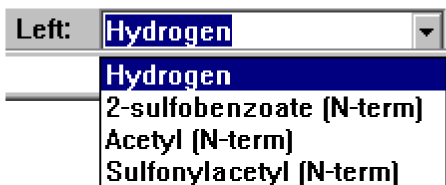
At the start of the program all cysteine states defined in the system are put into this combo box. The box is sorted alphabetically but the default cysteine state (set in the Modif.-Types) is always appearing on the top of the list. The selection of the current state can be performed dynamically. Each change of the cysteine state results in a new total mass computation of the amino sequence.

Selecting the "Oxidized" cysteine state while there are still free cysteine residues in the sequence, results in the Crosslink Editor dialog box appearing on the screen. The crosslink editor has only *Cys* residues and *Cys-disulfide* link type enabled. This Crosslink Editor appears when entering the oxidized cysteine state and any cysteine residues are still free.

Changing the cysteine state from *oxidized* to another one and when cys-disulfide crosslinks were defined, does not result in disappearing of the links. Instead, they will be shown in a fine gray color and become inactive (not considered while mass computation, etc.). The crosslinks can be restored if the state is again *oxidized*.

0.4.9 N-terminal

The combo box shown below is located in the toolbar of the SequenceEditor.



At the start of the program all N-terminal modifications in the system are put into this combo box. The box is sorted alphabetically but the default modification (set in the Modif.-Types) always appears on the top of the list. The selection of the current N-terminal modification can be performed dynamically. Each change of the modification results in a new total mass computation of the amino sequence.

0.4.10 C-terminal

The combo box shown below is located in the toolbar of the SequenceEditor.

At the start of the program all C-terminal modifications in the system are put into this combo box. The box is sorted alphabetically but the default modification (set in the Modif.-Types) always appears on the top of the list. The selection of the current N-terminal modification can be performed dynamically. Each change of the modification results in a new total mass computation of the amino sequence.

0.4.11 Sequence title

A single data file (sqs or seq) can contain several individual amino sequences. In order to distinguish the sequences you can use their titles. The titles can be specified within the sequence title field and will be later saved along with the sequence data.

0.5 Sequence area

Sequences are displayed in blocks of five residues and numbered in blocks of 10. Automatic line wrapping after each 10-group is implemented. The vertical scroll bar situated to the left of the sequence area is used to scroll the contents if it does not match into the window.

Modified residues appear in an another color that can be specified by the user. Crosslinks between residues are shown as colored lines (the color is defined by the user).

A cursor appears within the sequence area. Typing amino acid codes results in insert/overwrite the new entered residues into the sequence. The residues can also be deleted, cut, copied and pasted. The most changes made to the sequence can be easily undone.

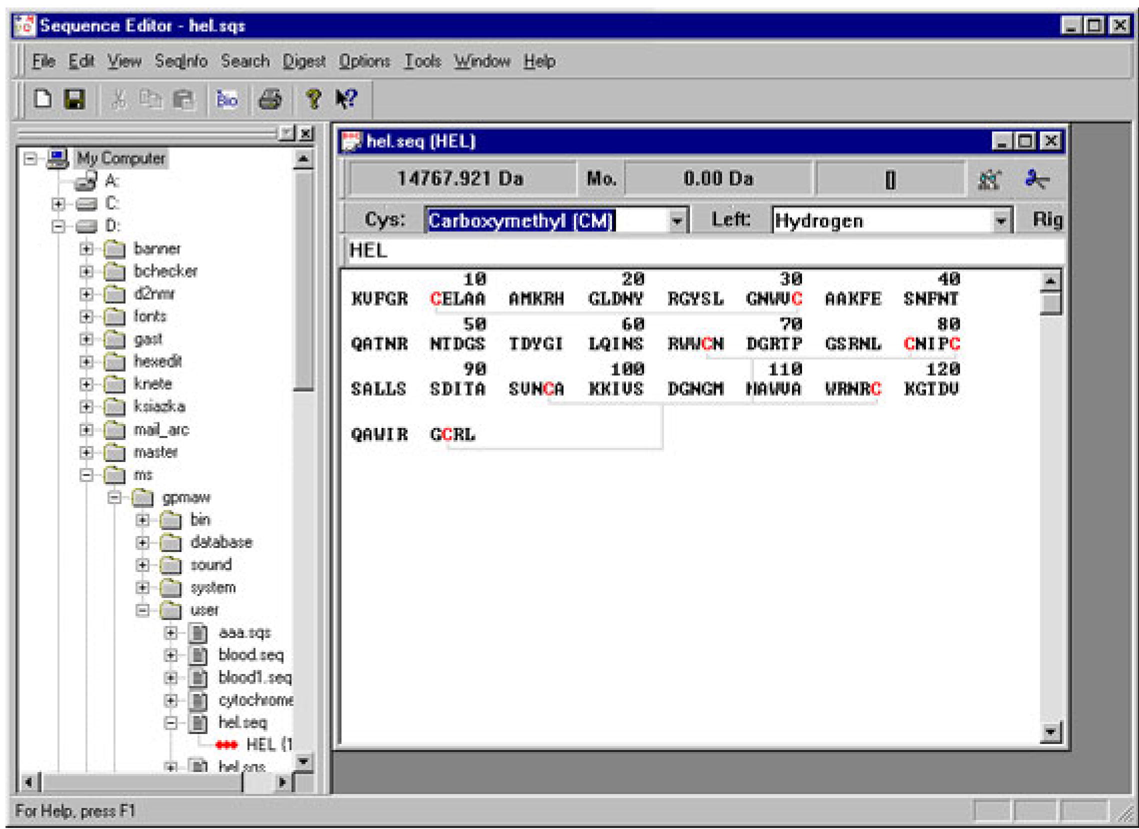
The total mass, the mass of the residue beneath the cursor (or of the selected group) and the residue number (or group range) are shown in the toolbar attached to each data window. Information like cysteine-state, N- and C-terminals appear in the toolbar.

0.6 Scrollbar

Displayed at the right edge of the document window. The scroll box inside the scroll bar indicates your vertical location in the document. You can use the mouse to scroll to other parts of the sequence.

0.7 Inactive cys-disulfide crosslinks

Entering a cysteine state different from *oxidized* and when any cys-disulfide crosslinks are defined, the crosslinks do not disappear but are shown in a fine gray color and deactivated.



Any operation considering the crosslinks (total mass computing, crosslinks definition, etc.) will be performed disregarding the inactive crosslinks. The links can be restored when returning to the cysteine state *oxidized*.

0.8 Status Bar

Use this command to display and hide the Status Bar, which describes the action to be executed by the selected menu item or depressed toolbar button, and keyboard latch state. A check mark appears next to the menu item when the Status Bar is displayed.

The status bar is displayed at the bottom of the SequenceEditor window. To display or hide the status bar, use the Status Bar command in the View menu.

If the mouse cursor is placed over a modified residue then on the left side is displayed information of local and global modifications of that residue, but no optional modifications will be shown.

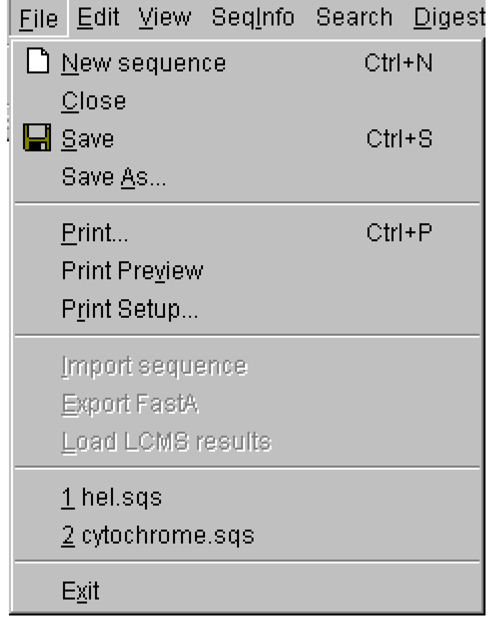
The right areas of the status bar indicate which of the following keys are latched down:

Indicator	Description
CAP	Caps lock key is activated.
NUM	NUM key is activated.

INS/OVR Toggles between INS mode and OVR mode. Toggle with the INS/OVR key on the keyboard between the insert and overwrite mode of the cursor.

1 File menu

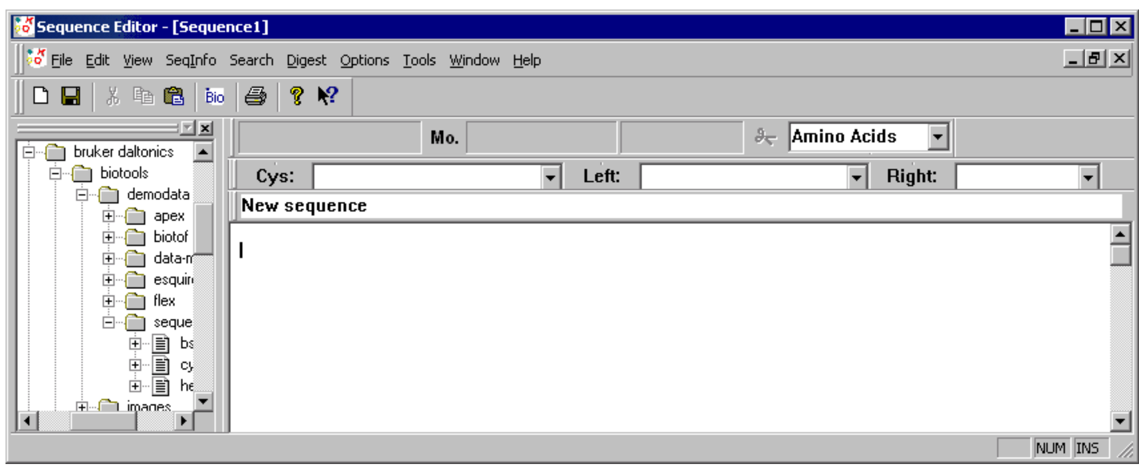
The File menu contains the following commands:



New sequence	Creates a new empty sequence.
Close	Closes an opened sequence.
Save	Saves an opened sequence using the same file name.
Save As	Saves an opened sequence to a specified file name.
Print	Prints a sequence.
Print Preview	Displays the sequence on the screen as it would appear printed.
Print Setup	Selects a printer and printer connection.
Import sequence	NOT YET IMPLEMENTED!
Export FastA	NOT YET IMPLEMENTED!
Load LCMS results	NOT YET IMPLEMENTED!
1, 2, 3, 4 commands	List of the last used files.
Exit	Exits the SequenceEditor.

1.1 New sequence

A new sequence window will be opened.



The default file title is set to "Sequence1", the cysteine state to "Reduced (SH)" and the N-terminal and C-terminal to "Hydrogen" and "Free Acid" respectively. A new sequence can be entered manually.

1.2 Close command

Use this command to close all windows containing the active sequence. SequenceEditor sugests that you save changes to your sequence before you close it. If you close a sequence without saving, you lose all changes made since the last time you saved it.

You can also close a sequence by using the Close icon on the document's window, as shown below:

1.3 Save command

Use this command to save the active sequence to its current name and directory. When you save a sequence for the first time, SequenceEditor displays the Save As dialog box so you can name your document. If you want to change the name and directory of an existing sequence before you save it, choose the Save As command.

Shortcuts

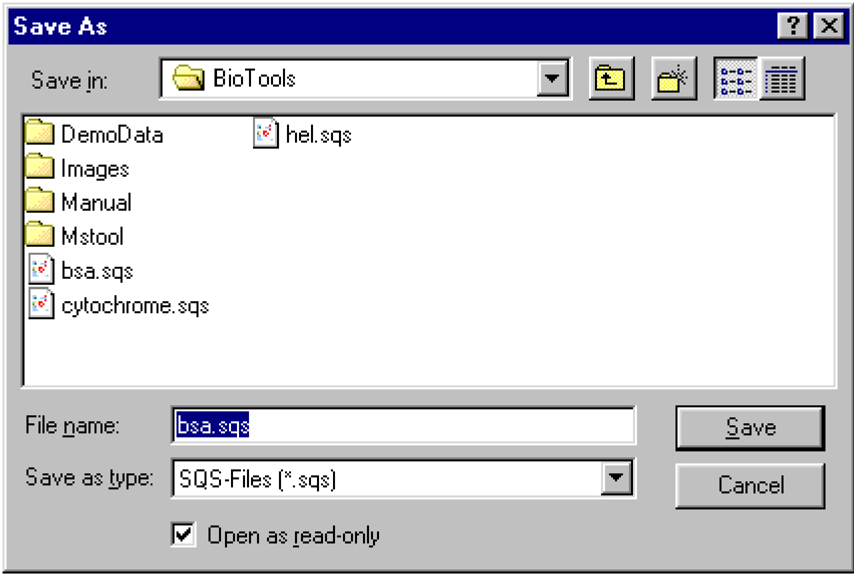
Toolbar:

Keys: CTRL+S

1.4 Save As command

Use this command to save and name the active sequence. SequenceEditor displays the *Save As* dialog box so you can name your document.

To save a sequence with its existing name and directory, use the Save command.



1.4.1 File name

The name of the file the sequence should be saved into.

1.4.2 Save in

Select the drive and the directory in which you want to store the sequence.

1.4.3 Files list

The list of data files existing in the chosen directory.

1.4.4 Save as type

The files list contains only the files with the extension of the chosen type. The save operation stores the sequence in a file with the chosen extension.

1.4.5 Default files

The SequenceEditor is delivered along with a set of default files. All of them have the name *BrukerDefault* and can not be altered within the program.

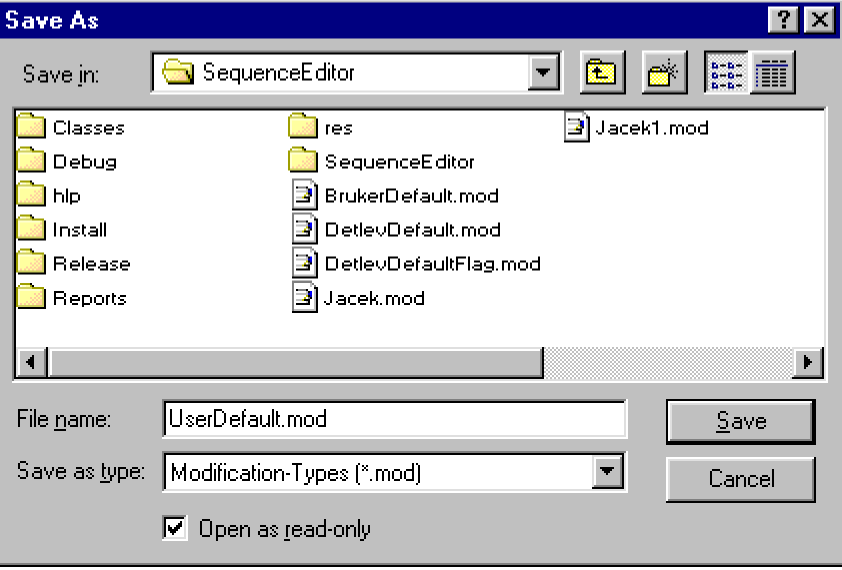
The default files are:

BrukerDefault.mod	for modification definitions,
BrukerDefault.clk	for cross-link definitions,
BrukerDefault.cns	for consensus definitions and
BrukerDefault.enz	for enzyme definitions.

Based on these files, a user defined file can be edited and used with the system.

1.4.6 Altering the default files

The default files delivered with the SequenceEditor can not be altered within the system. In the respective editor dialog box the 'Save' button is grayed and disabled. To create a set of user sepecified files, save the respective files under a new name. In the following example the default modification file *BrukerDefault.mod* is saved under the name *UserDefault.mod*.



1.5 Print command

Use this command to print a document. This command presents a Print dialog box, where you may specify the range of pages to be printed, the number of copies, the destination printer, and other printer setup options.

Shortcuts

Toolbar:

Keys: CTRL+P

The following options allow you to specify how the document should be printed:

Printer

This is the active printer and printer connection. Choose the Setup option to change the printer and printer connection.

Setup

Displays a Print dialog box, so you can select a printer and printer connection.

Print Range

Specify the pages you want to print:

All Prints the entire document.

Selection Prints the currently selected text.

Pages Prints the range of pages you specify in the From and To boxes.

Copies

Specify the number of copies you want to print for the above page range.

Collate Copies

Prints copies in page number order, instead of separated multiple copies of each page.

Print Quality

Select the quality of the printing. Generally, lower quality printing takes less time to produce.

Print Dialog

The Printing dialog box is shown during the time that <<YourApp>> is sending output to the printer. The page number indicates the progress of the printing.

To abort printing, choose Cancel.

1.6 Print Preview command

Use this command to display the active document as it would appear when printed. When you choose this command, the main window will be replaced with a print preview window in which one or two pages will be displayed in their printed format. The print preview toolbar offers you options to view either one or two pages at a time; move back and forth through the document; zoom in and out of pages; and initiate a print job.

Print Preview toolbar

The print preview toolbar offers you the following options:

Print

Bring up the print dialog box, to start a print job.

Next Page

Preview the next printed page.

Prev Page

Preview the previous printed page.

One Page / Two Page

Preview one or two printed pages at a time.

Zoom In

Take a closer look at the printed page.

Zoom Out

Take a larger look at the printed page.

Close

Return from print preview to the editing window.

1.7 Print Setup command

Use this command to select a printer and a printer connection. This command presents a Print Setup dialog box , where you specify the printer and its connection.

The following options allow you to select the destination printer and its connection.

Printer

Select the printer you want to use. Choose the Default Printer; or choose the Specific Printer option and select one of the current installed printers shown in the box. You install printers and configure ports using the Windows Control Panel.

Orientation

Choose Portrait or Landscape.

Paper Size

Select the size of paper that the document is to be printed on.

Paper Source

Some printers offer multiple trays for different paper sources. Specify the tray here.

Options

Displays a dialog box where you can make additional choices about printing, specific to the type of printer you have selected.

Network...

Choose this button to connect to a network location, assigning it a new drive letter.

1.8 Import sequence

NOT YET IMPLEMENTED!

1.9 Export FastA

NOT YET IMPLEMENTED!

1.10 Load LCMS results

NOT YET IMPLEMENTED!

1.11 1, 2, 3, 4 command

Use the numbers and filenames listed at the bottom of the File menu to open the last four sequences you closed. Choose the number that corresponds with the sequence you want to open.

1.12 Exit command

Use this command to end the SequenceEditor session. You can also use the Close command on the application Control menu. SequenceEditor prompts you to save sequences with unsaved changes, if selected in the menu Tools – Customize .

Press the *Yes* button to save the sequence under its old name or under a new name.

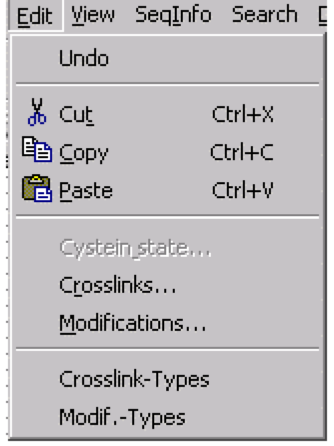
Shortcuts

Mouse: Double-click the application's Control menu button

Keys: ALT+F4

2 Edit Menu

The Edit menu contains the following commands:



Command	Description
Undo...	Reverse previous editing operation.
Cut	Deletes data from the sequence and moves it to the clipboard.
Copy	Copies marked data from the sequence to the clipboard.
Paste	Pastes data from the clipboard into the sequence.
Cysteine state	Not implemented yet.
Crosslinks	Specifies crosslinks within the sequence.
Modifications	Residues in the sequence can be altered by means of modifications.
Crosslink-Types	Crosslink Type Editor is opened.
Modif.-Types	Modification Type Editor is opened.

2.1 Undo

Use this command to reverse the last editing action, if possible. The name of the command changes, depending on what the last action was.

Shortcuts

Keys: CTRL+Z or

Introducing any change inside the sequence area, SequenceEditor performs all updating by itself. The total mass is calculated again and the modifications and crosslinks are updated. Most of the operations can be undone, considering the defined modifications and crosslinks. Currently, there are 8 levels of undoing.

2.2 Cut

Use this command to remove the currently selected data from the document and put it on the clipboard. This command is unavailable if there is no data currently selected.

Cutting data to the clipboard replaces the contents previously stored there.

Shortcuts

Toolbar:

Keys: CTRL+X

Move the cursor to the residue (or select a group) you want to cut and perform the operation. The cut function bases on the ASCII format. The selected data is deleted from the sequence area and moved to the clipboard without any information on the existing modifications or crosslinks.

2.3 Copy

Use this command to copy selected data onto the clipboard. This command is unavailable if there is no data currently selected.

Copying data to the clipboard replaces the contents previously stored there.

Shortcuts

Toolbar:

Keys: CTRL+C

Move the cursor to the residue (or select a group) you want to copy and perform the operation. The copy function is based on the ASCII format. The selected data is copied to the clipboard without any information on the existing modifications or crosslinks.

2.4 Paste

Use this command to insert a copy of the clipboard contents at the insertion point. This command is unavailable if the clipboard is empty.

Shortcuts

Toolbar:

Keys: CTRL+V

Move the cursor to the residue (or select a group) that should be the insert point and perform the operation. If any residue or group was selected the new data replace the selection, otherwise the data from clipboard is inserted. The paste function bases on the ASCII format. The selected data is pasted from the clipboard without any modification or crosslinks information. Any characters that are not the allowed amino acids are ignored.

2.5 Cysteine state

NOT IMPLEMENTED YET.

2.6 Crosslinks

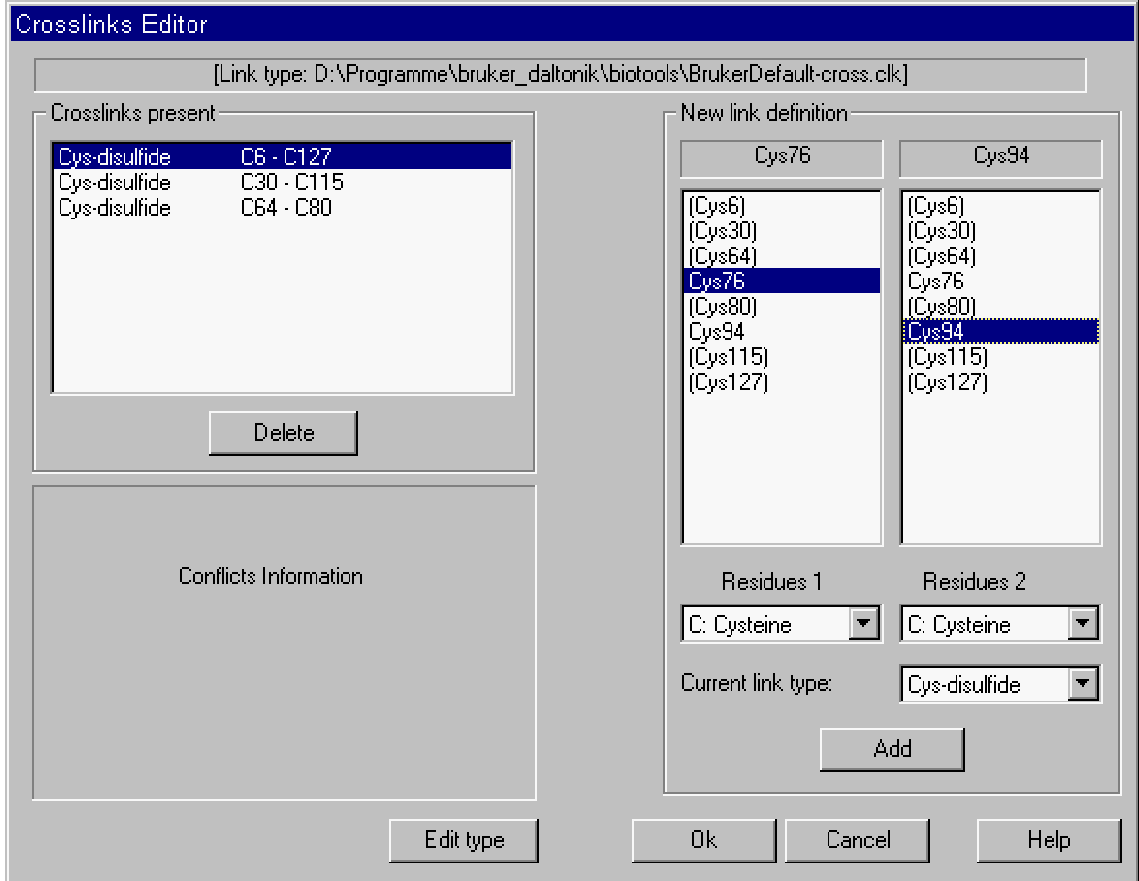
Crosslinks connect different residues within a single chain through the side-groups of the building blocks. The standard and the most prominent way to link different parts of a protein to each other is the disulfide bond linking cysteine (C) residues. Crosslinks are shown on the screen as colored lines binding the respective residues. Editing of the sequence like insert/overwrite, pasting, deleting or cutting consider the currently defined crosslinks and update them along with the changes made to the sequence.

SequenceEditor supports links delivered with the default crosslink file BrukerDefault.clk and any links defined by the user by means of the Crosslink Type Editor .

When switching from the cysteine state Oxidized (SS) to another state the existing disulfide crosslinks are not lost but are hidden.

The crosslinks are entered into the sequence by means of the Crosslinks Editor :

By clicking on Edit type button Edit type the Crosslink Type Editor will be started.



The color of the crosslinks can be changed by the user by means of the Options - Colors command.

2.6.1 Current Crosslink File

Shows the currently selected crosslink file.

The crosslinks defined for the system reside in a special ".clk" file. SequenceEditor is delivered along with a default file *BrukerDefault.clk*. The default crosslink file can not be changed in Crosslink Type Editor. Only a user created crosslink file can be edited and saved or a new file can be loaded by the user.

2.6.2 List of Crosslinks

Shows the list of all crosslinks specified in the system.

2.6.3 Delete a Crosslink

Select this button to delete the selected crosslink.

2.6.4 Conflicts information

Any conflicts in the editor (e. g. selecting of an already occupied residue) are monitored in this field.

2.6.5 New Link Definition

New crosslinks are defined here.

2.6.5.1 Residue 1

Monitors the first residue selected for a crosslink.

2.6.5.2 Residue 2

Monitors the second residue selected for a crosslink.

2.6.5.3 Residues list

These fields show all residues of the selected types existing within the sequence. Each field acts as a choice pool for the residue 1 and 2 respectively. Residues already linked (or modified in Modifications Editor) can not be used again and are shown in parentheses.

To define a new crosslink, select one residue in each field, specify the crosslink type and select the *Add* button.

2.6.5.4 Residue type 1

Selects the type of the first residue involved in a crosslink.

2.6.5.5 Residue type 2

Selects the type of the second residue involved in a crosslink.

2.6.5.6 Current link type

Selects the type of the crosslink to be appended.

2.6.5.7 Add a crosslink

Select this button to add a new crosslink. All relevant fields must be filled.

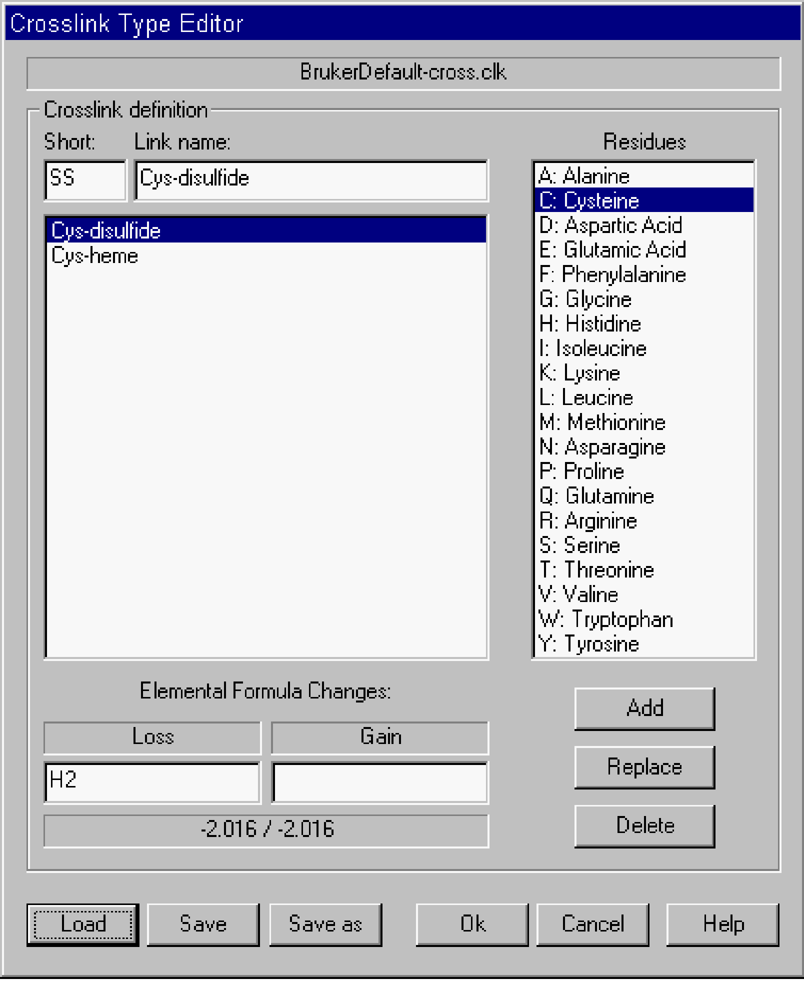
2.7 Crosslink-Types

This option opens the Crosslink Type Editor.

New crosslink types are entered by means of the Crosslink Type editor.

Any link can be dynamically defined, replaced or deleted along with its name, the short symbol, allowed residues, gains and loses.

Link definitions reside in a special system file supported with the program (BrukerDefault.clk). The crosslink types used in the system are stored in the currently selected crosslink file. Any crosslink file existing in the system can be loaded to support the work of individual users. New crosslinks can be specified and stored in the crosslink file (exception BrukerDefault.clk). The mass change caused by a crosslink insertion will be monitored while calculating the total mass of the sequence.



2.7.1 Current Crosslink File

The current crosslink is set within the Crosslink Type Editor. Once set, the crosslink type always appears at the top of the crosslink types list.

2.7.2 Crosslink Definition

New crosslinks are defined in this area.

2.7.2.1 Crosslink short name

The short name of the selected or new specified crosslink type.

2.7.2.2 Crosslink name

The full name of the selected or new specified crosslink type.

2.7.2.3 Crosslink types list

List of all crosslink types defined in the currently selected crosslink file.

2.7.2.4 Residues Selection List

These fields show all residues of the selected types existing within the sequence.

To add new residues to a crosslink choose from the list the crosslink type, select the residues and Replace the crosslink or Add a new one.

2.7.2.5 Formula changes: Loss

Inserting a crosslink into the sequence can cause gains and losses in the molecular weight of the sequence. Enter the loss into this field.

2.7.2.6 Formula changes: Gain

Inserting a crosslink into the sequence can cause gains and losses in the molecular weight of the sequence. Enter the gain into this field.

2.7.2.7 Resulting mass

The mass change resulting from inserting the selected crosslink into the sequence.

2.7.2.8 Add a crosslink type

This button adds a new crosslink type filling it with the contents of the defined fields.

2.7.2.9 Replace a crosslink type

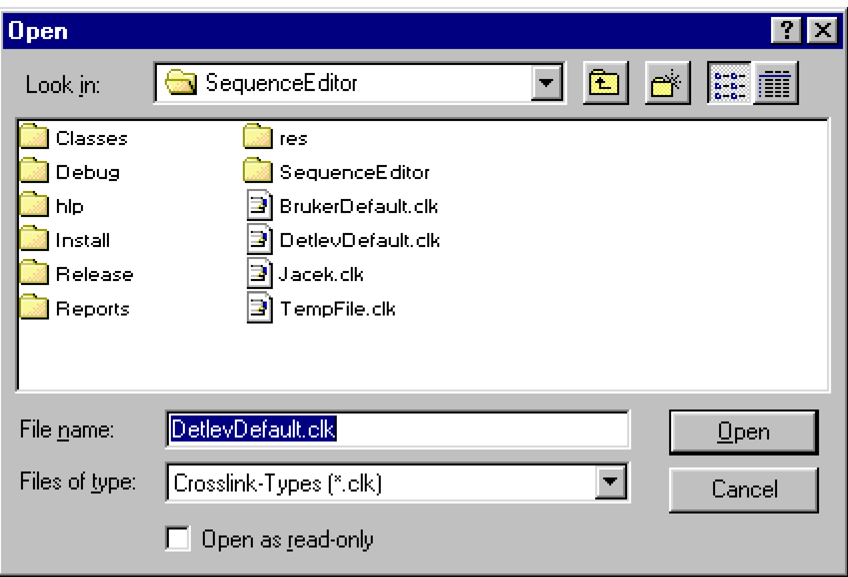
This button replaces the selected crosslink type with the content of the defined fields.

2.7.2.10 Delete a crosslink type

This button deletes the selected crosslink type.

2.7.3 Load a crosslink file

Select this button to load a new crosslink file. On selection, the following dialog box appears:

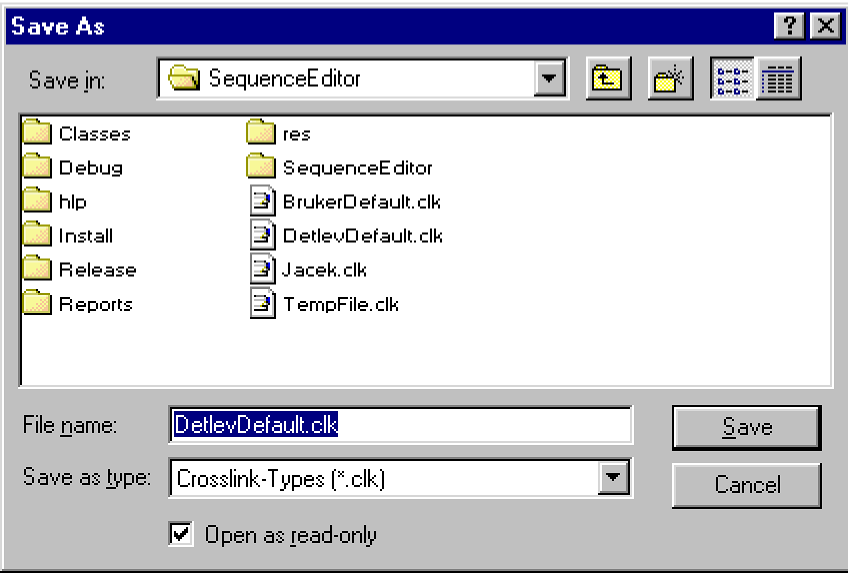


2.7.4 Save a crosslink file

Select the Save button Save to save the changes to the currently selected crosslink file.

2.7.5 Save a new crosslink file

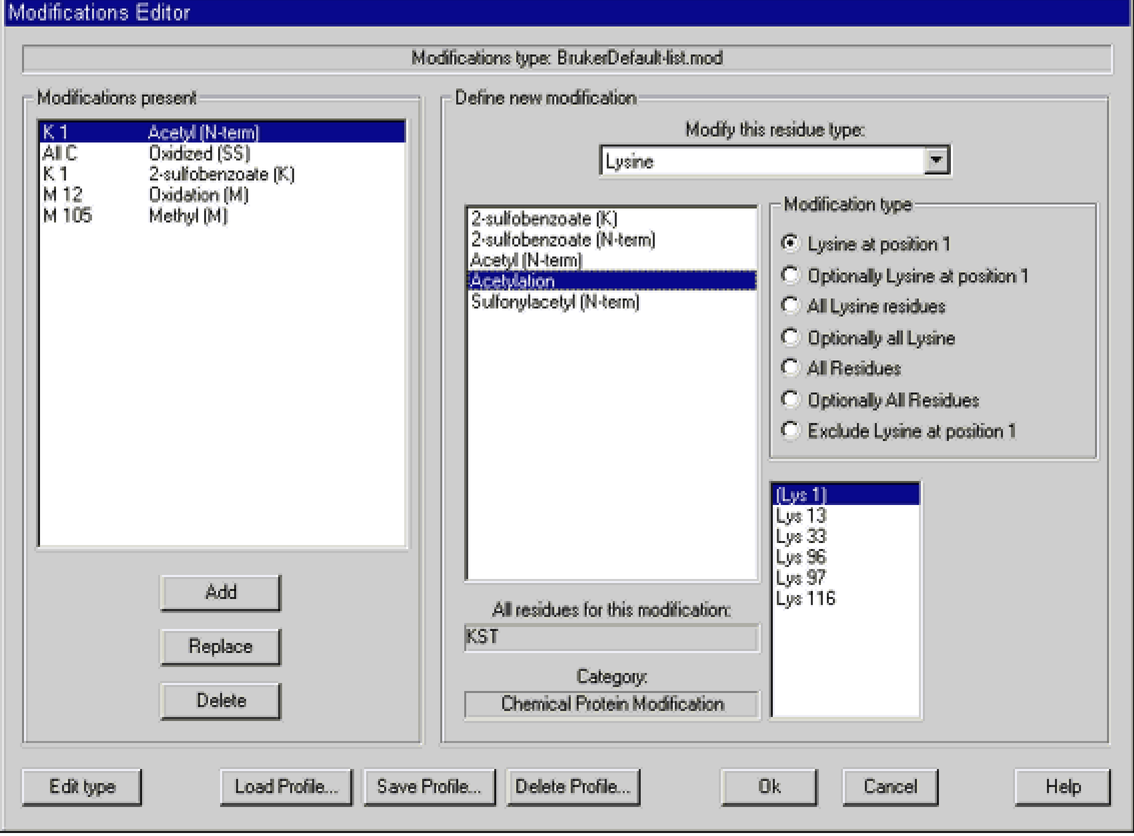
Select the Save As button [Save As] to save the changes to the file under another name. On selection the following dialog box appears:



2.8 Modifications

Residues in the sequence can be altered by means of modifications. The applied modifications are stored along with the sequence and are read by each calling of the sequence. If a read modification does not exist in the local data bank of *SequenceEditor* (modifications file) it can be easily appended to it. Modifications that differ in the notation but have the same chemical formula (e. g.: NH_2 and H_2N) are automatically recognized by the system.

In the SequenceEditor the user has the possibility to freely define the modifications. By means of the Modifications Editor the user can choose between different modification types. Like in the case of crosslinks, the modifications applied to the sequence can be stored along with the protein data.



There are three locations for modifications:

- N-terminal
- C-terminal and
- Side chain modifications.

Three types of modifications are distinguished:

- Local (specified residue selected)
- Global (all residues of one or more kinds, except for the locally modified residues)
- Optional modifications at certain residues or consensus sequence motifs.

Modifications can be specified only at the allowed positions (i.e. left- and right-terminal or inside the chain). No double modifications are possible. A terminal residue can however be modified with a terminal modification and simultaneously be applied a side chain modification.

By clicking on Edit type button [Edit type] the Modification Type Editor will be started.

The color of the modified residues can be changed by means of the Options - Colors command.

2.8.1 Current Modification File

Shows the currently selected modification file.

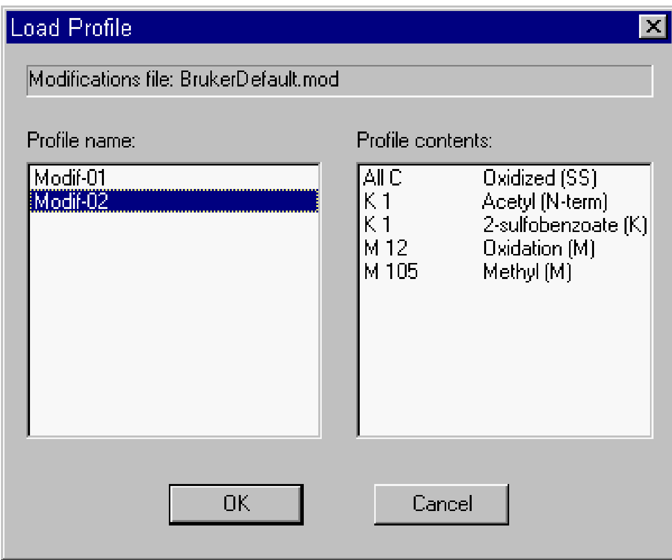
The modifications defined for the system reside in a special ".mod" file. SequenceEditor is delivered along with a default file *BrukerDefault.mod*. The default modification file can not be altered. Only a user created modification file can be edited and saved or a new file can be loaded by the user.

2.8.2 Modifications Profile

Modifications Profiles can be loaded, saved and deleted.

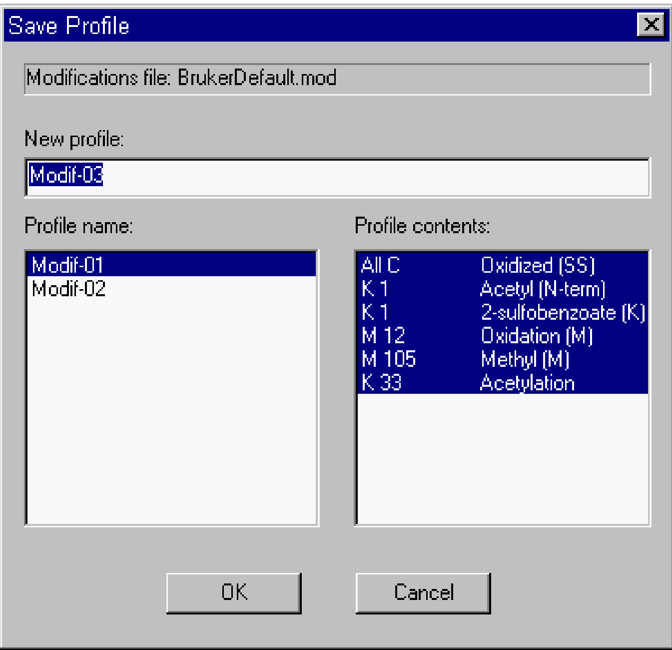
2.8.2.1 Load Profile...

A previously saved profile can be choosen for loading as current modifications profile.



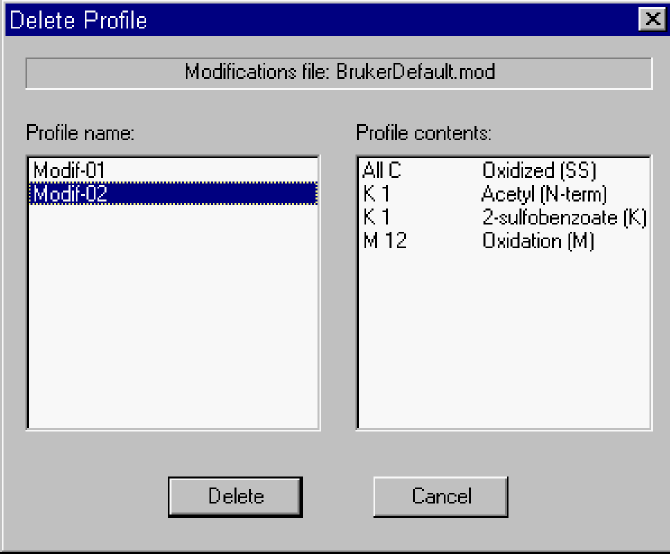
2.8.2.2 Save Profile...

A changed profile can be saved under a new profile name.



2.8.2.3 Delete Profile...

A profile can be deleted from the list of modifications profile.



2.8.3 Kinds of Modifications

2.8.3.1 Local modifications

Double clicking with the left mouse button on a residue opens the modifications dialog with a selection list of all modifications available for the selected residue.

2.8.3.2 Global modifications

Global modifications concern a specified type or types of residues and can be defined only for those residues that are not locally modified.

2.8.3.3 Optional modifications

They are defined like global modifications and used if a certain modification is suspected somewhere.Modifications present

2.8.3.4 Modifications list

This list contains all modifications applied to the current sequence.

2.8.3.5 Add Modification

Select this button to apply the selected modification to the sequence.

2.8.3.6 Replace Modification

Select this button to replace the modification selected in the left field with the one selected in the available modifications list.

2.8.3.7 Delete Modification

Select this button to delete the selected modification.

2.8.4 Define new modification

2.8.4.1 Modify this residue type

The residue type is selected here.

2.8.4.2 Available modifications

All modifications available for the selected residue type are listed here.

2.8.4.3 Modification types

Applying a modification the user has the ability to specify which residues are to be concerned by the modification. Click to define the needed modification type.

2.8.4.4 Residues list

All residues of the same type as clicked upon are listed in this box. The residues that have already been modified or crosslinked are shown in parentheses and are not available for a new modification.

2.8.4.5 Members list

The list of all residue types for which the selected modification is valid.

2.8.4.6 Category

The category of the currently selected modification.

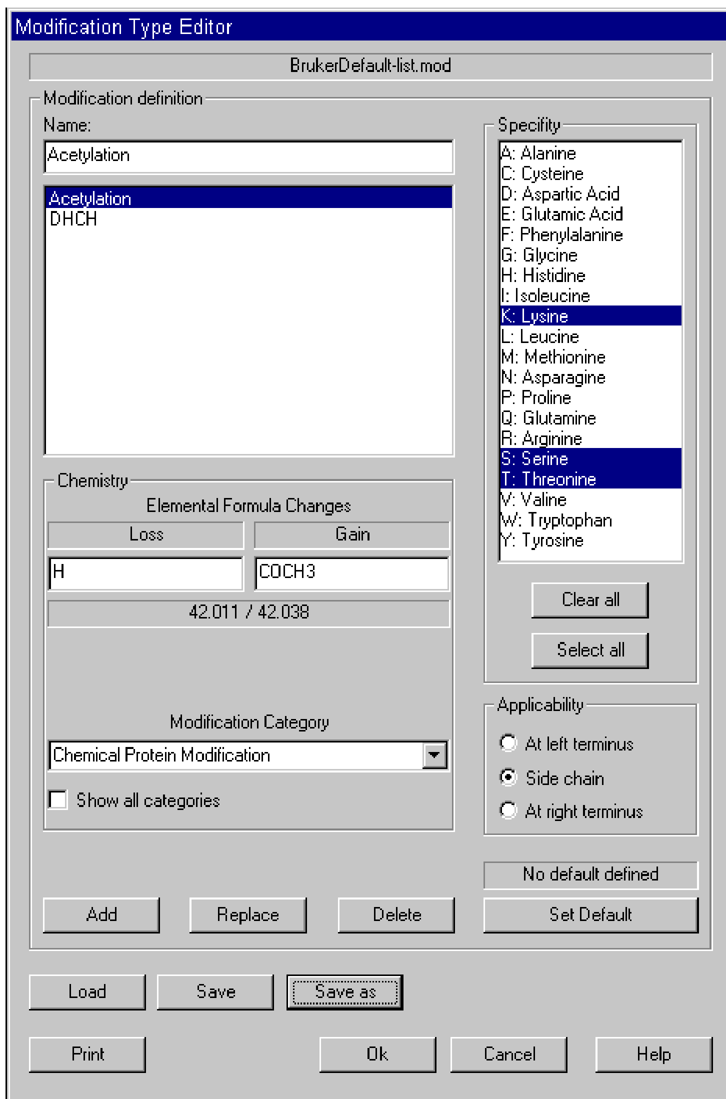
2.9 Modif.-Types

This option opens the Modification Type Editor.

New modification types are entered by means of the Modification Type Editor.

The modification types can be dynamically defined by the user and stored in a system file. Again, the default system file (BrukerDefault.mod) can not be changed but there is the possibility to save it under another name or to load any other system file.

New modification types are entered by means of the Modification Type Editor:



2.9.1 Current Modification File

The current modification is set within the Modification Type Editor. Once set, the modification type always appears at the top of the modification types list.

2.9.2 Modification name

Name of a new defined or currently selected modification type.

2.9.3 Modifications list

List of all modification types available in the currently selected category (or all categories if *Show all categories* is selected).

2.9.4 Chemistry

2.9.4.1 Elemental Formula Changes

Loses and gains introduced by this modification type.

2.9.4.2 Mass change

Mass change resulting from this modification type.

2.9.4.3 Modification category

Currently selected modification type category.

2.9.4.4 Show all categories

If selected all available modification types are shown, otherwise (default) the modification types available within the currently selected category.

Specifity

2.9.5 Specifity

2.9.5.1 Modification members

Residue types for which the modification type is valid.

2.9.5.2 Clear all

Use this button to clear all the member selections.

2.9.5.3 Select all

Use this button to select all members in the list.

2.9.6 Applicability

Defines the modification locations.

2.9.7 Add modification type

Select this button to add a new modification definition. All relevant fields must be filled.

2.9.8 Replace modification type

Select this button to replace the currently selected modification type with the data from the definition fields.

2.9.9 Delete modification type

Select this button to delete the currently selected modification type.

2.9.10 Default modification type

This field shows the default modification type for the currently selected category.

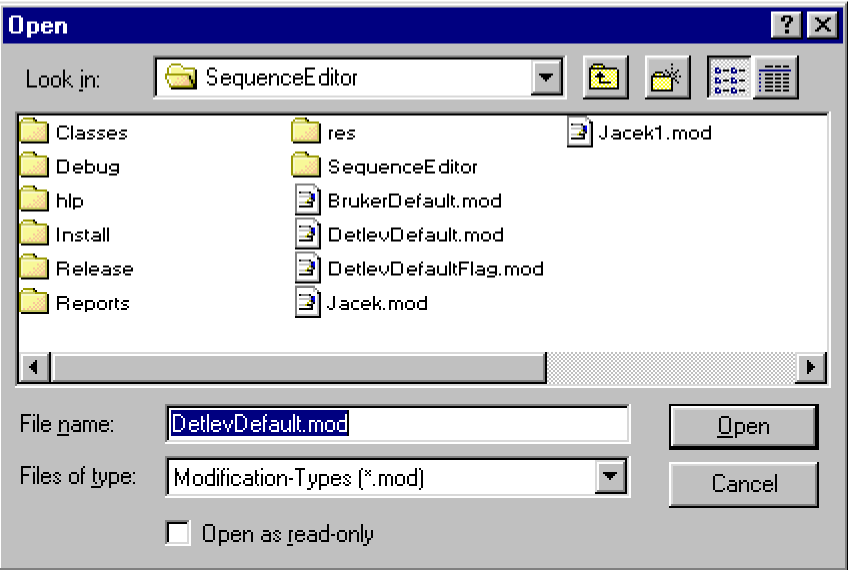
2.9.11 Set default

Select this button to set the currently selected modification type as the default for the current category.

2.9.12 Modification File

2.9.12.1 Load a modification file

Select this button to load a new modification file. On selection, the following dialog box appears:

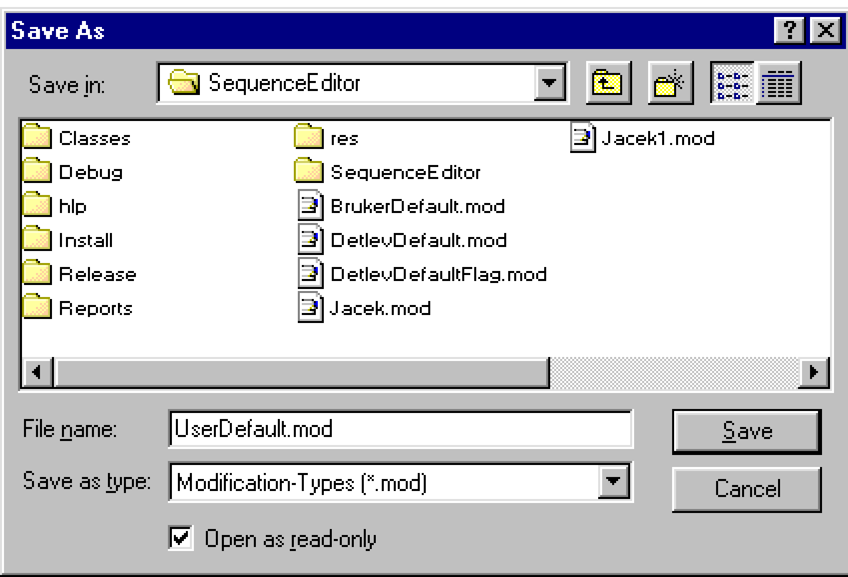


2.9.12.2 Save a modification file

Select this button to save the changes to the currently selected modification file.

2.9.12.3 Save a new modification file

Select the button Save As to save the changes to a new modification file. On selection the following dialog box appears:

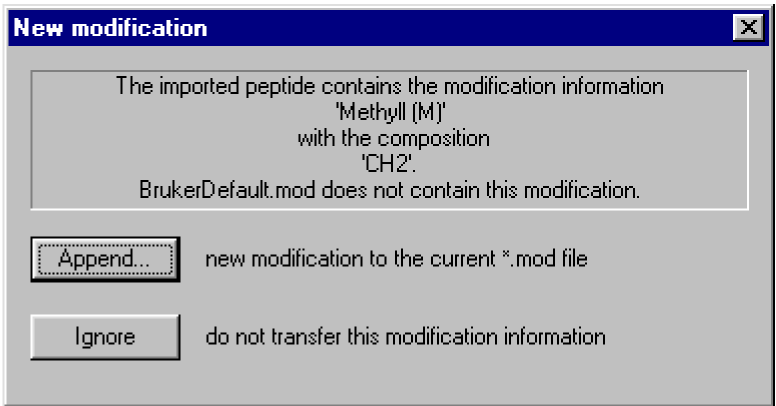


2.9.12.4 Unknown modification type

In case of an unknown modification type read along with an amino sequence the following dialog box appears and the user has the possibility with

Append to append the new modification type to the current modification file or with

Ignore to abort the modification information.



2.9.12.5 Print Modification File

Select this button to get a printout of the currently selected modification file.

3 View Menu

The View menu contains the following commands:

Toolbar	Shows or hides the upper part of the sequence toolbar.
Status Bar	Shows or hides the status bar .
Ave/Mono	Toggles mass display between Average and Monoisotopic.
Show residues	Shows defined residues within the sequence.
Show all types	Shows all predefined residue types within the sequence.
Hide all types	Hides all predefined residue types within the sequence.
Amino acids	Edits amino acids table.

3.1 Toolbar

The toolbar is displayed across the top of the main application window, below the menu bar and is accessible to all sequence windows currently existing. The toolbar provides quick mouse access to many tools used in SequenceEditor.

To hide or display the Toolbar, choose *View menu - Toolbar*.

Toolbar button	***Menu option***	**Shortcut F-key**	***Description***
	File – New Sequence	Ctrl + N	Creates a new sequence, also from the web (Sequences from the Web).
	File – Save	Ctrl + S	Saves the active sequence with its current name. If you have not named the sequence, your SequenceEditor displays the Save As dialog box.
	Edit – Cut	Ctrl + X	Removes selected data from the sequence and stores it on the clipboard.
	Edit – Copy	Ctrl + C	Copy the selection to the clipboard.
	Edit – Paste	Ctrl + V	Inserts the contents of the clipboard at the insertion point.
Bio	-		The modified sequence or a marked range of it will be Send data to BioTools for further processing.
	-		Prints the active data file immediately
	Help – About SequenceEditor…		Opens About SequenceEditor window
	-		Activates the context sensitive help.
	File – Print	Ctrl + P	Opens the printer dialogue
	File – Exit	ALT + F4	Terminates the program

3.2 Status Bar

Use this command to display and hide the Status Bar, which describes the action to be executed by the selected menu item or depressed toolbar button, and keyboard latch state. A check mark appears next to the menu item when the Status Bar is displayed.

The status bar is displayed at the bottom of the SequenceEditor window. To display or hide the status bar, use the Status Bar command in the View menu.

If the mouse cursor is placed over an modified residue then on the left side is displayed information of local and global modifications of that residue, but no optional modifications will be shown.

The right areas of the status bar indicate which of the following keys are latched down:

Indicator	Description
CAP	Caps lock key is activated.
NUM	NUM key is activated.

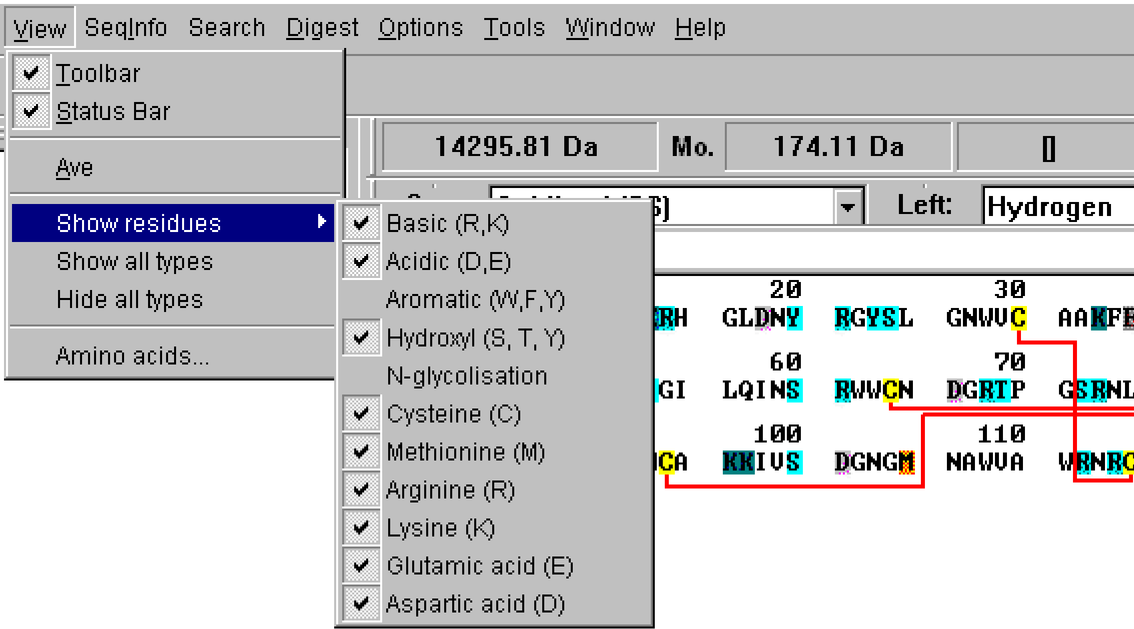
INS/OVR Toggles between INS mode and OVR mode. Toggle with the INS/OVR key on the keyboard between the insert and overwrite mode of the cursor.

3.3 Mono - Ave

With the option Ave/Mono the mass display can be set between monoisotopic mode and average mode. The actual mode is shown beneath the toolbar of the sequence window. The alternative mode is shown here in the Menu *View – Ave(Mono)* and you can toggle here or by clicking on the Mono-Ave-Button.

3.4 Show residues

The Show residues facility is a quick way to search for and to show pre-selected residues or their groups. Ten topics have been fixed in the program:



Selecting one of the above topics results in checking of the respective menu item and showing all occurrences of the specified residues in the background color according to the setting selected in Options - Colors.

Selecting the topic again toggles the menu item and the highlighting disappears.

3.5 Show all types

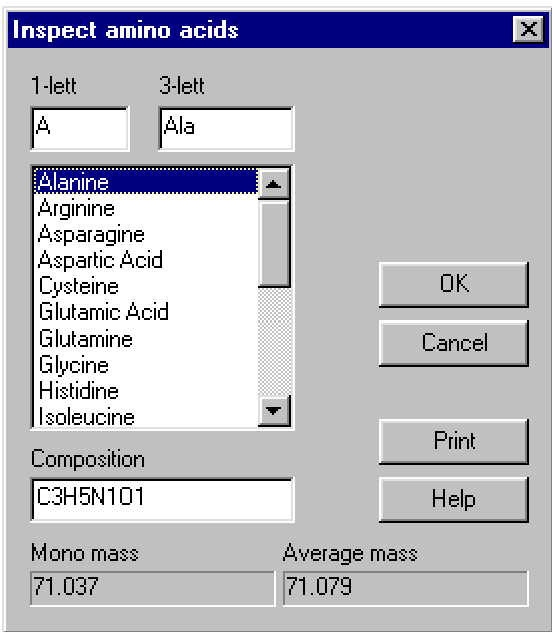
Shows all types of residues in the sequence window, simultaneously all types of residues are selected in the command Show residues.

3.6 Hide all types

Hides all types of residues in the sequence window, simultaneously all types of residues are de-selected in the command Show residues.

3.7 Edit amino acids

The amino acids table used by the SequenceEditor can be updated by the user. On this command the following dialog box appears:



The 1-letter code, the 3-letter code and the chemical composition of a selected amino acid can be altered and saved in the amino acids file (this file is delivered along with the SequenceEditor).

3.7.1 One-letter code

The 1-letter code of the selected amino acid is shown here and can be altered.

3.7.2 Three-letter code

The 3-letter code of the selected amino acid is shown here and can be altered.

3.7.3 Amino acids list

List of available amino acids.

3.7.4 Chemical composition

Chemical composition of the currently selected amino acid.

3.7.5 Amino acid mass

The mass of the currently selected amino acid.

3.7.6 Print Amino Acids Table

Select this button to get a printout of a list of Amino acids.

4 SeqInfo Menu

The Digest menu contains the following commands:

Chain names	NOT IMPLEMENTED YET!
Sample info	NOT IMPLEMENTED YET!
Isotopic pattern	NOT IMPLEMENTED YET!
Data Base entry text	NOT IMPLEMENTED YET!
Charge states	The information on the sequence masses will be shown.
Composition	The information on the composition of the selected sequence will be shown.

4.1 Chain names

NOT IMPLEMENTED YET!

4.2 Sample info

NOT IMPLEMENTED YET!

4.3 Isotopic pattern

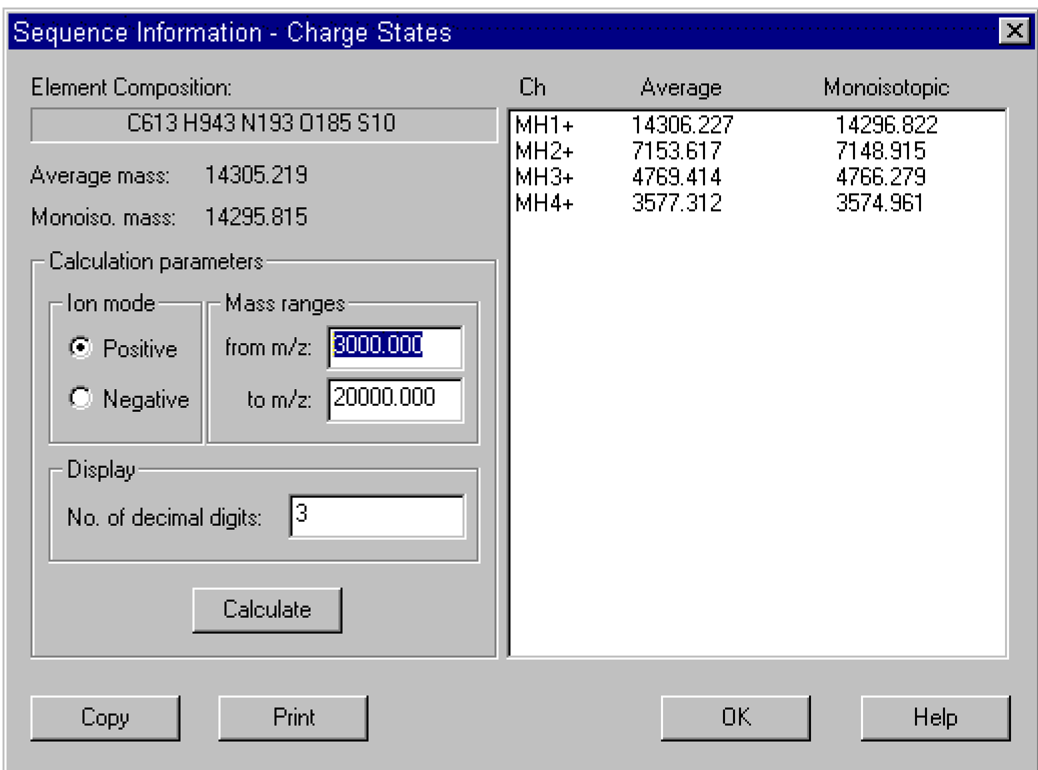
NOT IMPLEMENTED YET!

4.4 Data Base entry text

NOT IMPLEMENTED YET!

4.5 Charge States

The information on the sequence masses appears:



4.5.1 Element composition

The elemental composition of the whole sequence.

4.5.2 Total masses

Total mass shown as monoisotopic and average.

4.5.3 Calculation Parameters

After setting the desired values in Ion mode, Mass Range and Display Digits, the values in the mass list will be refreshed by clicking the Calculate button.

4.5.3.1 Ion mode

Switches between the MH^+ and MH^- mode.

4.5.3.2 Mass Ranges

The smallest and largest mass for calculation results can be set here. Masses outside this mass range will not be displayed in the list of calculated masses.

From m/z

Masses equal or larger than this limit will be shown. The lowest mass must be larger than 100 m/z, it will be set to 100 m/z if set lower.

To m/z

Masses equal or smaller than this limit will be shown.

4.5.3.3 Display Digits

The positions after decimal point used within this dialog box.

4.5.3.4 Calculate button

Refreshes the values in the mass list.

4.5.4 Calculated Masses

The calculated masses are listed here.

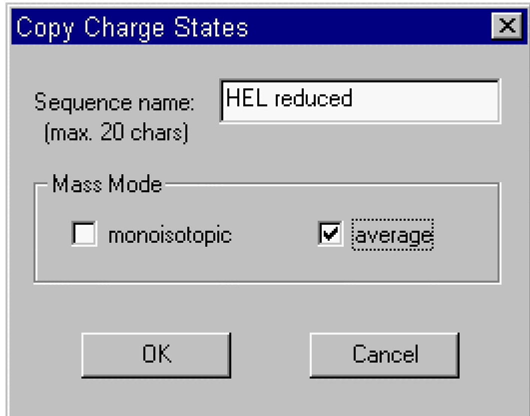
4.5.5 Copy button

The list of calculated masses is copied into the clipboard for further processing with acquisition programs (e. g., Bruker DataAnalaysis), or table calculating and text editing programs.

The dialog box Copy Charge States is opened for setting the sequence name and the desired mass mode.

4.5.6 Copy Charge States

The list of calculated masses is copied into the clipboard for further processing with acquisition programs (e. g., Bruker DataAnalaysis), or table calculating and text editing programs.



4.5.6.1 Sequence name

Enter a name for the sequence to be exported and copied into the clipboard.

4.5.6.2 Mass Mode

Select one or both mass mode (monoisotopic or average) for the exported sequence.

4.5.7 Print button

The list of calculated masses can be printed.

SequenceEditor

Date: 01/08/2002 Time: 15:59

FileName: D:\Programme\Bruker Daltonics\BioTools\DemoData\sequence\hel.sqs

Sequence: HEL reduced

Sequence Information - Masses

Element composition: C613 H967 N193 O185 S10

Monoisotopic mass: 14303.877

Average mass: 14313.283

Ion mode: positive

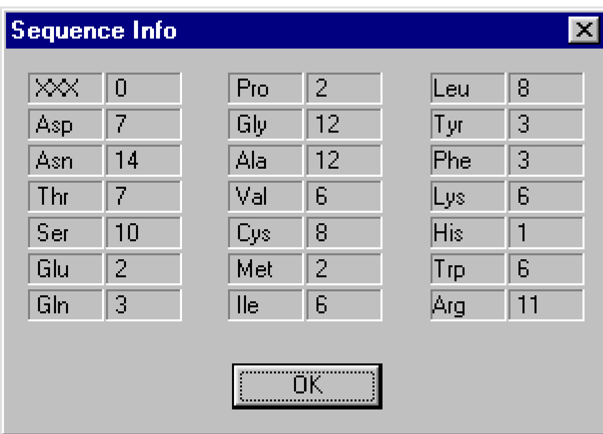
Range: 700.000 - 3000.000

No.	Ch	Average	Monoisotopic	No.	Ch	Average	Monoisotopic
1	MH5+	2863.664	2861.783	2	MH6+	2386.555	2384.987
3	MH7+	2045.762	2044.418	4	MH8+	1790.168	1788.992
5	MH9+	1591.372	1590.327	6	MH10+	1432.336	1431.395
7	MH11+	1302.215	1301.360	8	MH12+	1193.781	1192.997
9	MH13+	1102.029	1101.306	10	MH14+	1023.385	1022.713
11	MH15+	955.226	954.599	12	MH16+	895.588	895.000
13	MH17+	842.965	842.412	14	MH18+	796.190	795.667
15	MH19+	754.338	753.843	16	MH20+	716.672	716.201

SequenceEditor 2.1 ©2000 Bruker Daltonik GmbH Page -1-

4.6 Composition

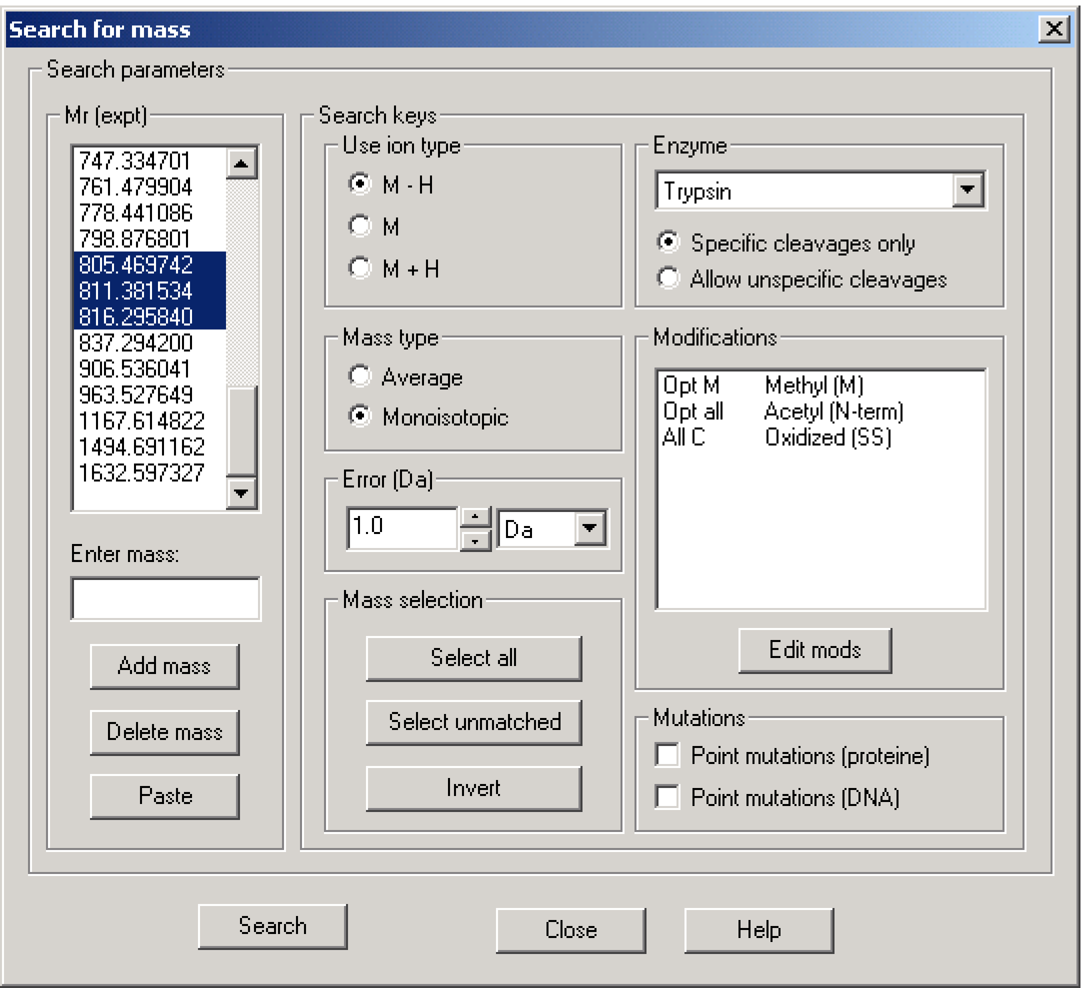
The information on the composition of the selected sequence appears:



5 Search menu

5.1 Mass search

When you obtain a mass value of a peptide (e.g. from a mass spectrum) you often need to know where in a given protein the peptide originated. In order to carry out a search this command can be selected.



There are two modes of the command activity: either you call it from the SequenceEditor (local mode) or the command will be activated by the BioTools

program (remote mode). If the search operation has been requested by BioTools, a new sequence window is created and the sequence in which the masses are to be searched shown.

In both cases the search will be performed for the masses specified in the following dialog box. In the case of the remote operation the mass list will be filled automatically.

Select the Search button [Search] to start the search and to open the Mass search results.

By clicking on Edit mods button [Edit mods] the Modifications Editor will be started.

5.1.1 Search Parameters

5.1.1.1 Masses

The masses (from BioTools or entered manually) to be searched for.

5.1.1.2 Enter mass

The mass that should be appended to the mass list.

5.1.1.3 Add mass

Adds the mass from the *Enter mass* field into the mass list.

5.1.1.4 Delete mass

Deletes selected masses from the mass list.

5.1.1.5 Paste

Pastes masses residing in the clipboard into the mass list.

5.1.1.6 Use Ion type

You can select either M-H, M or M+H. When selecting M+H the mass of a hydrogen will be subtracted from the values entered in the table before the actual search is performed.

5.1.1.7 Mass type

Depending on the sequence window either average or monoisotopic mass will be selected.

5.1.1.8 Error

Determines the precision of the mass search. Only found masses matching the range mass ± precision will be reported.

5.1.1.9 Error units

Selects the units of the precision entry.

5.1.1.10 Enzyme

Enzyme which specificity will be considered while reporting the mass search results.

5.1.1.11 Specific cleavage only

Only peptides that fit the digest specificity of the selected enzyme will be shown in the results.

5.1.1.12 Allow unspecific cleavages

Peptide terminals that fit the enzyme specificity will be marked with a colored '<' and '>' characters.

5.1.1.13 Select all

Selects all masses in the mass list. Only selected masses will be considered during the search.

5.1.1.14 Select unmatched

Selects all masses in the mass list. Only selected masses will be considered during the search.

5.1.1.15 Invert

The masses selection in the mass list will be inverted. Only selected masses will be considered during the search.

5.1.1.16 Modifications list

This list contains all modifications applied to the current sequence.

By clicking on Edit mods button Edit mods the Modifications Editor will be started.

5.1.1.17 Mutations

BioTools allows you to screen a sequence for single point mutations. Prerequisites are an MS/MS spectrum of a peptide from a fingerprint and a reference sequence, which was obtained in a previous fingerprint library search.

What you do:

- After MASCOT mass fingerprint search, import the candidate sequence with *Get Hit(s).*
- Start a *Search for masses* search from BioTools in Sequence Editor. Select only the one mass you want to match against the MS/MS spectrum.
- Click one of the *Mutations* options (Point mutation (proteine) or Point mutations (DNA)) and start the search.

Assumption in this search: You loaded a correct protein sequence into SE, but the sample of interest contained 1 diverging residue in the peptide, a "mutation". A mutation Q405D therefore means, the Q405 residue in the native sequence was actually identified as a D residue.

5.1.1.17.1 Point mutation (proteine)

All single amino acid residue exchanges are calculated.

Applications requiring this option:

1. Check for sequence errors in the database

Account for Point Mutations in site-directed mutagenesis experiments involving full codon exchanges in the investigated material with respect to the sequence in the SequenceEditor

5.1.1.17.2 Point mutations (DNA)

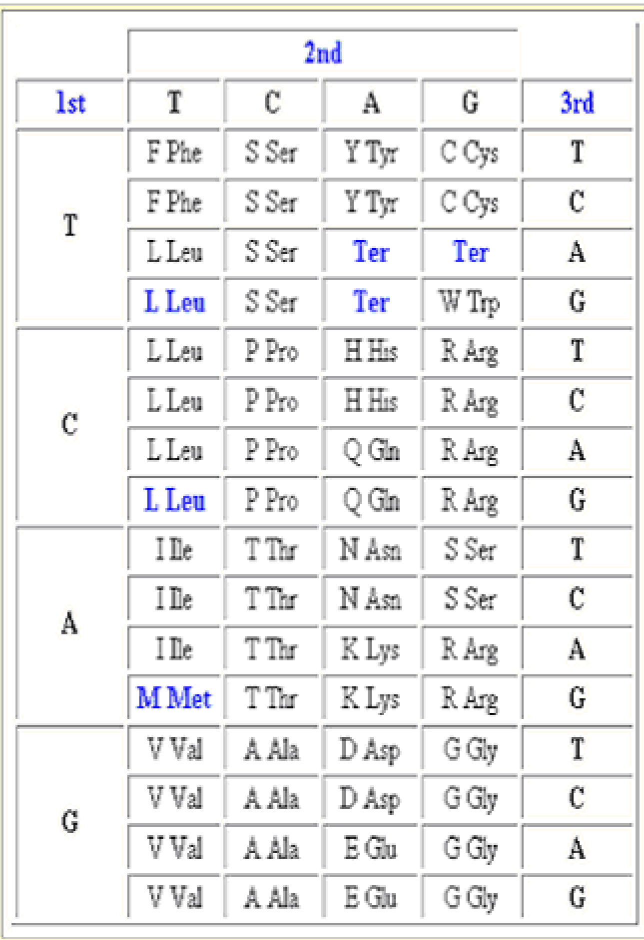
Only amino acid residue exchanges are calculated, which can be accounted for by a single base exchange on the DNA level. This corresponds to a single nucleotide polymorphous site (SNP).

Applications requiring this option:

1. Check sample for SNP with respect to the sequence in the database. This is currently only done on the protein level. Since the actual codon encoding an amino acid residue is not known, the analysis is not conclusive on the corresponding codon level. I.e., a point mutation in the Leu codon TTG could give rise to Trp (TGG). Leu (CTT) could not be transformed by a single nucleotide exchange into Trp, for which only the codon TGG is available.

2. Account for Point Mutations in site-directed mutagenesis experiments involving full codon exchanges in the investigated material with respect to the sequence in the SequenceEditor.

Handling of practical situations:



	2nd				
1st	T	C	A	G	3rd
T	F Phe	S Ser	Y Tyr	C Cys	T
	F Phe	S Ser	Y Tyr	C Cys	C
	L Leu	S Ser	Ter	Ter	A
	L Leu	S Ser	Ter	W Trp	G
C	L Leu	P Pro	H His	R Arg	T
	L Leu	P Pro	H His	R Arg	C
	L Leu	P Pro	Q Gln	R Arg	A
	L Leu	P Pro	Q Gln	R Arg	G
A	I Ile	T Thr	N Asn	S Ser	T
	I Ile	T Thr	N Asn	S Ser	C
	I Ile	T Thr	K Lys	R Arg	A
	M Met	T Thr	K Lys	R Arg	G
G	V Val	A Ala	D Asp	G Gly	T
	V Val	A Ala	D Asp	G Gly	C
	V Val	A Ala	E Glu	G Gly	A
	V Val	A Ala	E Glu	G Gly	G

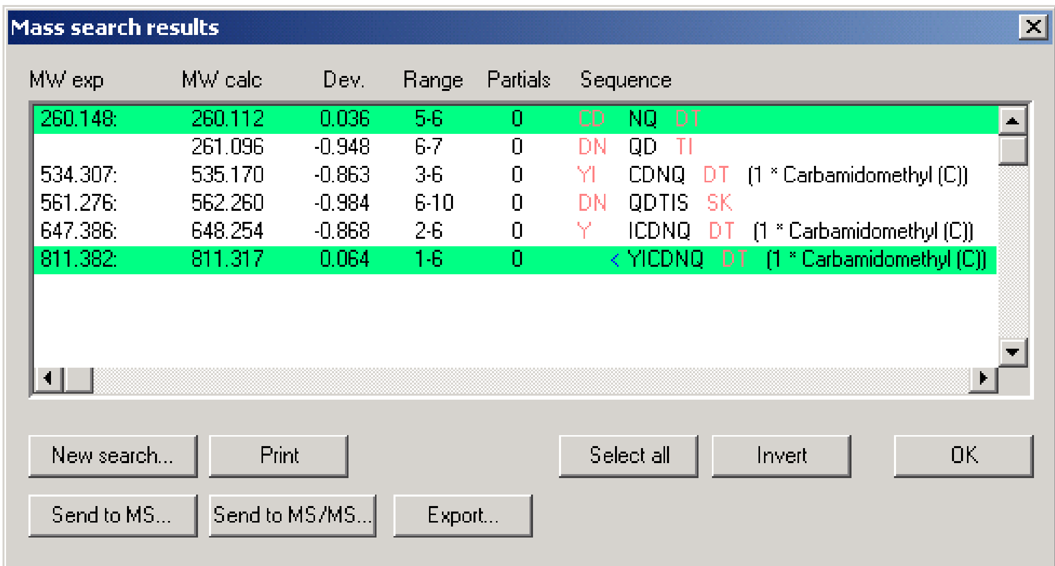
5.1.1.18 Search Button

Select the Search button Search to start the search and to open the Mass search results.

5.1.2 Mass Search Results

The search results are listed in increasing mass order.

Select the New Search button [New Search] to return to the Search for mass dialog box. The search operation can be repeated.



5.1.2.1 Results field

The results are shown as lines consisting of following fields:

- Mass to search for
- Found mass
- Deviation (search mass – found mass)
- Peptide range matching the search mass
- Partials
- Sequence shown in black; the preceding and following residues will be shown in a different color.

If an enzyme has been specified in the search parameters then only peptides that conform to the digest specifications for that enzyme will be shown if only Specific cleavage only has been checked.

If Allow unspecific cleavages has been checked the positions will be checked with a mark before ('<') and after ('>') the sequence as shown above.

With a click on a peptide, the corresponding peptide in the parent sequence window will be colored red and underlined.

5.1.2.2 New search

Select the New Search button [New Search] to return to the Search for mass dialog box. The search operation can be repeated.

5.1.2.3 Print

The list of calculated masses can be printed.

5.1.2.4 Select all

Selects all results in the results field.

5.1.2.5 Invert

The selections in the result results field will be inverted.

5.1.2.6 Send to MS

Select this button to send the results of the requested search operations to BioTools. Only the selected peptides will be send to BioTools.

5.1.2.7 Send to MS/MS

Select this button to send the results of the requested search operations to BioTools. Only the selected peptides will be send to BioTools.

5.1.3 Export theoretical digest results dialog

The list of calculated masses can be exported as csv-file (comma separated file) to get imported in other programs (e. g. table calculating programs).

Export table type

Here in it can be selected whether the whole table or only the list of masses should be exported as results.

Export format
Choose the format of the data:

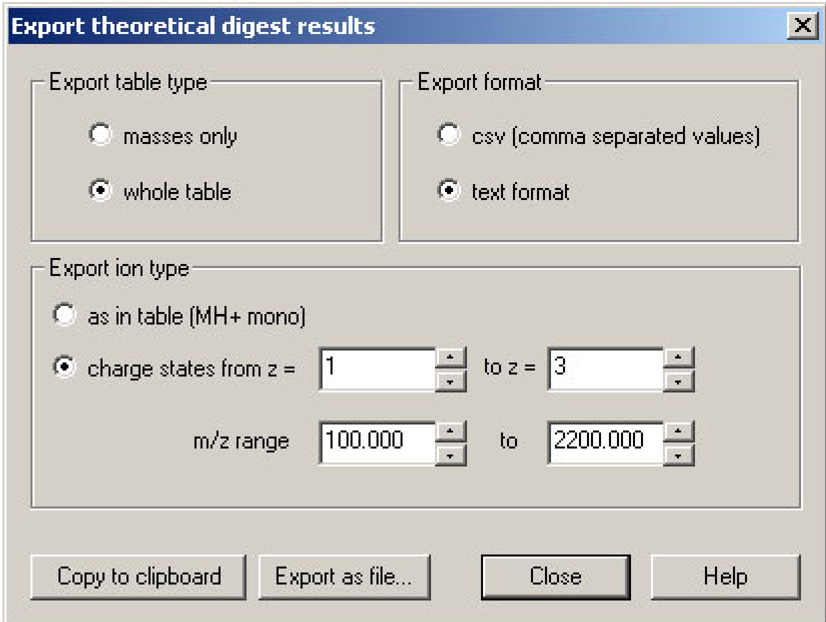
- with **csv (comma separated values)** the columns of the table will be separated by semi comma.
- with **text format** the columns of the table will be separated by tab stop.

Export ion type
Choose the ion type of the data:

- **as in table:** The data will be set as shown in the results window
- **charge states from:** To the results the multiple charged ions in the previous set, and here changeable mass range will be exported
- Copy to clipboard

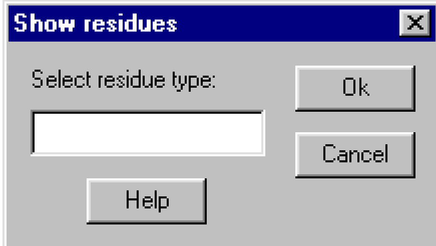
The results are copied into the clipboard for further processing in other programs.

Export as file
The results can be exported as csv-file for further processing in other programs.



5.2 Sequence

Residues or residue motifs specified in the following dialog box will be searched for and shown in invert colors. A left mouse button click within the sequence area results in disappearing of the highlight.



6 Digest menu

6.1 Perform digest

Call this command to perform an enzymatic digest or to define new digest agents.

Digest agents fragment a particular protein (or DNA or RNA) into smaller peptides after recognition of a particular amino acid residue or a more complex sequence motif. The task in these analyses is to match the mass spectrum with all its peaks to the protein sequence + modifications to account for as much structural information as possible. Artifacts make this task difficult sometimes: unspecific or irregular digests, unknown modifications, contaminations, autoproteplysis of the cutting enzyme, etc.

Calculation results can be used locally, like for the prediction of enzymatic peptide fragment masses or in conjunction with particular datasets: MALDI or LC-MS fingerprints, etc.

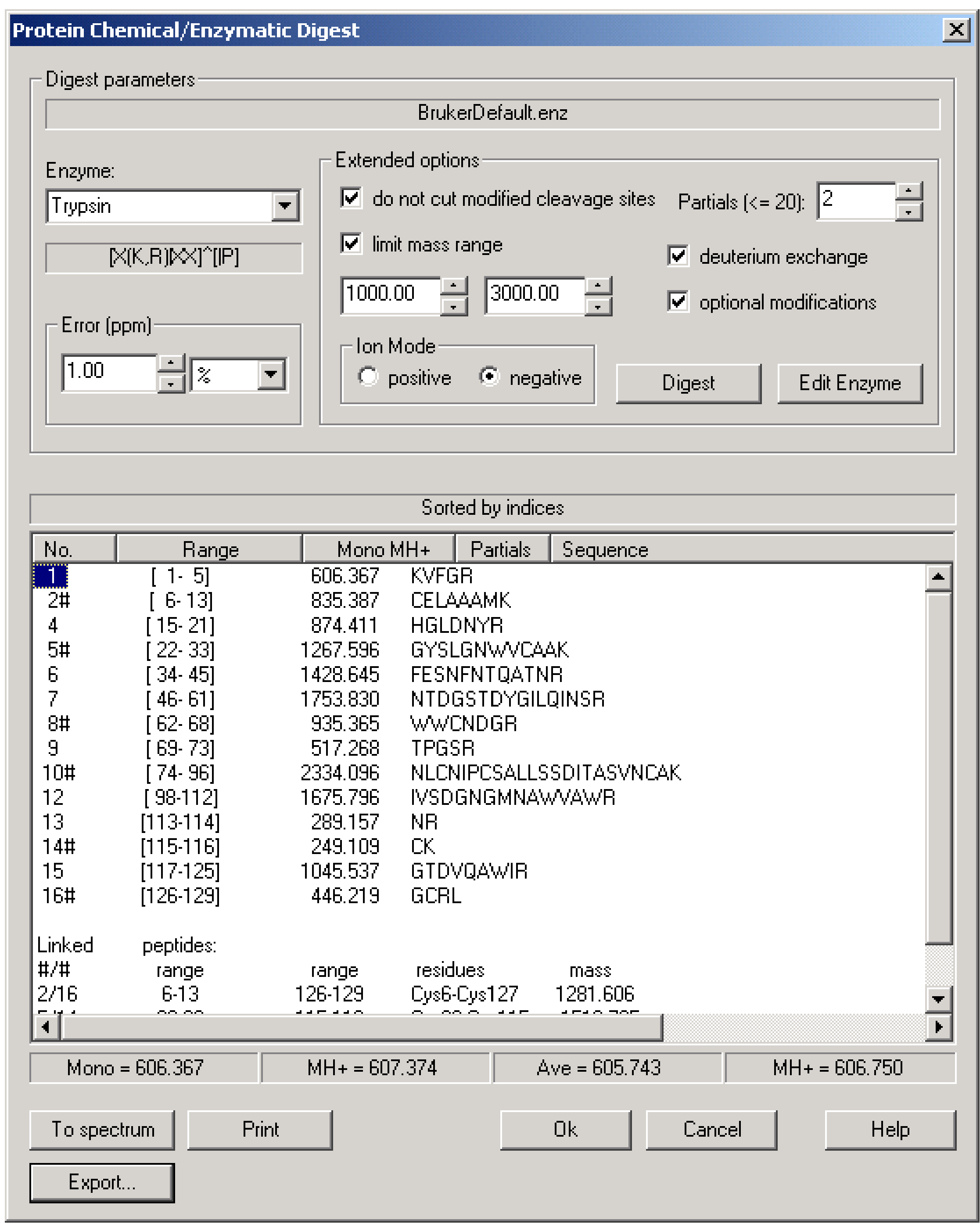
Defined digest agents can be used to digest the sequence. Along with the agent's name, the mass range and prohibition of cleaving at modified positions can be specified. The Partials option allows you to skip a number of cutting positions to check if the larger peptides are still within the mass range specified. Such larger fragments are marked with asterisk (*) in the result list.

In order to find some unknown masses, optional modifications are implemented. For each peptide containing residues defined as allowed for optional modifications, the mass is calculated with and without all possible modification configurations.

Both monoisotopic and average masses can be toggled in the result list. The results can be sorted by numbers (indices) or by masses. Fragments that are pseudo-independent but connected by crosslinks are listed in the table separately.

After setting the parameters select the Digest button Digest to start the digest operation and to get the results in the list of the window.

Select the Edit Enzyme button Edit Enzyme to start the Enzyme Specifity Editor.


Protein Chemical/Enzymatic Digest
Digest parameters
BrukerDefault.enz
Enzyme:
Trypsin
[X][K,R][X][X]^[P]
Error (ppm)
1.00
%
Extended options
do not cut modified cleavage sites
Partials (<= 20):
2
limit mass range
deuterium exchange
1000.00
3000.00
optional modifications
Ion Mode
positive
negative
Digest
Edit Enzyme
Sorted by indices
No.
Range
Mono MH+
Partials
Sequence
1 [1- 5] 606.367 KVFGR
2# [6- 13] 835.387 CELAAAMK
4 [15- 21] 874.411 HGLDNYR
5# [22- 33] 1267.596 GYSLGNWVCAAK
6 [34- 45] 1428.645 FESNFNTQATNR
7 [46- 61] 1753.830 NTDGSTDYGILQINSR
8# [62- 68] 935.365 WWCNDGR
9 [69- 73] 517.268 TPGSR
10# [74- 96] 2334.096 NLCNIPCSALLSSDITASVNCAK
12 [98-112] 1675.796 IVSDGNGMNAWVAWR
13 [113-114] 289.157 NR
14# [115-116] 249.109 CK
15 [117-125] 1045.537 GTDVQAWIR
16# [126-129] 446.219 GCRL
Linked peptides:
#/# range range residues mass
2/16 6-13 126-129 Cys6-Cys127 1281.606
Mono = 606.367
MH+ = 607.374
Ave = 605.743
MH+ = 606.750
To spectrum
Print
Ok
Cancel
Help
Export...

6.1.1 Digest parameters

6.1.1.1 Enzyme file

Shows the currently selected enzyme file.

The enzymes defined for the system reside in a special ".enz" file. *SequenceEditor* is delivered along with a default file *BrukerDefault.enz*. The default enzyme file can not be changed in enzyme types editor. Only a user created enzyme file can be edited and saved or a new file can be loaded by the user.

6.1.1.2 Enzyme list

Shows the list of all enzymes defined for this system. The currently defined enzymes are specified in the enzyme file. The recently used enzyme list is automatically loaded at each program start.

6.1.1.3 Error

Determines the precision of the mass search. Only found masses matching the range mass +- precision will be reported.

6.1.1.4 Error units

Selects the units of the precision entry.

6.1.1.5 Do not cut modified cleavage sites

Do not digest if the position at which the cutting should be performed is modified.

6.1.1.6 Limit mass range

If selected, only the peptides which masses lay within the specified mass range will be shown.

6.1.1.7 Mass Range

The smallest and largest mass for calculation results can be set here, if the limit mass range is selected. Masses outside this mass range will not be displayed in the list of calculated masses.

6.1.1.8 Ion mode

Switches between the MH^+ and MH^- mode.

6.1.1.9 Partials

This option allows to skip a number of cutting positions to check if the larger peptides are still within the mass range specified. Such larger fragments are marked with an asterisk (*) in the result list.

6.1.1.10 Deuterium exchange

Selecting results in hydrogen vs. deuterium change of exchangeable hydrogen atoms. The number of these atoms for each amino acid is read from a special list.

6.1.1.11 Optional modifications

Looking for some unknown masses, sometimes optional modifications have to be specified. If the amino sequence contains any optional specified modifications, enabling this checkbox results in applying this modification to all involved residues and testing if the resulting mass matches the one we are looking for. Found fragments containing optional modifications are marked (*) in the result list with an asterisk and a number specifying the number of optional modifications contained within the peptide.

6.1.1.12 Digest

Select this button to start the digest operation and to get the results in the window below.

No.	Range	Ave MH+	Partials	Sequence
1	[1- 5]	648.79	0	KVFGR (1 * 2-sulfobenzoate (K))
2#	[6- 13]	853.05	0	CELAAAMK (1 * Oxidation (M))
4	[15- 21]	874.94	0	HGLDNYR
5	[22- 33]	1268.47	0	GYSLGNWVCAAK (1 * Oxidized (SS))
6	[34- 45]	1429.50	0	FESNFNTQATNR
7	[46- 61]	1754.86	0	NTDGSTDYGILQINSR
8	[62- 68]	936.02	0	WWCNDGR (1 * Oxidized (SS))
9	[69- 73]	517.57	0	TPGSR
10	[74- 96]	2335.72	0	NLCNIPCSALLSSDITASVNCAK (3 * Oxidized (
12	[98-112]	1690.93	0	IVSDGNGMNAWVAWR (1 * Methyl (M))
13	[113-114]	289.32	0	NR
14	[115-116]	249.33	0	CK (1 * Oxidized (SS))
15	[117-125]	1046.18	0	GTDVQAWIR
16#	[126-129]	448.57	0	GCRL
Linked	peptides:			
#/#	ranges	mass		link
2/16	6-13/126-129	1298.60		Cys6-Cys127

6.1.1.13 Edit Enzyme

Select this button to start the Enzyme Specifity Editor.

6.1.2 Digest Results

6.1.2.1 Sorting information

The currently selected sorting method is shown here. The possibilities are: *Sorted by indices*, *Sorted by masses descendent* and *Sorted by masses asacendent*.

6.1.2.2 Sorting bar

Clicking on the No. partition of this bar results in switching the sorting method between Indices, Masses descending and Masses ascending.

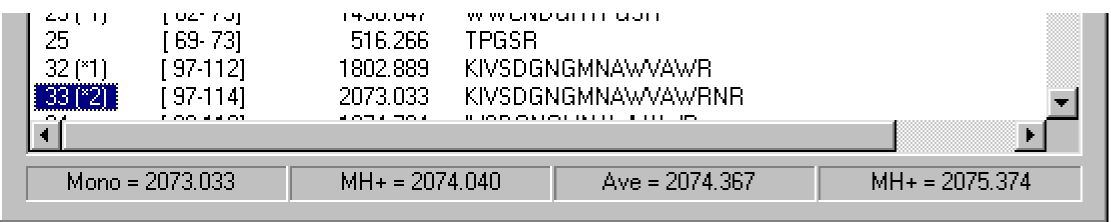
Clicking on the Mono partition results in switching the mass display between the modi Monoisotopic, Monoisotopic MH+, Average and Average MH+.

6.1.2.3 Digest results

The results of the digest operation are shown here.

6.1.2.3.1 Peptide mass

Shows the mass of a selected peptide as respectively: Monoisotopic, Monoisotopic MH+, Average and Average MH+.



6.1.2.3.2 Pseudo-independent fragments

The peptides connected with crosslinks are shown in the last part of the digest results:

No.	Range	Ave MH+	Partials	Sequence
1	[1- 5]	648.79	0	KVFGR (1 * 2-sulfobenzoate (K))
2#	[6- 13]	853.05	0	CELAAAMK (1 * Oxidation (M))
4	[15- 21]	874.94	0	HGLDNYR
5	[22- 33]	1268.47	0	GYSLGNWVCAAK (1 * Oxidized (SS))
6	[34- 45]	1429.50	0	FESNFNTQATNR
7	[46- 61]	1754.86	0	NTDGSTDYGILQINSR
8	[62- 68]	936.02	0	WWCNDGR (1 * Oxidized (SS))
9	[69- 73]	517.57	0	TPGSR
10	[74- 96]	2335.72	0	NLCNIPCSALLSSDITASVNCAK (3 * Oxidized (
12	[98-112]	1690.93	0	IVSDGNGMNAWVAWR (1 * Methyl (M))
13	[113-114]	289.32	0	NR
14	[115-116]	249.33	0	CK (1 * Oxidized (SS))
15	[117-125]	1046.18	0	GTDVQAWIR
16#	[126-129]	448.57	0	GCRL
Linked	peptides:			
#/#	ranges	mass		link
2/16	6-13/126-129	1298.60		Cys6-Cys127

6.1.2.3.3 Crosslinked peptide numbers

Numbers of residues connected by crosslinks (here residues #2 and #16).

6.1.2.3.4 Peptide ranges

Ranges of peptides connected by crosslinks.

6.1.2.3.5 Crosslinked mass

The total mass of two connected by crosslink peptides.

6.1.2.3.6 Crosslinked residues

Respective residues in both peptides that are connected with a crosslink.

6.1.3 To spectrum

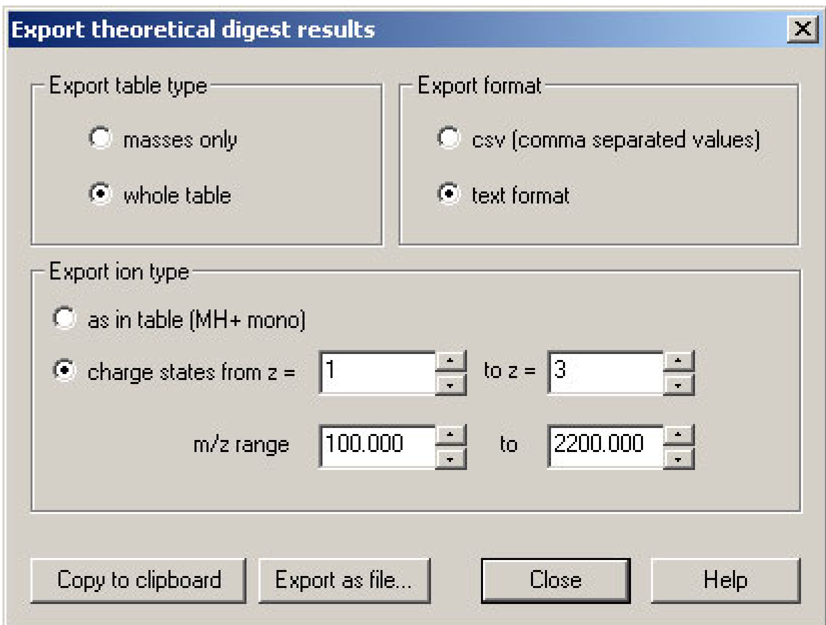
This option results in sending the selected results to the program BioTools. If the data does not match to the active spectrum error messages will appear to repeat the action.

6.1.4 Print

The complete list of digested masses can be printed.

6.1.5 Export theoretical digest results dialog

The list of calculated masses can be exported as csv-file (comma separated file) to get imported in other programs (e. g. table calculating programs).



6.1.5.1 Export table type

Here in it can be selected whether the whole table or only the list of masses should be exported as results.

6.1.5.2 Export format

Choose the format of the data:

- with **csv (comma separated values)** the columns of the table will be separated by semi comma.
- with **text format** the columns of the table will be separated by tab stop.

6.1.5.3 Export ion type

Choose the ion type of the data:

- **as in table:** The data will be set as shown in the results window
- **charge states from:** To the results the multiple charged ions in the previous set, and here changeable mass range will be exported

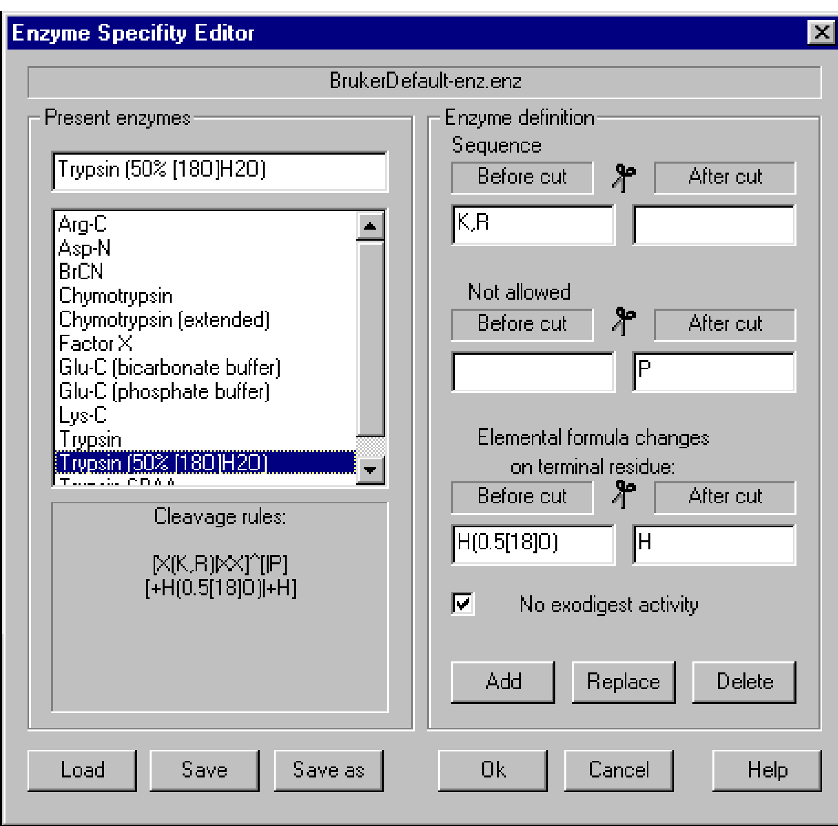
6.1.5.4 Copy to clipboard

The results are copied into the clipboard for further processing in other programs.

6.1.5.5 Export as file

The results can be exported as csv-file for further processing in other programs. To reduce the data file size select in the export dialog the desired data.

6.2 Edit enzymes



The Enzyme Specifity Editor is used to define the digest agents. An exact definition of the enzyme activity can be done. The name, recognition sequence, prohibition conditions and the changes of the elemental formula can be specified. To make sure the program works for all biopolymers, the standard options in the protein programs are not sufficient. Recognition patterns for restriction enzymes exceeding 8 residues need to be definable. Digest reactions can be done in isotopically labeled solutions, thus isotopic labels (e.g.: 50% ^{18}O labeling) can be introduced at the digest site. The proximity of a terminal is considered by means of eliminating exodigest activities. The features of each enzyme are coded in crypted text that can be edited by the user within this editor.

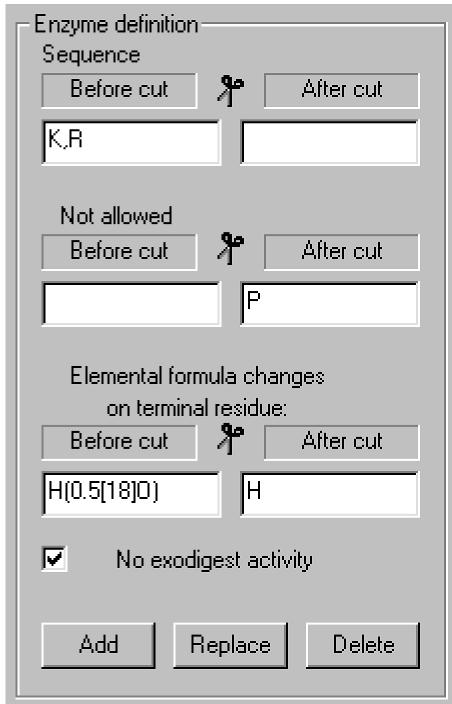
6.2.1 Enzyme name

The name of the currently selected enzyme or of a new enzyme that should be appended to the enzyme list or that replace the currently selected enzyme.

6.2.2 Digest rules

The crypted text describing the activity of the selected enzyme appears in this field.

6.2.3 Enzyme Definition



6.2.3.1 Sequence before cut

The amino acid residue or a more complex motif after which the digest will take place.

6.2.3.2 Sequence after cut

The amino acid residue or a more complex motif causing digest ahead of it.

6.2.3.3 Recognition sequence

An amino acid residue or a more complex motif after or in front of which the digest will take place. The following definition means that the recognition of the amino acid K or R or of the sequence DEG will result in a digest.

K,R,DEG

6.2.3.4 Not allowed before cut

The digest is not allowed if the specified residue(s) exist directly to the left of the cutting site.

6.2.3.5 Not allowed after cut

The digest is not allowed if the specified residue(s) exist directly to the right of the cutting site.

6.2.3.6 Prohibiting conditions

An amino acid residue or a more complex motif prohibiting digest if placed directly ahead or after the cutting site. The following definition means that the recognition of the amino acid P or of the sequence DEG will suppress digest.

P,DEG

6.2.3.7 Formula changes before cut

Changes entered by the enzyme to the left of the cutting site. The following definition:

-SCH3+O

means a loss of a mass respective to the group SCH3 and a gain respective to O.

Additionally, an isotopic labeling is possible. The definition:

H(0.5[18]O)

denotes a solution where, along with hydrogen, the oxygen part consists in 50% of monoisotopic and in 50% of O^{18} isotopic mass. Please note that the whole oxygen expression is enclosed in parentheses ( ) and the isotope specification is within brackets [ ].

6.2.3.8 Formula changes after cut

Changes entered by the enzyme to the right of the cutting site. The following definition:

-SCH3+O

means a loss of a mass respective to the group SCH3 and a gain respective to O.

Additionally, an isotopic labeling is possible. The definition:

H(0.5[18]O)

denotes a solution where, along with hydrogen, the oxygen part consists in 50% of monoisotopic and in 50% of O^{18} isotopic mass. Please note that the whole oxygen expression is enclosed in parentheses ( ) and the isotope specification is within brackets [ ].

6.2.3.9 No exodigest activity

Checking this box results in suppressing the digest at the sequence termini. In order for the cutting to be valid there must be at least two residues to the left or to the right of the digest site.

6.2.3.10 Add enzyme

Select this button to add the specified enzyme.

6.2.3.11 Replace enzyme

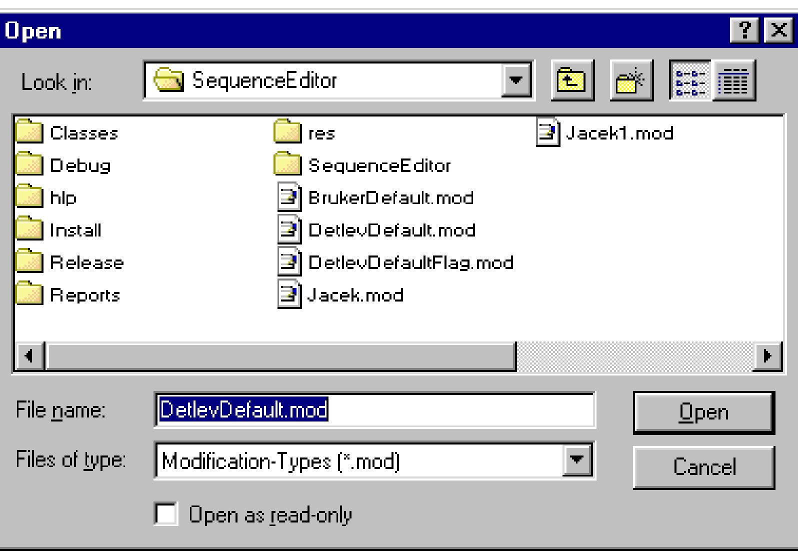
Select this button to replace the selected enzyme with the data entered into the fields.

6.2.3.12 Delete enzyme

Select this button to delete the selected enzyme.

6.2.4 Load an enzyme file

Select this button to load a new enzyme file. On selection the following dialog box appears:

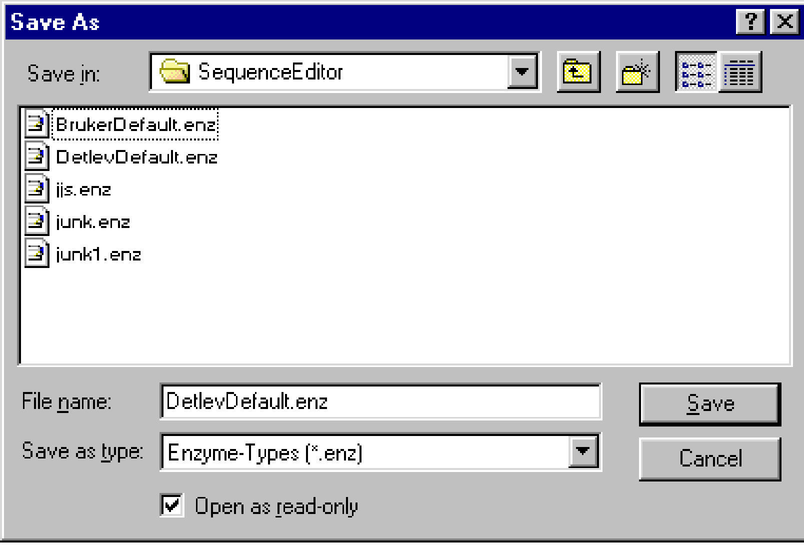


6.2.5 Save an enzyme file

Select this button to save the changes to the currently selected enzyme file under the same name.

6.2.6 Save as a new enzyme file

Select this button to save the changes to another enzyme file under another name. On selection the following dialog box appears:



7 Options menu

This Options menu contains the following commands.

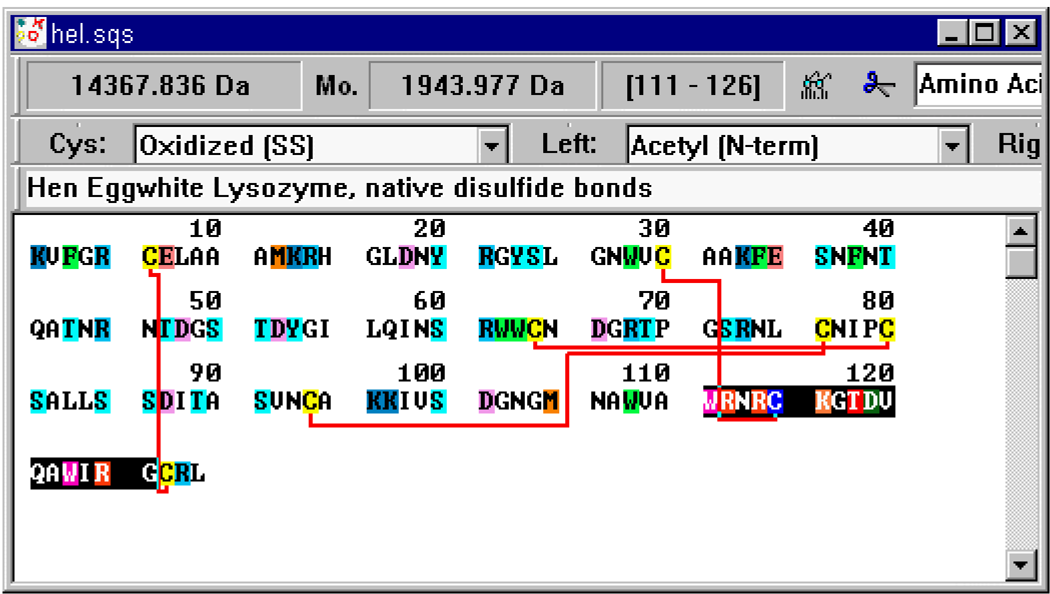
Colors

System

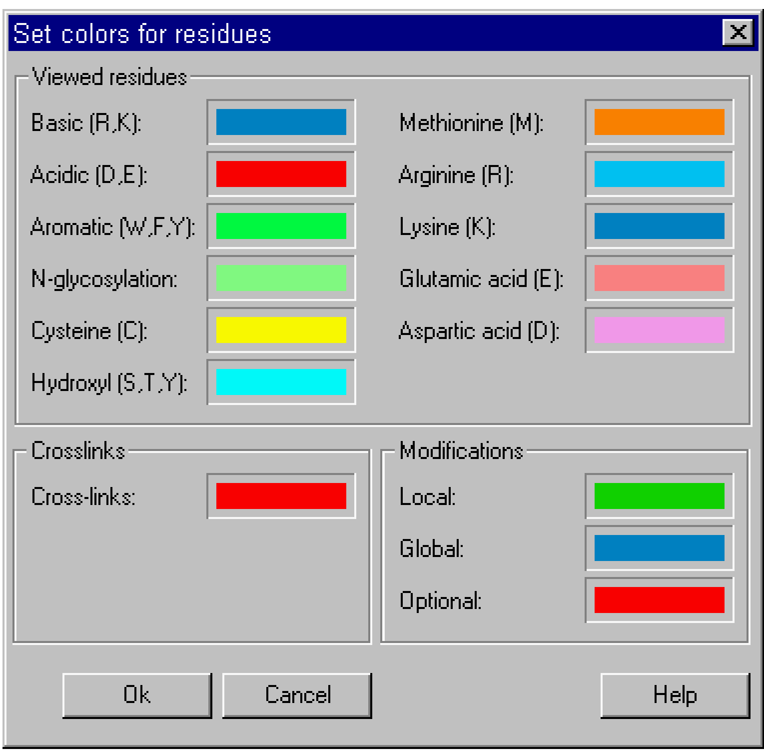
Default Building Blocks

7.1 Colors

The colors used to mark modifications, crosslinks or found consensus motifs can be easily set by the user. During selection for copying and deleting the colors will be displayed inverse.



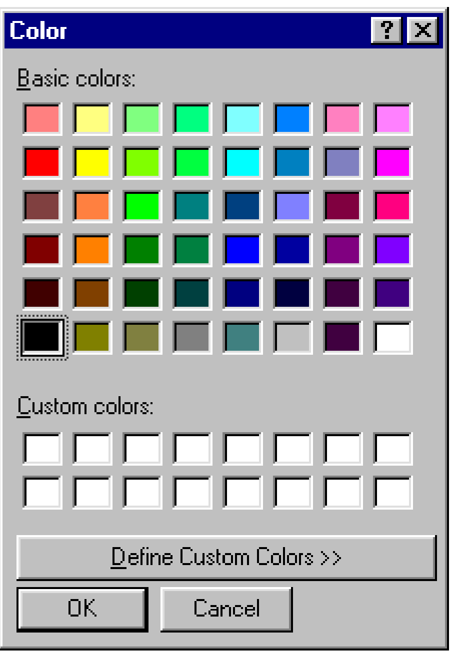
Using the command the following dialog box appears and the individual colors can be changed.

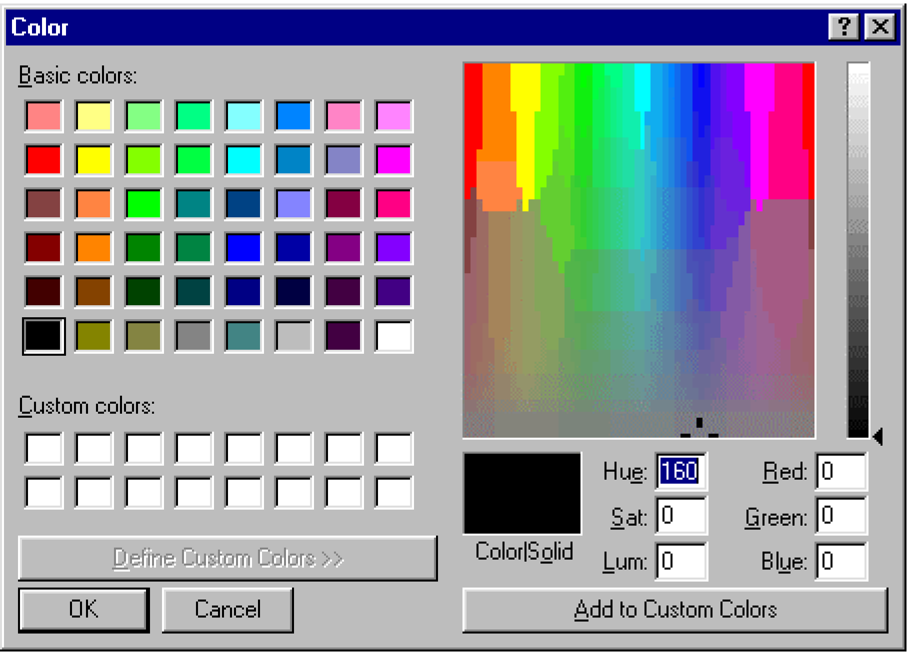


To change a color a click in the respective colored field is required. This results in the color dialog box is shown below:

The required color can either be directly chosen from the upper part of the dialog box or edited by the user. In order to edit the color the Define Custom Colors button should be selected. This selection leads to an expansion of the dialog box:

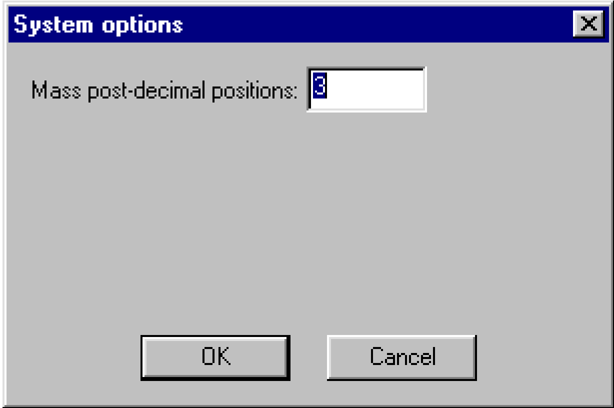
The required color can now be directly chosen from the right part of the dialog box and can be added with the Add to Custom Colors button to the Custom colors to be available for future use.


Color
Basic colors:
Custom colors:
Define Custom Colors >>
OK
Cancel


Color
Basic colors:
Custom colors:
Define Custom Colors >>
OK
Cancel
Color|Solid
Hue: 160
Sat: 0
Lum: 0
Red: 0
Green: 0
Blue: 0
Add to Custom Colors

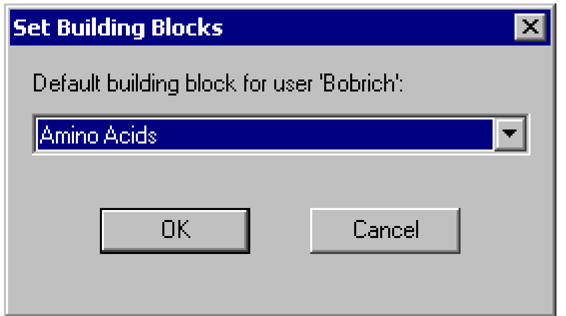
7.2 System

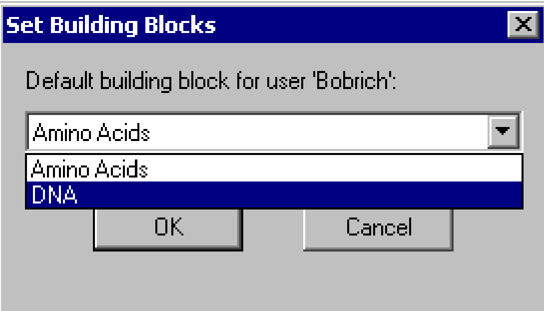
Some system options can be freely defined by the user.



7.3 Default Building Blocks

With this option the default building block (Amino acids or DNA) can be set. The entries for the default building block can be selected from the drop-down list.

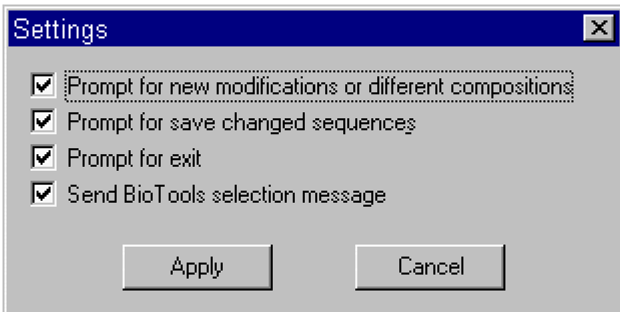




8 Tools menu

8.1 Customize

Some system defaults can be set within this dialog box. Click the Apply button [Apply] to confirm the changes made.



8.1.1 Prompt for New Composition

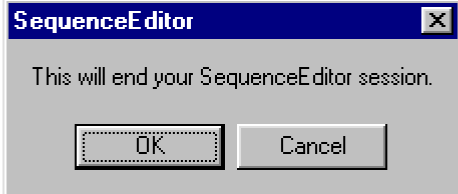
A dialog box asking for saving of unknown modifications appears if this option is checked.

8.1.2 Prompt for save

A dialog box asking for saving of changed sequences while exiting the program or closing a sequence window appears if this options is checked.

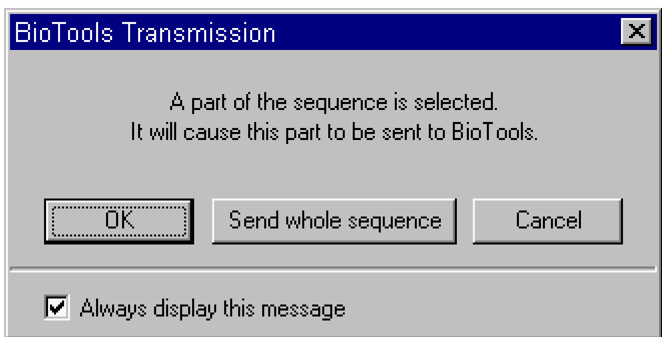
8.1.3 Prompt for exit

The following dialog box appears on exit if this option is checked.



8.1.4 Send BioTools selection message

A dialog box asks for sending the previously marked part of the sequence, or the whole sequence, to BioTools.



9 Window Menu

The WINDOW menu contains the following commands, which enable you to arrange multiple views of multiple documents in the application window:

New Window	Creates a new window that views the same sequence.
Cascade	Arranges windows in an overlapped fashion.
Tile	Arranges windows in non-overlapped tiles.
Arrange Icons	Arranges icons of closed windows.
Active Windows	Goes to specified window.

9.1 New window

Use this command to open a new window with the same contents as the active window. You can open multiple document windows to display different parts or views of a document at the same time. If you change the contents in one window, all other windows containing the same document reflect those changes. When you open a new window, it becomes the active window and is displayed on top of all other open windows.

9.2 Cascade windows

Use this command to arrange multiple opened windows in an overlapped fashion.

9.3 Tile windows

Use this command to arrange multiple opened windows in a tiled fashion.

9.4 Arrange Icons

Use this command to arrange the icons for minimized windows at the bottom of the main window. If there is an open document window at the bottom of the main window, then some or all of the icons may not be visible because they will be underneath this document window.

9.5 List of active Data Files

The SequenceEditor displays a list of currently opened document windows at the bottom of the Window menu. A check mark appears in front of the document name of the active window. Choose a document from this list to make its window active.

10 Help menu

The Help menu contains the following commands:

Context Help	With this option the context sensitive help will be activated
Help Topics	Offers an index to topics on which you can get help
About SequenceEditor	Displays the version number of this application

10.1 Context sensitive Help

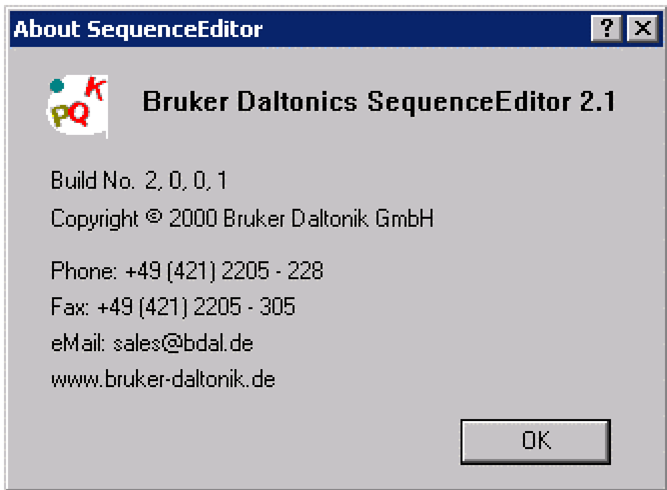
Use the Context Help command to obtain help on some portion of the SequenceEditor. When you choose the Toolbar's Context Help button, the mouse pointer will change to an arrow and question mark. Then click somewhere in the SequenceEditor window, such as another Toolbar button. The Help topic will be shown generally for the item you clicked.

Shortcut

Keys: SHIFT+F1

10.2 About SequenceEditor

Use this command to display the copyright notice and version number of your copy of the SequenceEditor.



About SequenceEditor window

10.3 Help Topics

Use this command to display the opening screen of Help. From the opening screen, you can jump to step-by-step instructions for using the SequenceEditor and various types of reference information.

Once you open Help, you can click the Contents button whenever you want to return to the opening screen.

11 System Menu

The System menu contains the following commands:

Restore

Move

Size

Minimize

Maximize

Close

11.1 Restore command

Use this command to return the active window to its size and position before you chose the Maximize or Minimize command.

11.2 Move command

Use this command to display a four-headed arrow so you can move the active window or dialog box with the arrow keys.

Note: This command is unavailable if you maximize the window.

Shortcut

Keys: CTRL+F7

11.3 Size command

Use this command to display a four-headed arrow so you can size the active window with the arrow keys.

After the pointer changes to the four-headed arrow:

1. Press one of the DIRECTION keys (left, right, up, or down arrow key) to move the pointer to the border you want to move.
2. Press a DIRECTION key to move the border.

3. Press ENTER when the window is the size you want.

Note: This command is unavailable if you maximize the window.

Shortcut

Mouse: Drag the size bars at the corners or edges of the window.

11.4 Minimize Window

Use this command to reduce the SequenceEditor window to an icon.

Shortcut

Mouse: Click the minimize icon on the title bar.

Keys: ALT+F9

11.5 Maximize command

Use this command to enlarge the active window to fill the available space.

Shortcut

Mouse: Click the maximize icon on the title bar; or double-click the title bar.

Keys: CTRL+F10 enlarges a document window.

11.6 Close command

Use this command to close the active window or dialog box.

Double-clicking a Control-menu box is the same as choosing the Close command.

Note: If you have multiple windows open for a single document, the Close command on the document Control menu closes only one window at a time.

Shortcuts

Keys: CTRL+F4 closes a document window

I Index

A

About SequenceEditor 10-2
Add modification 2-14
Allow unspecific cleavages 5-3
Amino mass 0-5
Average mass 4-2

B

Building block 0-6

C

Calculate button 4-3
Calculated Masses 4-3
Calculation Parameters 4-3
Category 2-15
Changes after cut 6-11
Changes before cut 6-10
Charge States
- copy 4-3
- print 4-5

Close Window 11-2
Colors 7-1, 7-2
Composition 4-6
Context sensitive Help 10-1
Copy 2-2
Copy button 4-3
Copy Charge States 4-3, 4-4
Copy to clipboard 6-8
Crosslink Definition 2-7
Crosslink Type Editor 2-6
Crosslinks 2-3, 2-4
Crosslinks editor 0-6, 2-3
Crosslink-Types 2-6
C-terminal 0-7
Current Crosslink File 2-4, 2-7
Current Modification File 2-11, 2-17
Customize 8-1
Cut 2-2
Cysteine state 0-6, 0-7, 2-2

D

Default Building Blocks 7-4
Define new modification 2-14
Delete a Crosslink 2-4
Delete Profile... 2-13
Digits 4-3
Directories 1-3
Display Digits 4-3
Do not cut modified cleavage sites 6-3

E

Edit Enzyme 6-5
Edit Menu 2-1
Element composition 4-2
Elemental Formula Changes 2-17
Enzyme 5-3
Enzyme Definition 6-9
Error 5-3, 6-3
Error units 5-3, 6-3
Esport FastA 1-8
Exit 1-8
Export 5-7
Export as file 6-8
Export FastA 1-8
Export format 6-7
Export ion type 6-8
Export table type 6-7
Export theoretical digest results dialog 6-7

F

File Menu 1-1
files
- managing 1-2, 1-3
- managing 1-1, 1-8

Formula changes after cut 6-11
Formula changes before cut 6-10
From m/z 4-3, 6-3

H

Help Menu 10-1
Help Topics 10-2

I

Import sequence 1-8
Inactive cys-disulfide crosslinks 0-9
Invert 5-3, 5-7

L

List of active Data Files 9-2
List of crosslinks 2-4
Load LCMS results 1-8
Load Profile... 2-12

M

Mass Mode 4-4
Mass Range 4-3, 6-3
Mass Search Button 0-6
Mass Search Results 5-6
Mass type 5-2
Maximize Window 11-2
Menu
  Edit 2-1
  Help 10-1
  View 3-1
  Window 9-1
Menu bar 0-1
Minimize Window 11-2
Modif.-Types 2-16
Modification Type Editor 2-16
Modifications 2-10, 2-11
Modifications Editor 2-10
Modifications list 5-3
Modifications Profile 2-12
Modify this residue type 2-14
Mono-Ave-Button 0-5
Monoisotopic mass 4-2
Move Window 11-1
Mutations 5-4

N

New Link Definition 2-4
New search 5-7
New sequence 1-2
New window 9-1
N-terminal 0-7

O

Optional Modification Transfer Dialog 0-3
Options menu 7-1

P

Paste 2-2
Perform Enzymatic Digest Button 0-6
Point mutation proteine 5-4
Point mutations DNA 5-4
Print 1-5, 4-5, 5-7, 6-7
Print Amino Acids Table 3-6
Print Charge States 4-5
Print Dialog 1-5
Print Modification File 2-20
Print Setup 1-7
printing and print preview 1-5, 1-6, 1-7
Profile 2-12
Prompt_for_New_Composition 8-1

R

Range field 0-5
Residues list 2-4
Residues Selection List 2-4, 2-7
Restore Window 11-1
Results field 5-6

S

Save a crosslink file 2-8
Save a new crosslink file 2-9
Save a new modification file 2-19
Save as a new enzyme file 6-12
Save in 1-3
Save Profile... 2-12
Scrollbar 0-8
Search Button 5-5
Search Parameters 5-2
Select all 5-3, 5-7
Select unmatched 5-3
Send BioTools selection message 8-2
Send data to BioTools 0-3
Send to MS 5-7
Send to MS/MS 5-7
Sequence 5-8
Sequence area 0-8
Sequence name 4-4
Sequence title 0-7
Sequence Toolbar 0-5
SequenceEditor Operator Manual i
SequenceEditor window 0-1
Sequences from the Web 0-4
Size Window 11-1
Specific cleavage only 5-3

Status Bar 0-10, 3-3
System Menu 11-1
System Options 7-4

T
Table of changes iii
To m/z 4-3, 6-3
Toolbar 0-2, 3-2
Total mass 0-5
Total masses 4-2
Tree structure 0-4

U
Undo 2-1
Use Ion type 5-2

V
View Menu 3-1

W
Window Menu 9-1