MassLynx NT BioLynx & ProteinLynx Guide

Version 4.0

10 December, 2001

Waters Part No -715000391

Micromass Part No - 6666664

MassLynx NT BioLynx & ProteinLynx Guide

The software described in this book is furnished under a license agreement and may be used only in accordance with the terms of that agreement.

Copyright Notice

Micromass

Copyright © 1993-2002 Micromass UK Ltd. All Rights Reserved.

No part of this publication may be copied without the express written permission of Micromass UK Ltd.

Apache

This product includes software developed by the Apache Software Foundation (http://www.apache.org/).

Copyright © 1999-2002 The Apache Software Foundation. All rights reserved.

Trademarks

Micromass ® is a registered trademark of Micromass UK Limited (Reg. U.S. Pat. & Tm. Off.).

Windows is a trademark of Microsoft Corporation. Other product names mentioned in this manual may be trademarks or registered trademarks of their respective companies and are hereby acknowledged.

Table of Contents

MassLynx NT BioLynx & ProteinLynx Guide i
MassLynx NT BioLynx & ProteinLynx Guide ii
Copyright Notice ii
Trademarks ii

Chapter 1 BioLynx - Proteins 1-1
Contents 1-2
Figures 1-3
Overview 1-5
The Sequence Window 1-5
Display Options 1-6
Disulfide Bridge Editor 1-7
The Proteins Toolbar 1-8
The Chain Editor 1-9
Editing Text 1-11
Terminal Group Modifications 1-13
Post Translational Modifications 1-15
User-Definable Elements 1-17
Miscellaneous Editing 1-18
Search Routines 1-18
Residue Search 1-18
Multiple Mass Search 1-19
Find Residue Mass Combination 1-22
Digests 1-23
Digest Simulation 1-23
Mass Spectral Fragments 1-28
Spectrum Interpretation 1-33
Hydropathic Profiles 1-36
Printer Support 1-37
General Information 1-40
References 1-42

Chapter 2 BioLynx - Peptide Sequencing 2-1
Contents 2-2
Figures 2-3
Overview 2-5
Peptide Fragmentation 2-5
The Scoring Algorithm 2-5
Creating a New Document 2-6
Specifying the Spectrum 2-6
Specifying Sequencing Parameters 2-7
Peptide Parameters Page 2-7
Sequence Parameters Page 2-9
Modifications Page 2-9
Digest Reagents Page 2-11
Analyzing the Spectrum 2-11
Troubleshooting 2-12
Isotopic Labelling 2-13
Peptide Sequencing Documents 2-14
The Main Window 2-14
Child Windows 2-15
The Sequence Pane 2-15
The Spectrum Pane 2-16
The Tree Views Pane 2-18
The Sequence Report Pane 2-19

The Toolbar 2-19
Menu Commands 2-21
File Menu 2-21
Report Options 2-21
View Menu 2-23
Options 2-24
Tools 2-24
Window Menu 2-25
References 2-26

Chapter 3 BioLynx - Database Searching 3-1
Contents 3-2
Figures 3-2
Overview 3-3
Using the Database Searching Program 3-3
Creating a New Document 3-3
Loading a Spectrum 3-4
Database Setup 3-4
Exclude from search dialog 3-6
Saving/Loading Query parameters 3-6
Carrying out a Search 3-7
The ProteinProbe™ Interface 3-7
ProteinProbe Documents 3-7
The Main Window 3-7
Child Windows 3-7
Toolbar 3-10
The Toolbar 3-11
Menu Commands 3-13
File Menu 3-13
View Menu 3-14
Options Menu 3-15
Window Menu 3-15
References 3-16

Chapter 4 BioLynx - CarboTools 4-1
Contents 4-2
Figures 4-2
Overview 4-3
Toolbars 4-3
Menu Commands 4-4
Creating a New Document 4-5
Configuration 4-5
Loading a Spectrum 4-6
Display Settings 4-6
Calculating Compositions 4-7
User Defined Modifiers and Terminal Groups 4-8
Manually Identifying Residues 4-9
Series Assignments 4-9
Configuring the Results Display 4-10
Deleting Results 4-10
Saving/Loading Parameters and Results 4-11

Chapter 5 BioLynx - Nucleic Acids 5-1
Contents 5-2
Figures 5-2
Overview 5-3
The Sequence Window 5-3
Display Options 5-3
The Nucleic Acids Toolbar 5-3

The Sequence Editor 5-4
Editing Text 5-6
Terminal/Linkage Group Modifications 5-7
Modified Bases 5-9
To Exit the sequence editor 5-11
Miscellaneous Editing 5-11
Sequence Name and Source 5-11
Clipboard 5-11
Calculations 5-12
Base Composition 5-12
Printer Support 5-13
References 5-14

Chapter 6 BioLynx - Oligonucleotide Sequencing 6-1
Contents 6-2
Figures 6-2
Overview 6-3
Oligonucleotide Fragmentation 6-3
The Sequencing Algorithm 6-3
Using the Oligonucleotide Sequencing Program 6-5
Accessing the Oligonucleotide Sequencer 6-5
Creating a New Document 6-5
Specifying the Spectrum 6-5
Specifying Sequencing Parameters 6-6
Analyzing the Spectrum 6-9
Troubleshooting 6-9
The Oligonucleotide Sequencing Interface 6-11
Oligonucleotide Sequencing Documents 6-11
The Main Window 6-11
Child Windows 6-11
Toolbar 6-13
Menu Commands 6-19
References 6-24

Chapter 7 ProteinLynx 7-1
Contents 7-2
Figures 7-2
Introduction 7-3
Automated Data Processing & Output 7-3
Automated Database Searching using ProteinLynx Global Server 7-3
Data Browser for Rapid Results Review 7-3
Overview 7-3
Output File Format 7-4
Data processing 7-5
Installation 7-6
Installing ProteinLynx on the Instrument PC 7-7

Chapter 8 ProteinLynx - Setup 8-1
Contents 8-2
Figures 8-2
Overview 8-3
Accessing ProteinLynx Setup 8-3
Getting Started 8-3
Specifying Setup Parameters 8-4
Instrument Type 8-4
Combine and Mass Measure 8-5
Peptide Selection & Centroiding 8-9
Select Output 8-11
Database Setup 8-14

Chapter 9 ProteinLynx - Results 9-1
Contents 9-2
Figures 9-2
Overview 9-3
Creating Report Files 9-3
Accessing ProteinLynx Results 9-3
The ProteinLynx Workspace Screen 9-3
The Plate Pane 9-4
The Workspace Toolbar 9-5
Getting Started 9-6
Menu Commands 9-7
Workspace Menu 9-7
Workspace Properties 9-7
Properties - General Page 9-7
Properties - Default Plate Page 9-8
Properties - Colors Page 9-9
Properties - Report Options Page 9-10
Properties - Sample and Results Pages 9-11

Chapter 1 BioLynx - Proteins

Contents

Figures 1-3
Overview 1-5
The Sequence Window 1-5
    Display Options 1-6
    Disulfide Bridge Editor 1-7
The Proteins Toolbar 1-8
The Chain Editor 1-9
    Editing Text 1-11
    Terminal Group Modifications 1-13
    Post Translational Modifications 1-15
    User-Definable Elements 1-17
    Miscellaneous Editing 1-18
Search Routines 1-18
    Residue Search 1-18
    Multiple Mass Search 1-19
    Find Residue Mass Combination 1-22
Digests 1-23
    Digest Simulation 1-23
Mass Spectral Fragments 1-28
Spectrum Interpretation 1-33
Hydropathic Profiles 1-36
Printer Support 1-37
General Information 1-40

Figures

Figure 1.1 Sequence window showing the amino acid sequence of trypsinogen with intact disulfide bridges 1-5
Figure 1.2 MassLynx Color and Font Editor 1-6
Figure 1.3 Display Options dialog 1-6
Figure 1.4 Disulfide bridge editor dialog 1-7
Figure 1.5 Chain Editor showing text editor, masses and composition 1-9
Figure 1.6 Elemental Composition dialog 1-10
Figure 1.7 Information dialog 1-10
Figure 1.8 Find and Replace dialog editor 1-12
Figure 1.9 Conversion of preproinsulin into insulin (Ref. 1) 1-12
Figure 1.10 Chain Editor showing A chain 1-13
Figure 1.11 Chain Editor showing A and B chains 1-13
Figure 1.12 Pentapeptide showing the amino and carboxy termini 1-14
Figure 1.13 The terminal group editor dialog 1-14
Figure 1.14 Create new terminal group dialog 1-15
Figure 1.15 Post-translational modification dialog editor 1-15
Figure 1.16 Create new post-translational modification dialog editor 1-16
Figure 1.17 User definable elements dialog 1-17
Figure 1.18 Chain name and Database source information 1-18
Figure 1.19 Residue search dialog 1-19
Figure 1.20 Residue search highlighted in sequence window 1-19
Figure 1.21 Mass search parameters 1-20
Figure 1.22 Entering a new mass 1-20
Figure 1.23 Editing the mass list 1-21
Figure 1.24 Results of a mass search 1-21
Figure 1.25 Result of mass search with residues highlighted in sequence window 1-22
Figure 1.26 Result of find residue combination 1-23
Figure 1.27 Entering parameters for automatic digest dialog box 1-23
Figure 1.28 Cleavage Parameters dialog 1-24
Figure 1.29 Peptide mapping of a tryptic digest of a beta lactoglobulin 1-26
Figure 1.30 Annotated chromatogram peaks indicating peptide fragment elution 1-27
Figure 1.31 List of digest fragments with digest residues highlighted in sequence window. 1-27
Figure 1.32 Mass Spectral Fragmentation dialog 1-28
Figure 1.33 Ion Display Setup 1-29
Figure 1.34 Match mass spectral fragment ions dialog 1-30
Figure 1.35 Matched mass spectral fragmentation ions showing customised ion series 1-31
Figure 1.36 Annotated spectrum displaying matched mass spectral fragment ions 1-32
Figure 1.37 Backbone and Side-Chain fragment ions according to a combination of the Roepstorff and Biemann nomenclatures (Ref. 4 and 5) 1-33
Figure 1.38 Spectrum Interpretation dialog box 1-34
Figure 1.39 MS sequinator results of unknown peptide 1-35
Figure 1.40 Hydropathic Profiles dialog 1-36
Figure 1.41 Hydrophobicity plot according to Kyte&Doolittle 1-37
Figure 1.42 Print report dialog box used for obtaining hard copies of various analyses 1-38
Figure 1.43 Charge State Parameters dialog 1-39

Overview

This Chapter provides an overview of **BioLynx** and its component programs. The Proteins, Nucleic Acids, Oligonucleotide Sequencer, CarboTools and ProteinProbe programs are accessed from the MassLynx Bar. All five programs can be active simultaneously and interact with other parts of the MassLynx data system allowing for intuitive and easier sample characterization.

The Protein/Peptide Editor is a biological software toolset used to analyze and manipulate protein or peptide sequences so as to facilitate the interpretation of data obtained from mass spectrometry. Various matching routines allow data to be automatically matched against known protein or peptide structures.

The Sequence Window

Selecting **BioLynx - Protein/Peptide Editor** from the MassLynx Shortcut bar brings up a sequence window. An empty chain with normal N-Terminal (amino terminus) and C-Terminal (carboxy terminus) groups will be displayed the first time this window is accessed, after this the last file used will be displayed.

The sequence window is one of several windows within BioLynx that uses a Multiple Document Interface (MDI) display allowing multiple windows (called documents) to be displayed simultaneously. The sequence window also features a toolbar located at the top of the window for easy access to common commands and a status bar to display useful information.

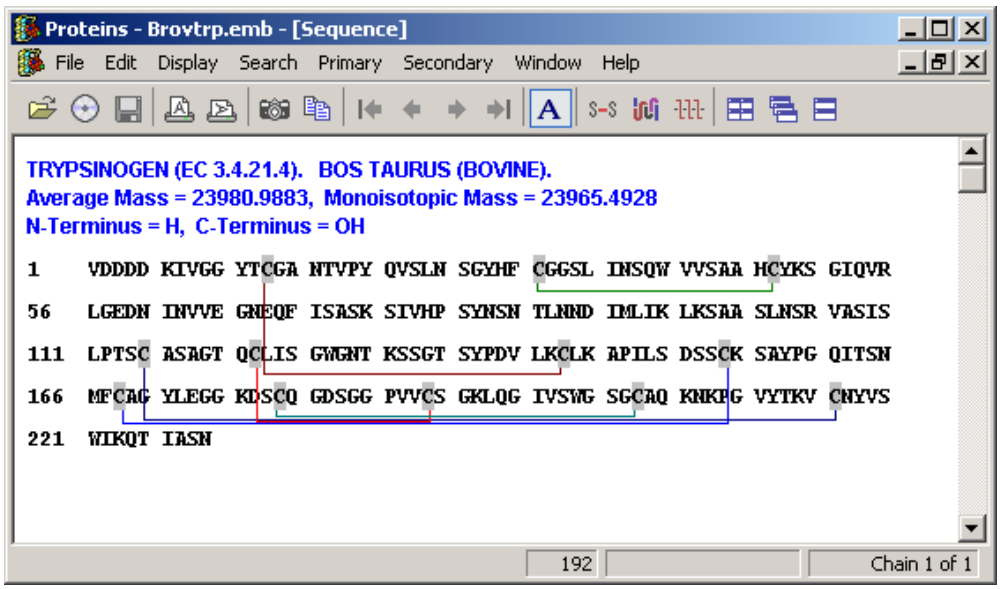


Figure 1.1 Sequence window showing the amino acid sequence of trypsinogen with intact disulfide bridges

Display Options

The font type and size can be customized for display purposes. This would be useful when wanting to capture for e.g. Figure 1.1 and paste the picture into an external software package. To customize the font type and size, select **Tools**, **Colors and Fonts** from the MassLynx top-level menu bar.

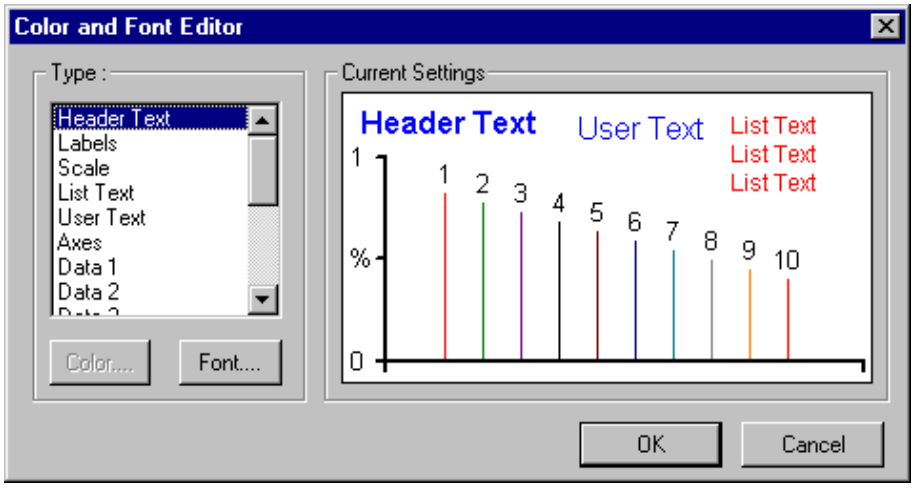


Figure 1.2 MassLynx Color and Font Editor

To amend the Protein/Peptide display, select **Display, Options** from the Protein/Peptide Editor menu.

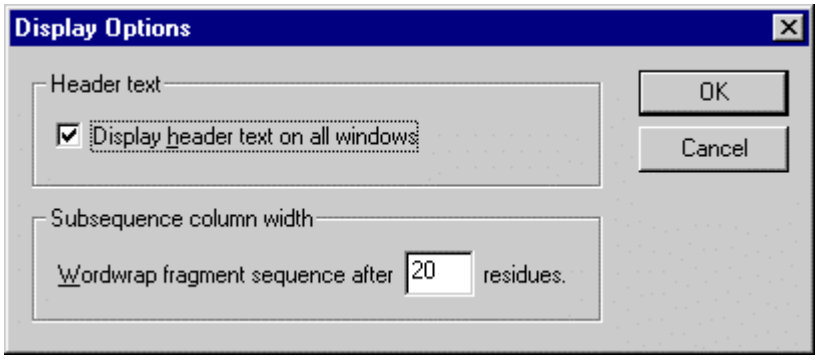


Figure 1.3 Display Options dialog

The header text displayed at the top of all windows can be removed which could be useful if you want to copy a sequence with disulfide bridges to an external application.

For digest fragment lists the displayed sub-sequence fragments will be word-wrapped after the specified number of residues.

Disulfide Bridge Editor

The Disulfide Bridge Editor allows you to add or remove intra-/intermolecular disulfide bridges from sequences. Intramolecular disulfide bridges are bridges between cysteine residues on the same chain, whilst intermolecular bridges are disulfide bridges between cysteine residues on separate chains.

The Disulfide Bridge Editor is a modeless dialog box, this means that you can continue working in the Sequence window while the Disulfide Bridge Editor dialog box is displayed.

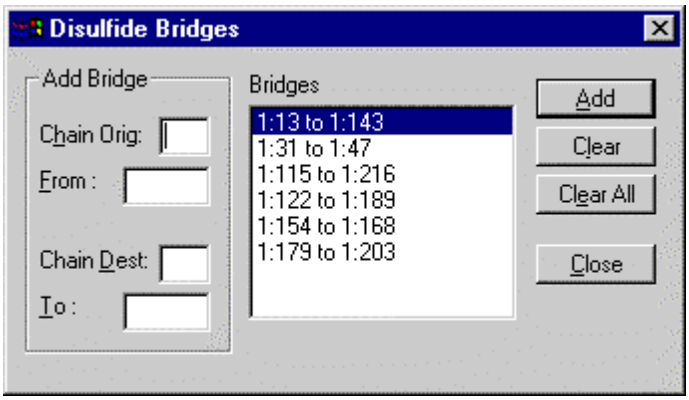


Figure 1.4 Disulfide bridge editor dialog

To Add a bridge

1. Select **Display, Disulfide bridge editor...** from the Proteins menu bar.

2. Enter the relevant chain and residue numbers in the edit boxes using the keyboard.

   -or-

   Click on the cysteine residue for the chain origin and again for the chain destination. These values will appear in the Disulfide Bridges dialog.

3. Press **Add** to enter the values in the list box.

4. Steps 2 and 3 can be repeated as required to add more bridges.

To View a bridge

To make disulfide bridges visible (oxidized state) or not (reduced) in the Sequence Display window toggle the

toolbar button or choose **Reduced Disulfide Bridges** or **Oxidized Disulfide Bridges** from the Proteins **Display** menu. The average and monoisotopic masses are adjusted to account for the loss of two hydrogens for each intramolecular bridge.

To Clear a Bridge(s)

1. Select **Display, Disulfide bridge editor** from the Proteins menu bar.
2. Select the disulfide bridge to be removed in the **Bridges** list box.
3. Press the **Clear** button to remove the disulfide bridge.
4. Steps 2 and 3 can be repeated as required
5. To remove all the disulfide bridges from the sequence press the **Clear All** button.

Pressing the Close button will close down the Disulfide Bridge Editor.

The Proteins Toolbar

The Toolbar is displayed at the top of the window and allows you to perform some common operations with a single click of the appropriate Toolbar button.

Press to open a peptide file.

Press to import an EMBL file.

Press to save file in peptide format.

Press to print report of selected analysis in portrait format.

Press to print report of selected analysis in landscape format.

Press to send a picture of current window to the clipboard.

Press to copy contents of the current window to the clipboard.

Press to display the first chain of a set of chains or press the **HOME** key on the keyboard.

Press to display the previous chain of a set of chains or press the ← key on the keyboard.

Press display the next chain of a set of chains or press the → key on the keyboard.

Press to display the last chain of a set of chains or press the **END** key on the keyboard.

Press to toggle between displaying residues in single or three letter code. This button operates for both the chain sequence and digest fragment sequence windows.

Toggle to display reduced/oxidized state of intra/intermolecular disulfide bridges.

Press to display the automatic digest dialog box.

Press to display the mass spectral fragments dialog box.

Press to arrange the windows in a tiled view.

Press to arrange the windows in a cascaded view.

Press to arrange the windows in a stacked view.

The Chain Editor

The chain editor allows basic text editing of the protein or peptide sequence with informative displays.

To Access the Chain Editor

Select **Edit**,**Current Chain** or **New Chain** (for multi-chain support)

-or-

Double click on the Sequence Display window.

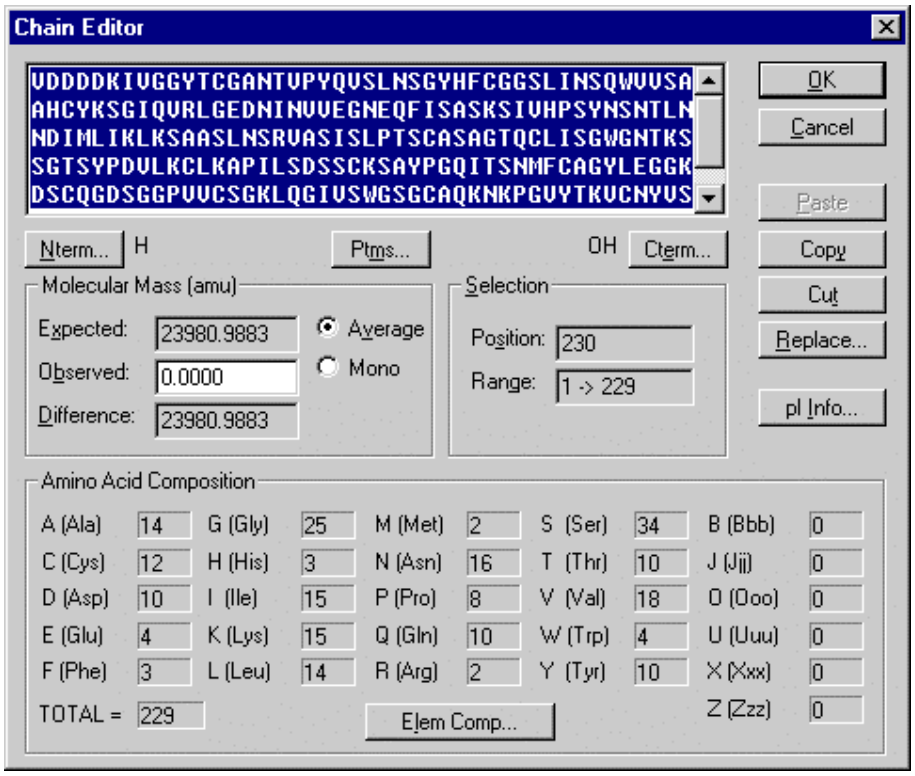


Figure 1.5 Chain Editor showing text editor, masses and composition

Molecular Mass

The **expected** mass is the mass of the highlighted part of the text and takes into account whether disulfide bridges are displayed in the Sequence window. An observed mass can be entered, using

the keyboard, in the **Observed** edit control. The **Difference** mass reflects the difference between the observed and expected masses

Selection

The position of the cursor in the text edit window is indicated in the **Position** control. If there is no text in the text editor this position will be 1. If text is selected i.e. highlighted, the range of the selection will be indicated in the **Range** control.

Amino Acid Composition

Shows the number of each amino acid present in the current chain including modified residues

Elem Comp Displays of the elemental composition of the sequence (Figure 1.6) Ten of the most common elements are displayed showing the number of carbons, hydrogens etc. These calculations are based on zero charge species and, as for the chain editor, take into account the number of disulfide bridges visible in the sequence display window.

The displayed formula can be copied into the **Isotope Model** program in **Spectrum** by highlighting the formula and selecting **Copy**.

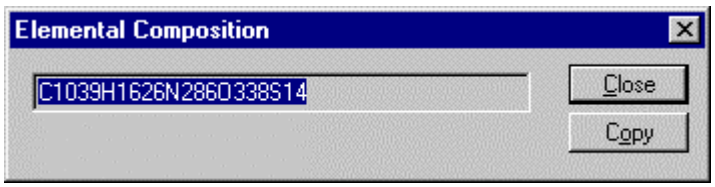


Figure 1.6 Elemental Composition dialog

pI Info Displays the Isoelectric Point (pI) value, HPLC Index, Bull&Breese values and Extinction Coefficient value at 280nm for the current chain. These values reflect residues highlighted in the chain editor. If no residues are highlighted the values are for the whole chain.

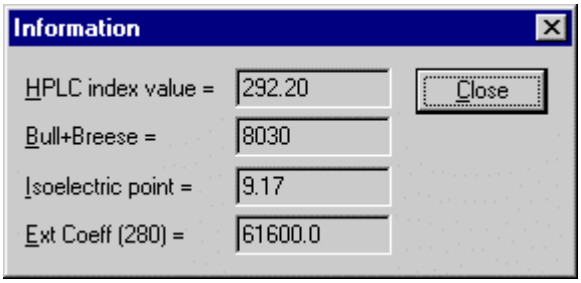


Figure 1.7 Information dialog

Cut, Copy, Paste and Replace Buttons used for editing text and are described in the next section

Nterm, Cterm and PTMs Buttons load the N-Terminal, C-Terminal and Post Translational Modification Group Editor.

Editing Text

Only alphabetical text can be entered in the edit window and the unassigned letters (B,J,O,U,X,Z) are represented as zero mass until assigned a user defined residue. Expected mass, difference, position and amino acid composition are constantly updated to reflect changes made to the current chain. When residues are highlighted the range is displayed as well as the appropriate mass plus H_2O or in the case of amino or carboxy terminal amino acids the appropriate terminal group.

To Add Text

Using the keyboard, place the cursor at the required position and type in single letter amino acid codes.

Using the mouse, place the cursor at the required position then click on the relevant amino acid in the **Amino Acid Composition** part of the Chain Editor window.

From the clipboard, place the cursor at the required position and press the **Paste** button. Text stored on the clipboard will be pasted into the Chain Editor.

To Delete text

Place the cursor at the required position and press the **BACKSPACE** key to delete the character to the left of the cursor, or the **DELETE** key to delete the character to the right of the cursor.

To delete a section of text, highlight the text and press the **DELETE** key.

To Select Text

Using the keyboard, place the cursor at the required position hold down the **SHIFT** key and press one of the arrow keys.

Using the mouse, place the cursor at the required position then click and drag across the required text. A double click in the Chain Editor, will select all the text.

Selected text can be deleted, cut or copied. Cut removes the text from the editor and places it onto the clipboard. Copy copies the selected text to the clipboard without removing it from the editor.

To Replace Text

Using the keyboard, select the text to be replaced and type in single letter amino acid codes.

Using the mouse, select the text to be replaced then click on the relevant amino acid in the **Amino Acid Composition** part of the Chain Editor window.

From the clipboard, select the text to be replaced and press the **Paste** button. Text stored on the clipboard will overwrite the current text in the Chain Editor.

Find and Replace Operation

Residues can easily be replaced by modified residues or highlighted to show their occurrence in a particular sequence.

- **Find Next** highlights the next occurring specified residue.
- **Replace** replaces the specified residue with the new residue.
- **Replace All** replaces all occurrences of a specified residue with the new residue.

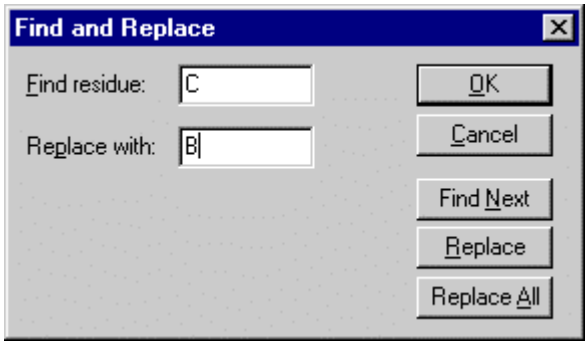


Figure 1.8 Find and Replace dialog editor

Example of use of the CUT, COPY and PASTE Commands

Many proteins imported from databases are precursors of the active molecules. Proinsulin is the biosynthetic precursor of insulin. Proinsulin is a single polypeptide chain containing a sequence of about 30 residues that is absent from insulin. Proinsulin is not the earliest form of the hormone. The nascent polypeptide chain, called preproinsulin, contains additional residues at the amino terminus known as the signal sequence.

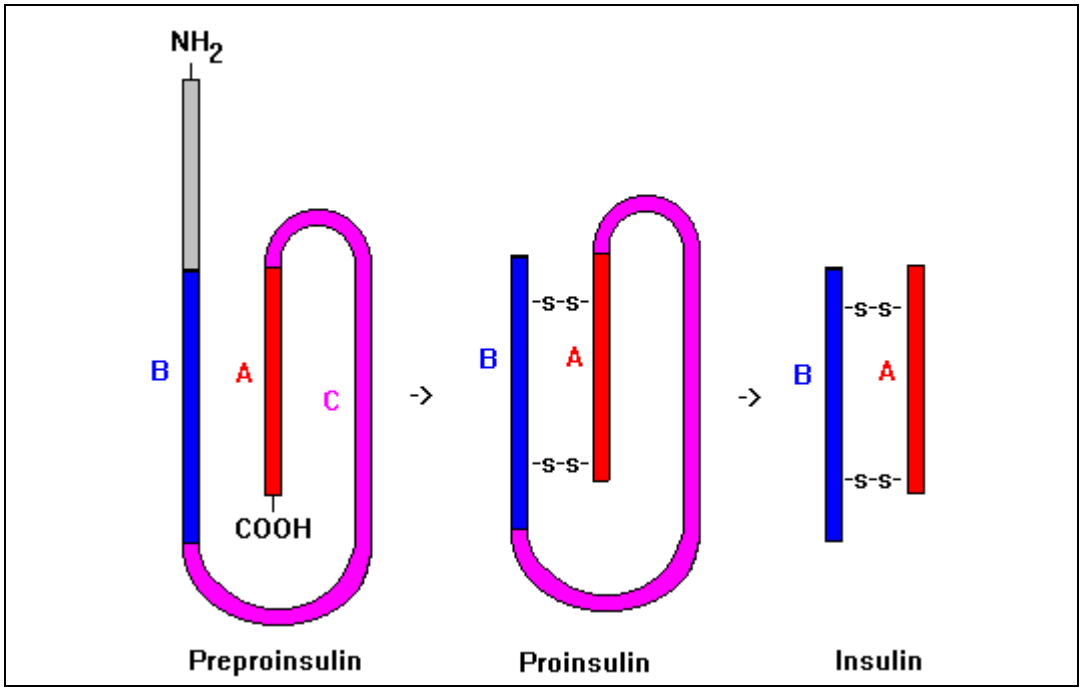


Figure 1.9 Conversion of preproinsulin into insulin (Ref. 1)

1. The signal sequence of 24 residues of preproinsulin can be removed by highlighting the residues in the chain editor (1-24) and choosing **Cut** or **Delete** or **Backspace** from the keyboard.

2. The remaining residues (proinsulin) consist of chain **B** (1-30), a connecting peptide (31-65) and chain **A** (66-86). The connecting peptide can similarly be removed by highlighting the relevant residues (31-65) and choosing **Cut**.

3. Chains **A** and **B** constitute the insulin molecule. These are separated by highlighting either residues 1-30 or 31-51 depending on which chain will be chain **1** or chain **2**. To make chain **A** chain **1** highlight residues 1-30 and choose **Cut**.

4. Press **OK** to exit the chain editor.

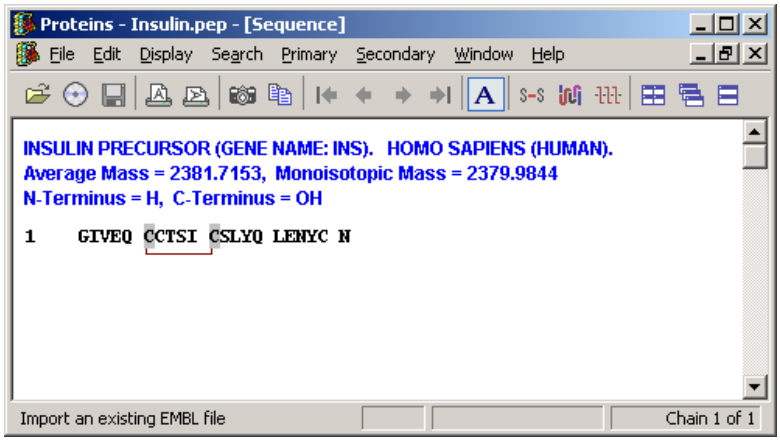


Figure 1.10 Chain Editor showing A chain

5. The status bar will display chain 1 of 1 and chain **A** will be displayed in the sequence window. Choose **Edit New Chain** from the Proteins menu this invokes Figure 1.5.

6. In the empty chain editor choose **Paste** to copy the contents of the clipboard into the chain editor i.e., chain **B**.

7. Press **OK**.

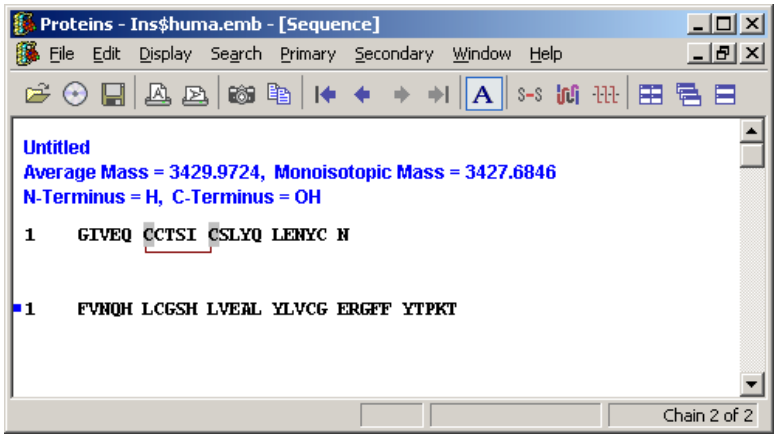


Figure 1.11 Chain Editor showing A and B chains

8. The status bar now displays Chain 2 of 2. The Sequence window displays both chain **A** and chain **B** and the toolbar buttons become enabled to allow navigation through the chains.

Terminal Group Modifications

The default amino and carboxy terminal groups are H and OH respectively, these can be modified by editing and creating new groups.

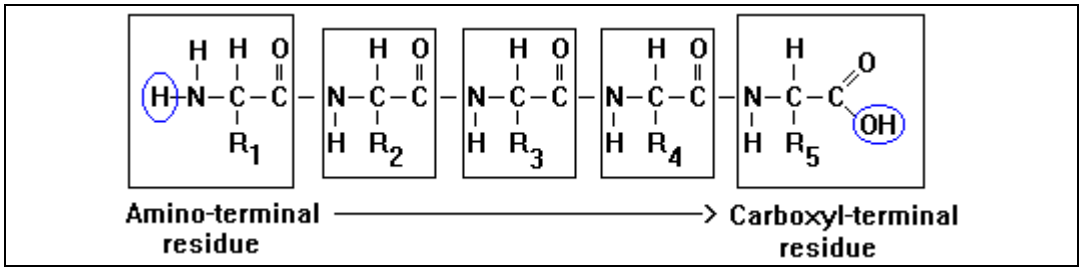


Figure 1.12 Pentapeptide showing the amino and carboxy termini

To Modify the Terminal Group

Click on the **Nterm… or Cterm…** button in the Chain Editor (Figure 1.5) this invokes the N – Terminal groups dialog box (Figure 1.13) or the C- Terminal Groups dialog box

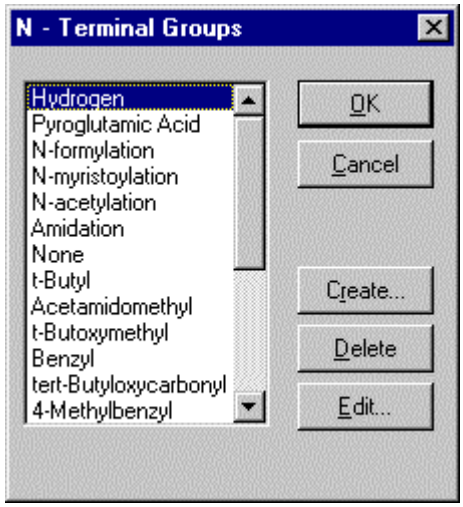


Figure 1.13 The terminal group editor dialog

1. Press the **N-term** or **C-term** button.
2. Click on the required terminal group in the list box.
3. Press **OK** to modify the terminal group.
4. The chain editor mass displays are automatically updated.

To Create a New Terminal Group

1. Press the **Create** button, this invokes Figure 1.14.
2. Enter a unique name for the group (maximum 50 characters).
3. Enter a unique symbol e.g. H for Hydrogen (maximum 50 characters). Press the **User elements...** button to allow for isotopes such as deuterium etc.
4. Enter a formula using IUPAC nomenclature e.g. Cl for Chlorine.
5. Press **OK** to enter group in database. The new group is displayed in the list box.

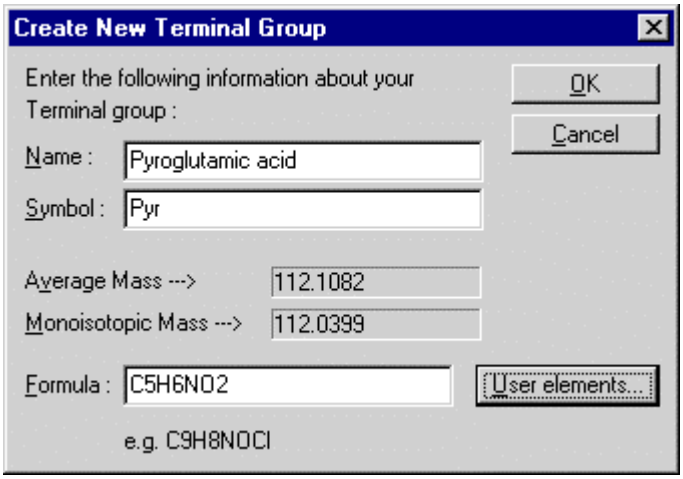


Figure 1.14 Create new terminal group dialog

To Edit a Terminal Group

1. Select the required terminal group in the list box.
2. Press the **Edit** button.
3. Change parameters and press the **OK** button.

To Delete a Terminal Group

1. Select the required terminal group in the list box.
2. Press the **Delete** button.
3. Press **YES** to confirm deletion.

Post Translational Modifications

To modify a residue or create a post-translational modification, press the **PTM** button in the chain editor.

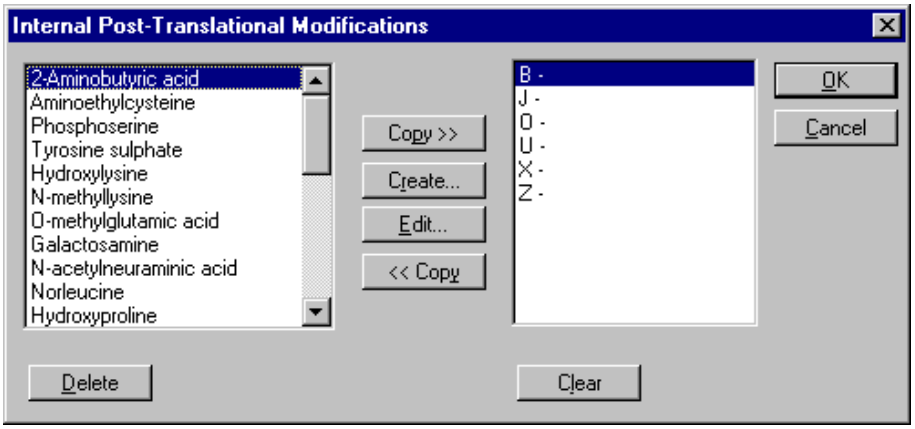


Figure 1.15 Post-translational modification dialog editor

Modified residues are listed in the list box with the input focus on the currently selected modified residue.

To Create a Modified Residue

1. Press the **Create** button.
2. Enter a unique name for the group (maximum 50 characters).
3. Enter a unique **three-letter** symbol.
4. Enter a formula using IUPAC nomenclature or leave blank for cyclic peptides.
5. Press the **OK** button.

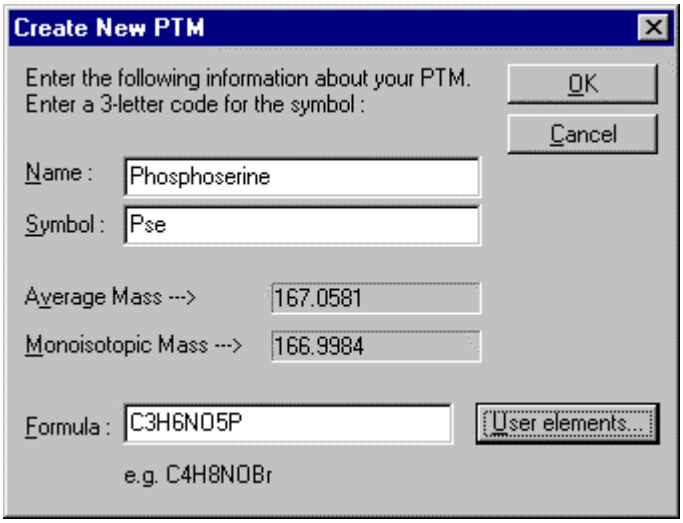


Figure 1.16 Create new post-translational modification dialog editor

To Edit a Modified Residue

1. Select the required residue in the list box.
2. Press the **Edit** button.
3. Change parameters and press the **OK** button.

To Delete a Modified Residue

1. Select the required residue in the list box.
2. Press the **Delete** button.
3. Press **YES** to confirm deletion.

To Copy a Modified Residue to Modified Residue List

1. Select the residue you want to copy, in the residue list box.
2. Select one of the unused letters (B-Z).
3. Press **Copy >>**.
4. The residue name appears next to the selected letter. Press **OK** to confirm the copy.

To Copy a Modified Residue from Modified Residue List to Database

This allows modified residues created on an alternative data system to be copied onto the database.

1. Copy the required file into the Pepdata folder using NT Explorer.
2. Choose **Open** from the Proteins Sequence **File** menu, or press the toolbar button, and select the required file. The filename will appear next to the first unused letter.
3. Select the modified residue from modified residue list.
4. Press **Copy <<** and the modified residue will appear in the residue list box.
5. Press **OK** to confirm the copy.

To Clear a Modified Residue(s)

This clears the residue associated with a letter.

1. Select the modified residue from modified residue list.
2. Press **Clear**.
3. Press **OK** to enter the modified groups into the database.
4. The chain editor mass displays are automatically updated.

User-Definable Elements

Other elements can be defined from the Create New Terminal Group and Create New PTM dialog boxes (Figure 1.14 page 1-15) from the Molecular Mass calculator (available on the MassLynx ToolsShortcut Bar).

Defining user elements

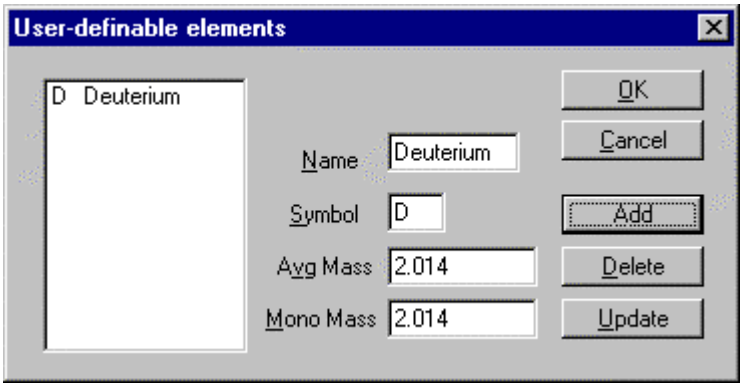


Figure 1.17 User definable elements dialog

1. Press **User elements** button.
2. Enter the parameters and press **Add** to enter the group in the list. **Update** can be used to edit a particular element or group. **Delete** removes the highlighted group in the list box.
3. Up to 10 elements, isotopes, molecules can be added to the list.

4. The list is saved in the masslynx.ini file for future use on pressing **OK**.

Miscellaneous Editing

Sequence Name and Source

The name of the current chain can be edited by choosing **Edit Chain Name**. Insulin consisting of two chains could be named the **A-chain** for the first chain and the **B-chain** for the second chain. The maximum length allowed for a name is 100 characters and, although it can be the same, should not be confused with the **File Save As** name given to the sequence for disk storage. Files imported from the EMBL will have information displayed as in Figure 1.18. Any modifications to sequences can be entered in the **Sequence Modifications** edit control. This is displayed on hard copy formats.

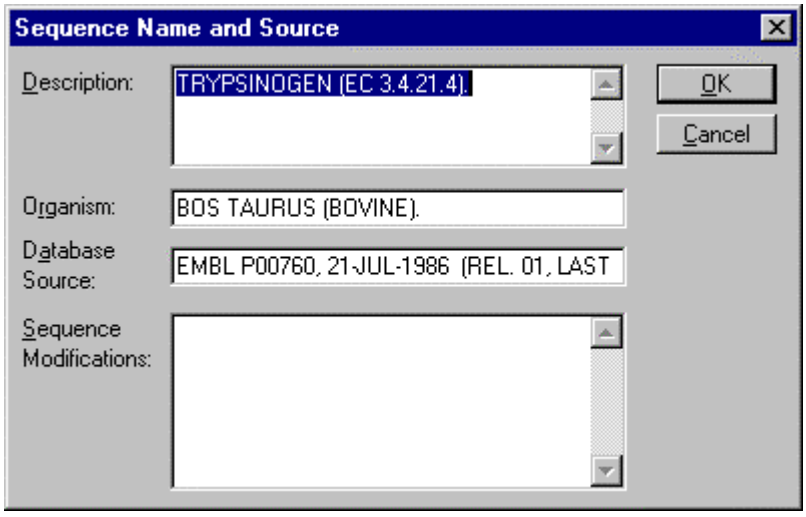


Figure 1.18 Chain name and Database source information

Clipboard

The currently active MDI window can be copied to the clipboard as an enhanced metafile by choosing **Edit, Copy** from the menu or pressing the toolbar button.

Edit, Copy, List or the toolbar button copies the contents of the currently active window to the clipboard i.e. lists of multiply charged digest fragments or sequences can be copied to the clipboard.

Search Routines

Residue Search

To carry out a search for a sub-sequence of residues choose **Search, Find Residue** this invokes the Residue Search dialog box (Figure 1.19).

Three sub-sequences with maximum of 100 (single-letter code) residues can be entered in each box. Any unknown residues can be replaced with a '**?**'. Press **OK** to carry out the search.

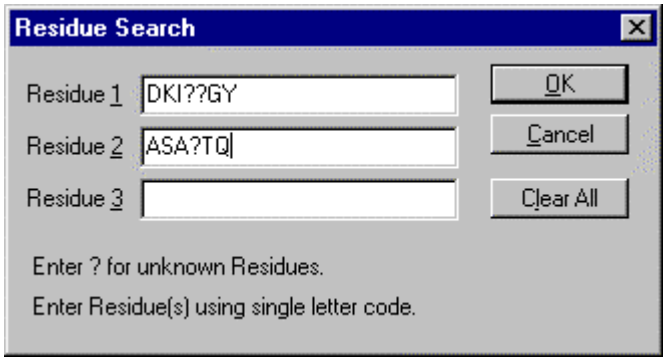


Figure 1.19 Residue search dialog

All matches between the sub-sequences of interest and the sequence are highlighted in the "Sequence" and "Digest fragment" display windows (Figure 1.20).

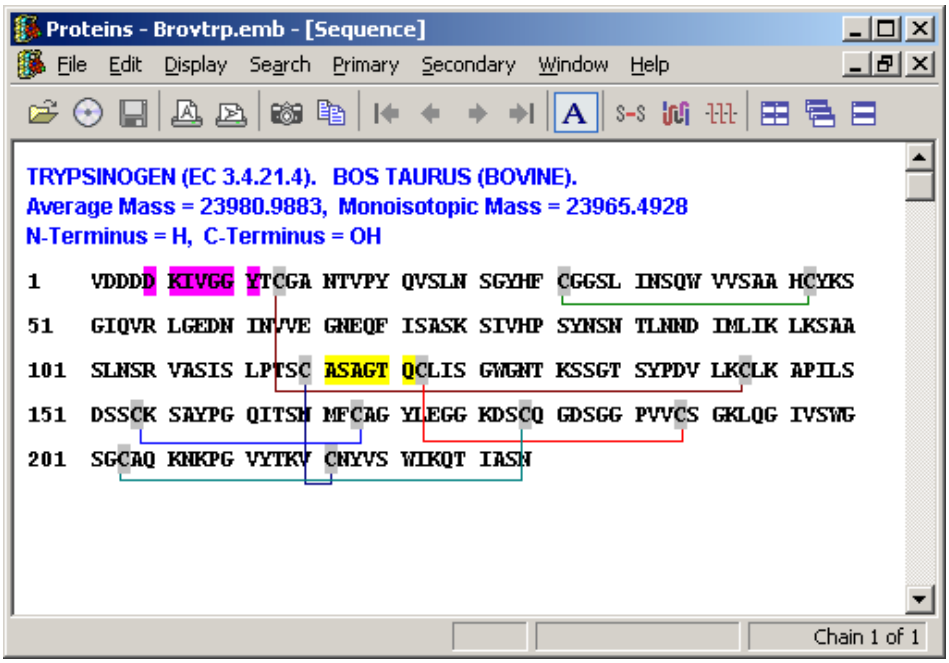


Figure 1.20 Residue search highlighted in sequence window

Multiple Mass Search

To carry out a mass search for a sub-sequence or sub-sequences linked by disulfide bridges select **Search, Find Mass**. This utility is useful for identifying and matching components in a spectrum that are the result of a partial cleavage i.e. a tryptic digest that has not resulted in cleavage at all specified sites.

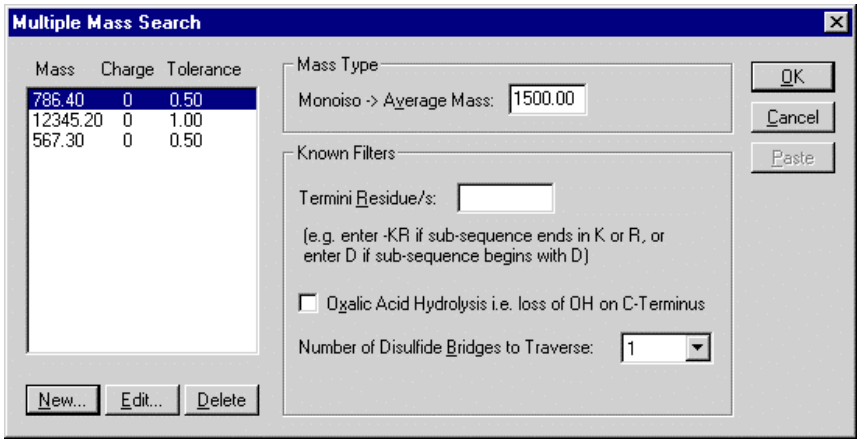


Figure 1.21 Mass search parameters

Mass Search Parameters

The default mass changeover from monoisotopic to average is 1500. The actual mass being searched is calculated based on the entered mass and charge. Known filters such as termini, whether or not a loss of OH as a result of hydrolysis should be allowed and the number of disulfide bridges to traverse can all be specified.

If the disulfide bridges are not displayed in the sequence window then the **number of disulfide bridges** option will be grayed out. The more bridges the program has to traverse the longer the processing time.

Entering a List of Masses or Single Mass

There are three methods of entering a list of masses. You can enter the specified mass by pressing the **New** button and entering the mass, charge state and tolerance. The charge state and tolerance are defaulted, to 0 and 1.0 respectively, if not entered.

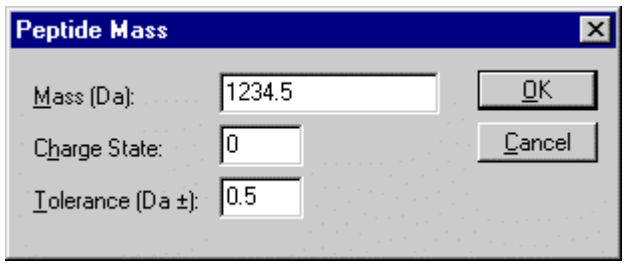


Figure 1.22 Entering a new mass

-or-

A list of masses can be entered by right clicking on/near peaks of interest in the spectrum. These masses are copied directly into the list box in the mass search dialog box. The charge state and tolerance are defaulted to 0 and 1.0 respectively.

-or-

From the Spectrum display a list of peaks can be copied to the clipboard and pasted into the mass search list box by pressing the **Paste** button.

Editing Masses

The mass, charge state or tolerance of a single mass can be modified by clicking on the mass in the list box, and pressing the **Edit** button.

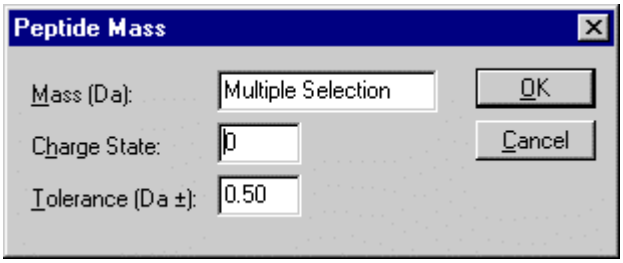


Figure 1.23 Editing the mass list

The **Charge State** and **Tolerance** of a list of masses can be changed simultaneously. and drag down the list of masses then press the **Edit** button. The mass field is displayed as Multiple Selection, change the **Charge State** and **Tolerance** as required.

Deleting Masses

Highlight the mass or list of masses and press the **Delete** button.

Carrying Out a Search

Press **OK** to carry out the search. A status box indicates processing and allows the operation to be cancelled at any time. If more than one mass has been entered the status box indicates the current mass being searched for. The output is displayed as below on a per mass basis. The **assignment** column indicates S-S linkages for fragments joined by disulfide bridges.

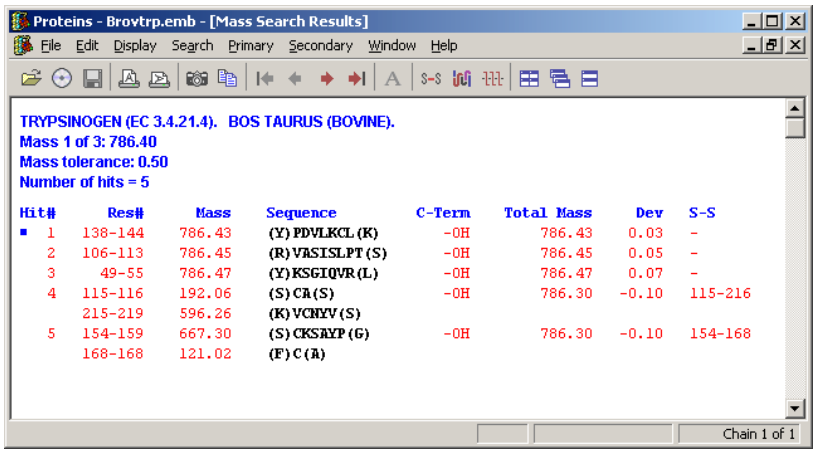


Figure 1.24 Results of a mass search

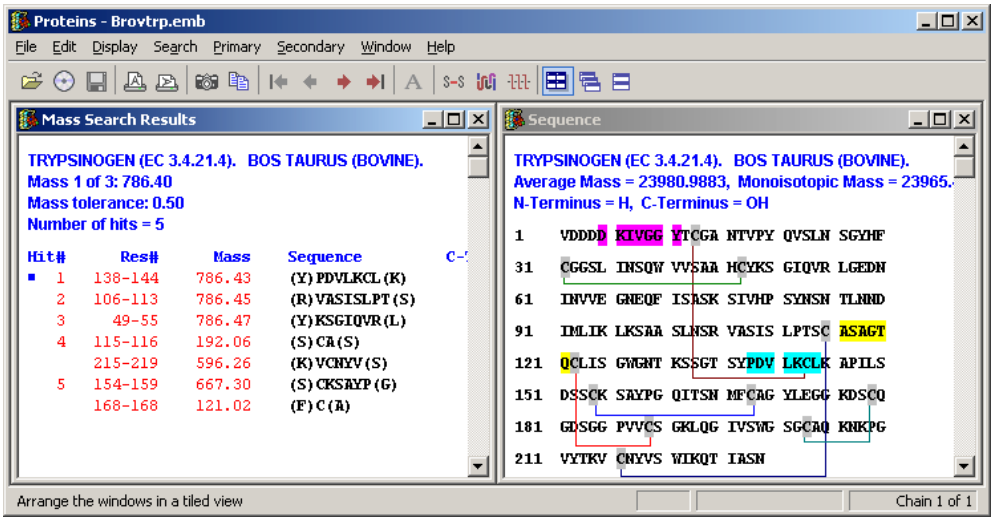


Figure 1.25 Result of mass search with residues highlighted in sequence window

The best-fit mass search result is listed at the top i.e. smallest deviation, with a maximum of 100 results in the list.

Viewing the Mass Search Results

To view the Mass Search Results and the Sequence window on the same screen, choose **Window, Tile**, **Cascade** or **Stack** from the menu or the corresponding toolbar buttons.

To view a selected hit in the sequence window click on one of the hits in the Mass Search Results window. A small square appears to the left of the selected hit and the sequence is highlighted in the Sequence window.

Viewing the Results of Multiple Mass Searches

To view results for a hit you can:

- Click the buttons to display the first, next and last mass of the searched masses.
- Use the arrow keys on the keyboard.
- Click on the hit in the list.

Find Residue Mass Combination

To perform a residue mass combination search, choose **Search, Find Residue Mass Combination**. For a given mass this routine displays all possible combinations of the 20 common amino acids.

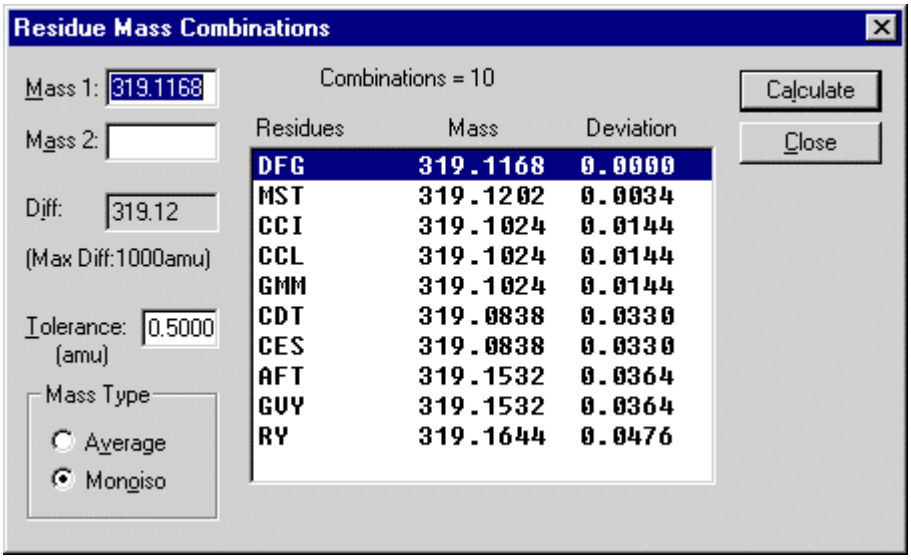


Figure 1.26 Result of find residue combination

Enter a mass, a tolerance (default 0.5) and either average or monoisotopic mass. A right mouse click on/near a peak of interest in Spectrum will enter the relevant mass in **Mass 1**, clicking again will enter a second mass in **Mass 2**. Press **Calculate,** or the enter key if the cursor is on mass1 or mass2, to perform the calculation.

A maximum of 100 combinations is sorted in the list with the smallest deviation at the top of the list.

Digests

Digest Simulation

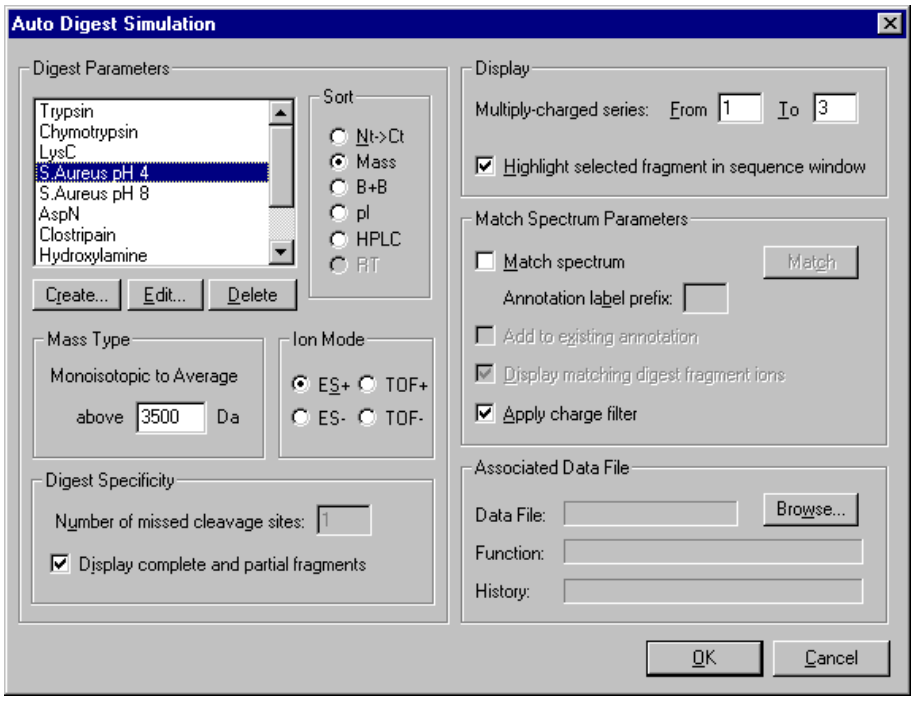


Figure 1.27 Entering parameters for automatic digest dialog box

Choose **Auto Digest Simulation** from the **Primary** menu or press the toolbar button to invoke Figure 1.27.

A list of cleavage reagent rules is provided which can be edited. New reagent rules can be added to the list by pressing the **Create** button. Pressing the **Edit** button brings up a dialog as displayed in Figure 1.28. The **Delete** button deletes a set of rules and removes the reagent from the list.

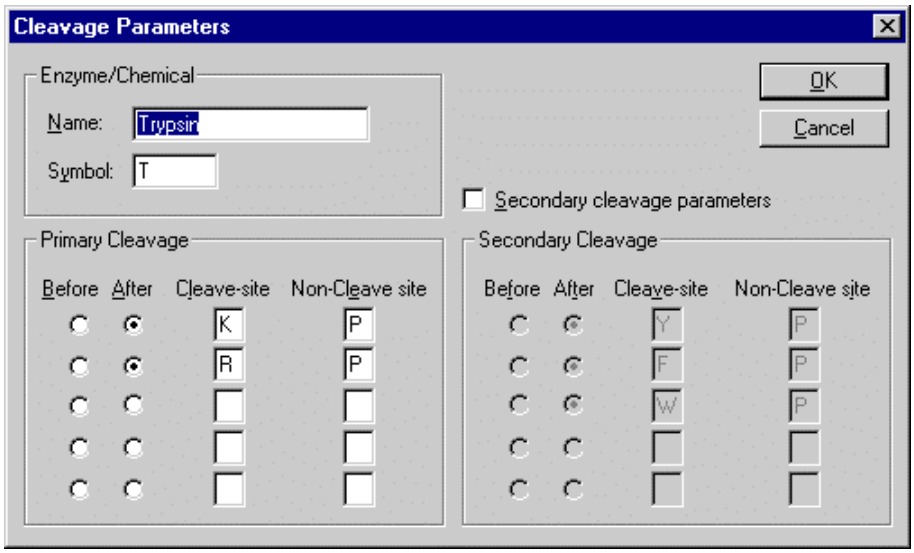


Figure 1.28 Cleavage Parameters dialog

To Create a New Set of Rules

1. Press the **Create** button.
2. Enter a unique name for the reagent.
3. Enter a unique symbol (1-letter). This is the letter, which appears before the fragment number in the Digest Fragment List.
4. Enter the residues to be cleaved or not cleaved.
5. Select cleave before or after for each cleave rule.

The **Secondary** check box allows for enabling secondary digests i.e., trypsin behaving like chymotrypsin or for simulating two digests one after the other. To cleave between two residues select the before and after buttons of adjacent residues.

Simulated Digests do not modify residues. e.g. CnBr digest normally converts Met into homoserine lactone or homoserine. This facility has not been implemented in this version. To modify Met into homoserine or homoserine lactone use the **PTM** button in the chain editor before carrying out the digest. This similarly applies to Cys Cyanylation or any other modification prior to digestion.

Entering Digest Parameters

1. Select the required cleavage type from the list box.

2. Choose the method of sorting

   Nt->Ct From amino terminus (amino acid number 1) to carboxy terminus.

   Mass In ascending mass**.**

   B&B (Bull&Breese) In ascending Bull and Breese free energy value (Ref. 2) which is a measure of the partition of a peptide between an aqueous and hydrophobic phase. Peptides with a highly negative B&B value tend to give stronger protonated molecular ions than those with a positive or less negative value.

   PI In ascending isoelectric point which is the pH at which the peptide has zero charge.

   HPLC In ascending HPLC value according to Browne, Bennet and Solomon (Ref. 3) which is an indication of the retention of the peptide on a RP-HPLC column. Values correspond to the percentage of acetonitrile required to elute the peptide from a C18 micro-Bondpak column using a TFA:water:acetonitrile buffer system. Values are based on the above authors system and will differ depending on the column and conditions used.

   RT This feature is enabled once digest targeting has been carried out. Fragments are sorted according to their corresponding retention time identified from the chromatogram peaks.

3. Choose the mass type i.e., **average**, **monoisotopic** or **both**. Enter a changeover mass if **both** is selected. Average masses are used above the changeover mass.

4. Choose the ion mode i.e., **ES+, ES-, Tof+** or **Tof-**. Choosing the **ES** ion mode results in the display of multiply-charged ion masses. **Tof** displays singly-charged ion masses with the HPLC, B&B and pI values for each peptide fragment.

5. Enter a number for the **Number of missed cleavage sites**. Using Trypsin as an example:

   Entering **1** results in cleavage at all specified sites (ACK / EGYR / SALPK / DLK / LSD).

   Entering **2** results in partial cleavage i.e., combining two fragments (ACKEGYR / EGYRSALPK / SALPKDLK / DLKLSD).

   Entering **3** results in partial cleavage i.e., combining three fragments (ACKEGYRSALPK / EGYRSALPKDLK / SALPKDLKLSD).

   If the **Display all combinations** check box is on then all possible combinations of complete and partial fragments, up to and including the **combination** value, will be displayed. Similarly if the **Display all combinations** check box is off then only the selected combination will be displayed.

6. Enter the **From** and **To** parameters for the **multiply charged series** to be displayed i.e., entering 3 to 8 results in triply charged masses up to and including 8 charges to be displayed for the respective peptide fragments.

7. Choose whether you want to highlight the digest fragments in the sequence display window. A tick in the check box indicates highlighting.

Digest Matching and Spectrum Annotation Parameters

Digest fragment matching against a spectrum (Electrospray or M@LDI-TOF) is possible. Also multiple sequence files can be matched against the same spectrum.

1. To carry out matching check the **Match Spectrum** box and press the **Match** button. Specify a threshold and mass error (tolerance) and press **OK**.

2. If multiple sequences are being matched to the same spectrum check the **Add to existing annotation** checkbox. Also to prefix the annotation on the spectrum add a letter e.g. **a** or **b** to the **Annotation label prefix** box.

3. The **Display matching digest fragment ions** checkbox will display matching multiply-charged ion fragment masses in bold text.

4. **Apply charge filter** is available for 3 or more charges. A rule is applied to discriminate against arbitrary matches based on the composition of the peptide under investigation. E.g. small peptides with only 2 basic sites are unlikely to hold 5 charges.

5. Select a **centroided** data file from the **Spectrum Data Browser** using the **Browse** button or right click on a centered or MaxEnt Sequence spectrum.

Digest Targeting and Chromatogram Annotation

Matching of digest fragments analyzed by LC/MS is carried out from chromatogram - see the **Components - Auto Find and Edit Worklist** features in the Chromatogram section of the MassLynx User Guide. In future releases the matching of digest fragments will be a two-way process involving looking in the data and using known digest fragment masses. Select a data file from the **Chromatogram Data Browser** using the **Browse** button or right on the specified chromatogram. The selection is displayed in the dialog box and is required for displaying retention time, observed masses for matching components and digest fragments.

To Carry out a Simulated Digest

Press the **OK** button in the Automatic digest dialog box.

-or-

Double click on the selected cleavage reagent in the list box.

A list of digest fragments with their expected multiply-charged masses are displayed as well as evidence for identifying its elution time (from component worklist in chromatogram).

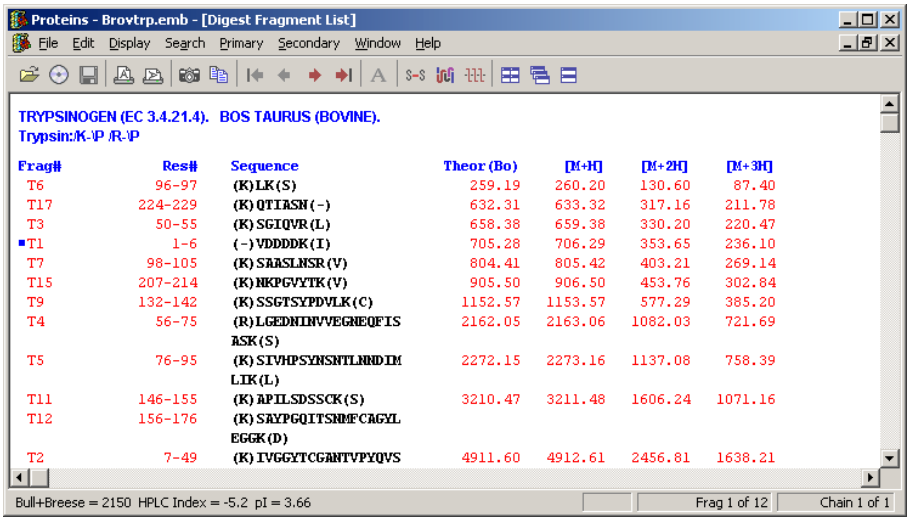


Figure 1.29 Peptide mapping of a tryptic digest of a beta lactoglobulin

If a data file has been selected using the **Browse** button then the name of the data file and the scanned mass range appears at the top of the digest fragment list window. It is possible to display **summed mass chromatograms** of individual digest fragments by right on a selected digest fragment. The uppermost trace in Figure 1.30 was generated in this way.

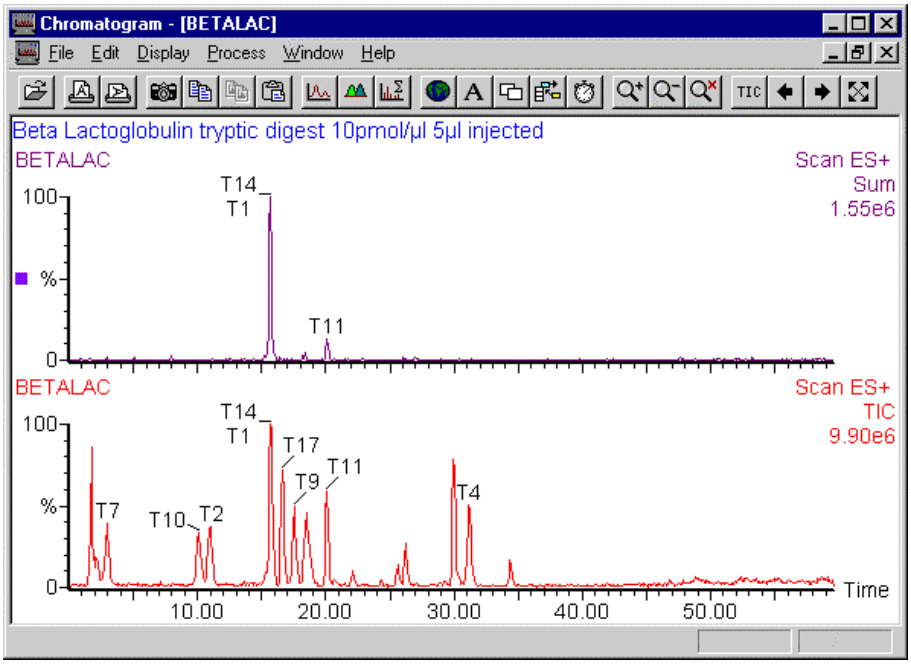


Figure 1.30 Annotated chromatogram peaks indicating peptide fragment elution

Peaks, on the chromatogram, are annotated with the fragment digest label if this type of annotation has been selected (for more information see the Chromatogram Chapter of the MassLynx User Guide).

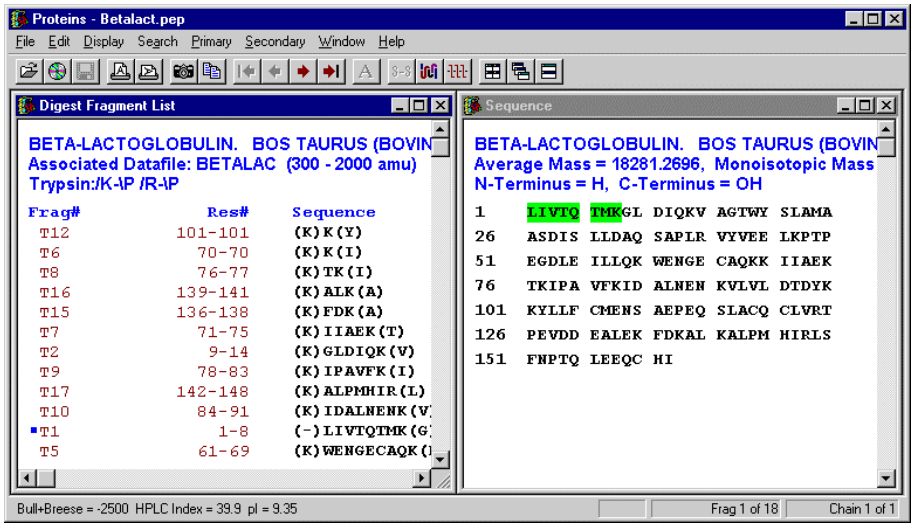


Figure 1.31 List of digest fragments with digest residues highlighted in sequence window.

Viewing the Sequence of the Currently Selected Digest Fragment

Choose **Window, Tile** to view the highlighted fragment in the main sequence window. The HPLC index, pI and B&B values for the selected fragment are displayed on the status bar.

Viewing Sequences of Other Digest Fragments

- Click the buttons to display the first, next and last fragment
- Use the arrow keys on the keyboard.
- Select the fragment.

The number of the currently selected digest fragment and the total number of fragments are displayed on the status bar i.e. **Frag 3 of 32**. A small square, to the left of the Fragment Label, indicates the currently selected digest fragment. If another fragment is selected all windows are updated with the newly selected digest fragment details.

Mass Spectral Fragments

A simulated MS-MS fragmentation or Cone Voltage Fragmentation (CVF) pattern of a peptide is useful for identifying an ion series and hence the structure of a particular molecule.

Select **Primary, Mass Spectral Fragments** or press the toolbar button to activate the dialog box.

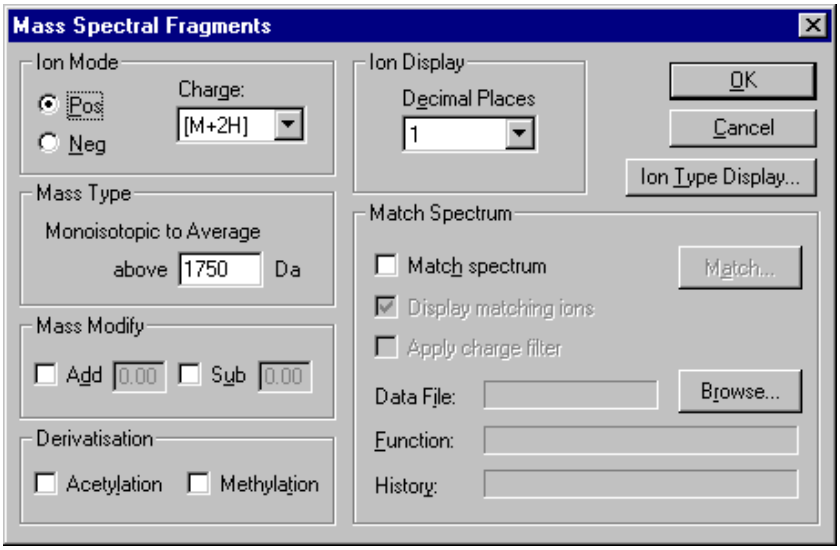


Figure 1.32 Mass Spectral Fragmentation dialog

Ion Display Setup

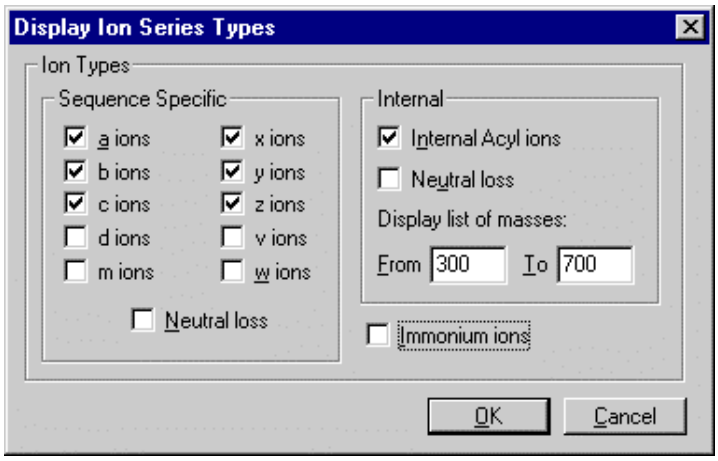


Figure 1.33 Ion Display Setup

1. Press the **Ion Type Display** button to select ion types to be displayed.

2. Select the sequence ion types to be displayed. 'b' and 'y' ions are displayed by default if none are selected. The **Neutral Loss** ions are NH_3 and H_2O.

3. Select **Internal Acyl ions** - these ions result from fragmentation at either end of the molecule. **Neutral Loss** ions are NH_3, H_2O and CO. Enter the mass range (default of 300 - 700). This is useful for large protein sequences.

4. Press **OK**.

Automatic Matching and Spectrum Annotation

1. Choose the ion mode **ES+** or **ES-** and the **Charge** to be taken into account in the calculation. Up to M+20H or M-20H can be selected.

2. Enter a changeover mass for the Mass Type.

3. Choose whether to **add** and/or **subtract** a mass from all sequence ions. This is used for cases when a group is lost prior to fragmentation resulting in all sequence ions being out by a specific mass.

4. Choose whether you want to derivatize specific groups. **Acetylation** acetylates the N-Terminal amino group if it is not already acetylated and results in acetylation of all lysine residues (increase of 42 amu). **Methylation** methylates the C-Terminal carboxy group if it is not already methylated and methylates all aspartic and glutamic residues (increase of 14 amu).

5. Select the number of **Decimal Places** required for the display (1 to 4).

6. If spectrum matching is required the **Match Spectrum** check box should be checked, this enables the **Match** button.

7. Press the **Match** button to invoke the **Spectrum Parameters** dialog.

   Enter a **mass window** (tolerance) value. The default is 1 amu.

   Enter a **threshold** for the matching process. This value specifies a minimum intensity of peaks that the program considers. It is either specified as a **percentage** or actual **intensity** of the intensity of the most intense peak in the spectrum.

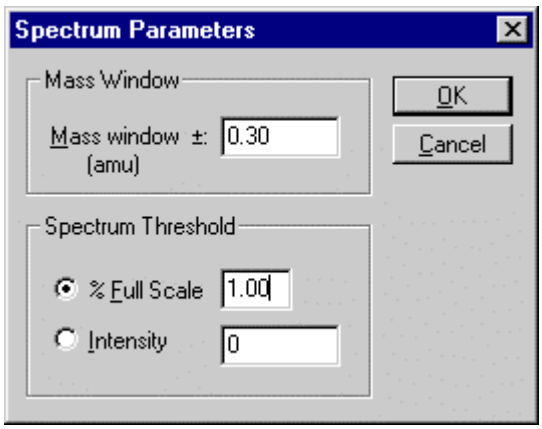


Figure 1.34 Match mass spectral fragment ions dialog

8. The **Display matching ions** check box works in conjunction with the **Match Spectrum** check box and should be checked if you want to view matching ions.

9. The **Apply charge filter** check box is only enabled for charge states of 3 and greater. This filter uses the number of acidic or basic residue rule and discriminates against coincidental matches.

10. To select a spectrum for matching press the **Browse** button and select a file, or a right mouse click on a displayed spectrum carries out the same operation. Matching can only be carried out on centroided or MaxEnt Sequence data files.

11. Press **OK** to display the mass spectral fragment ions.

If matching has been selected then a **status box** is displayed which provides an update on progress and allows the operation to be halted. All possible mass to charge ratio ions are compared to the original data including the selected charge state.

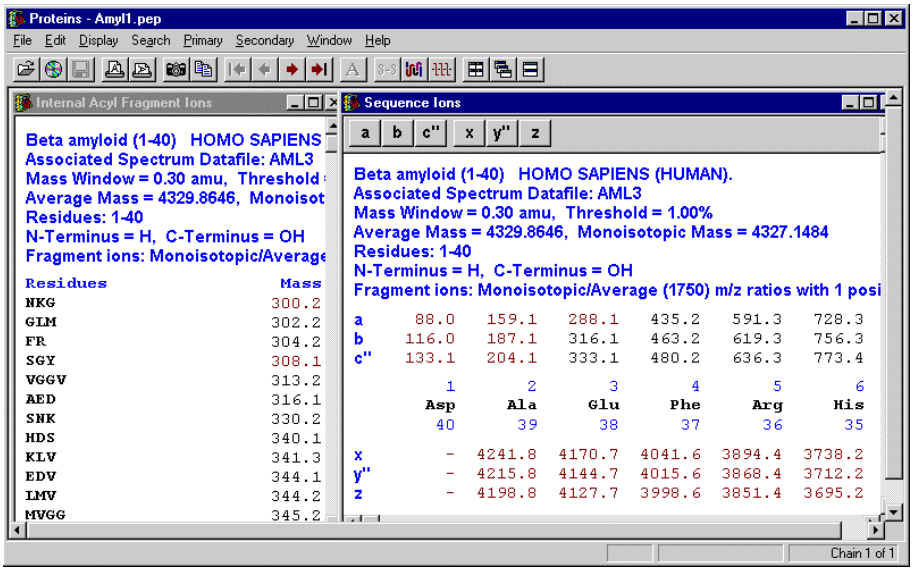


Figure 1.35 Matched mass spectral fragmentation ions showing customised ion series

Predicted fragment ions that match observed mass spectral fragment ions are displayed in black text in the sequence and internal acyl ions windows of BioLynx.

To View a Different Mass To Charge Ion Ratio

Click one of the toolbar buttons or use the arrow keys on the keyboard when either the sequence ion window or the internal acyl window is current.

Annotation of the relevant spectral peaks within the spectrum windows is carried out simultaneously if this type of annotation has been requested (see Controlling the appearance of peak labels in the Spectrum chapter). The labels used are based on the Roepstorff / Biemann nomenclature and are stored with each raw data file as **.ion**, **.int** and **.tab** files. This means that annotation of a particular spectrum need only be carried out once.

Simulated mass spectral fragmentation can be carried out on synthetic peptides or digest fragments. If a digest has been carried out then the observed fragmentation pattern will be of the currently selected digest fragment. Only digest fragments with reduced disulfide bridges are currently supported.

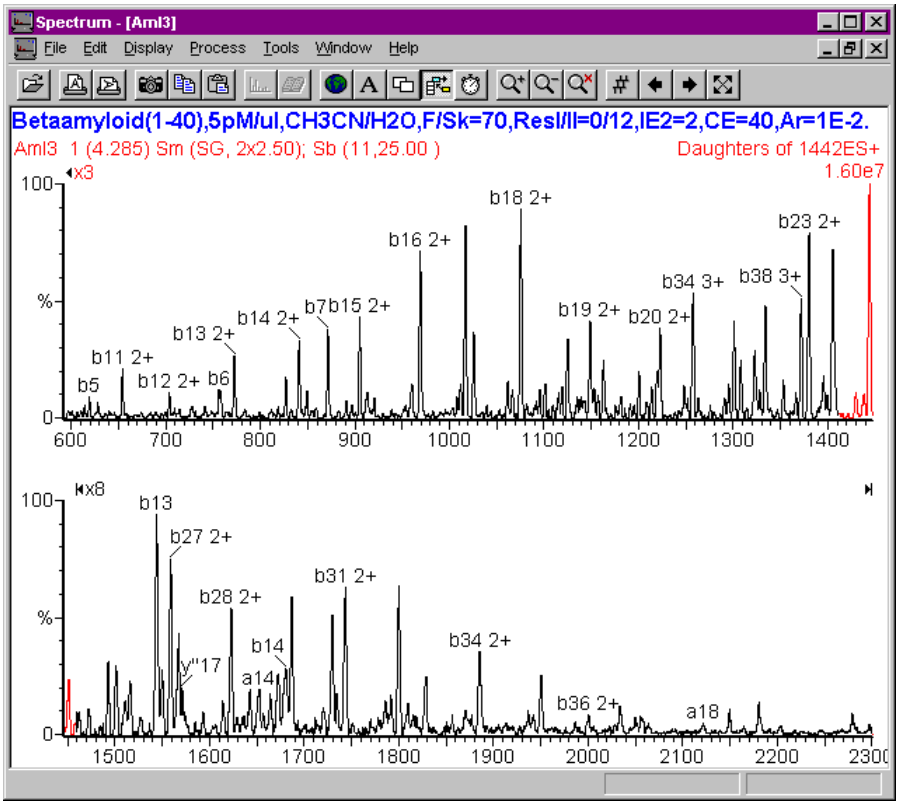


Figure 1.36 Annotated spectrum displaying matched mass spectral fragment ions

Customizing the ion series

Clicking on one of the buttons in the **Sequence Ions** window, simulates the addition of a hydrogen due to hydrogen transfer. E.g. if the **y** button is clicked the button changes to **y'** and the y ion series masses increase by 1.

Clicking the button a second time gives **y''** and adds 1 to the masses again, i.e. transfer of two hydrogens.

Clicking the button a third time resets back to the start, i.e. **y** and the original masses.

The side-chain ions **d, v and w** are included for high energy CID spectra which are important for identifying post-translational modifications and residues such as Ile and Leu.

Carrying Out a Fragmentation on Other Digest Fragments

If a digest has been carried out then the observed fragmentation pattern will be of the currently selected digest fragment. Any change to the digest windows i.e., as a result of selecting a different digest fragment (described in the **Digest** section) will result in the mass spectral fragments windows being updated with fragmentation of the currently selected digest fragment.

Overview of Fragmentation Processes

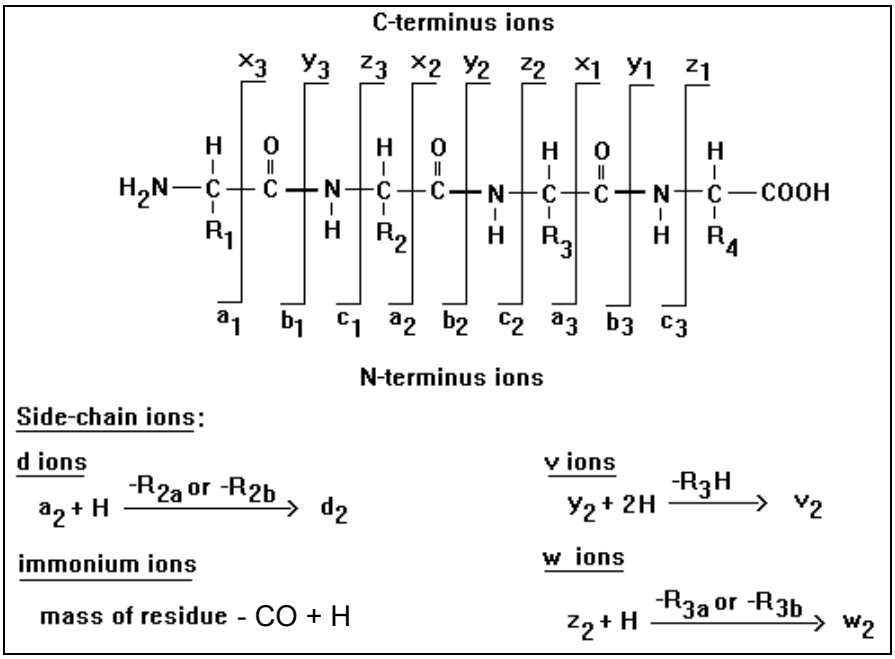


Figure 1.37 Backbone and Side-Chain fragment ions according to a combination of the Roepstorff and Biemann nomenclatures (Ref. 4 and 5)

Spectrum Interpretation

MS/MS data or in-source low energy CID electrospray MS spectra of peptides, including doubly charged product ions, can be interpreted without prior information. These ions can be generated by electrospray or fast atom bombardment ionization techniques. The **MS sequinator** program is suitable for interpreting the correct amino acid sequence of small peptides or tryptic-like peptides.

Based on graph theory considerations the program finds the most probable sequences, even if the composition is unknown, by scoring mass differences. If a part of the sequence is not well represented in the spectrum the program explores all possibilities which best explain the 'gap' in the sequence.

To Interpret Spectra

1. Choose **Primary**, **Interpret Spectrum** to activate the dialog box.

2. Press the **Browse** button to invoke the **Spectrum Data Browser** and select a centered data file from the **History**, or click with the right mouse button on a displayed centered spectrum.

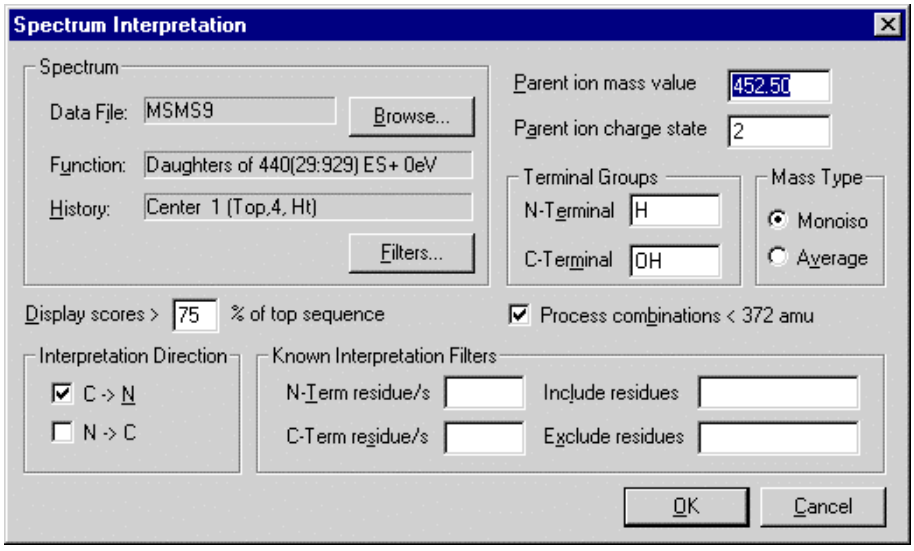


Figure 1.38 Spectrum Interpretation dialog box

3. Press **Filters** to set the % threshold or actual threshold and mass window parameters.

   The **% threshold** value should exclude background noise. The more peaks above the threshold the longer the interpretation process. Start with a relatively high threshold and progressively lower the threshold until the proposed sub-sequences masses match that of the peptides molecular mass.

   Enter a **mass window (tolerance)**. A value less than 1 amu is desirable (default is 1 amu). To get an idea of the accuracy of the data drag the left mouse button in the high mass end between two major peaks in the **Spectrum** window. The status bar displays the mass difference. Find two peaks differing by one amino acid residue and compare this to the actual monoisotopic mass for that residue. This value is important since the algorithm proposes sub-sequences based on mass differences between peaks. The inclusion or exclusion of a residue is therefore directly affected by the mass window (tolerance) setting.

4. Enter the **Parent ion mass value** i.e. the mass-to-charge ratio of the intact parent ion of interest. In the above example this is 452.5 for des-Arg Bradykinin.

5. Enter the **Parent ion charge state** i.e. singly or doubly charged. In the above example this is 2.

6. Enter the **N** and **C termini**. This is H and OH by default but can be changed by entering the formula for modified termini.

7. Choose **average** or **monoisotopic** mass. For peptides less than 1500 amu monoisotopic should be selected. Nominal masses are used in the sequencing algorithm to remove ambiguity and speed processing. A correction factor is therefore applied to all peaks.

8. Enter a value for the **Display scores > than % of top score**. If this value is 75% then only sequences scoring 75% of the top scoring sequence, or above, will be displayed

9. Check the **Process Combinations** box if you want the algorithm to consider combinations of residues less than 372 amu. This occurs when the proposed sub-sequence is not complete and the calculated mass of the peptide has not been reached.

10. Indicate the **Interpretation Direction** i.e., whether interpretation starts at the N-terminus or C-terminus or both. For tryptic-like peptides the **y ion series** is usually dominant so interpretation should start at the C-terminus end. If both check boxes are checked the program will carry out interpretation from both ends. This process will take longer but should give higher scores for well represented sequences.

11. Enter any **Known Interpretation Filters** as this speeds up processing. If the sequence starts with PPG then enter this in the **N-term residues** control or if the sequence ends in K or R enter K or R in the **C-term residue** control. From the low mass immonium ions the presence of certain residues might be evident, e.g. P at mass 70 or F at mass 120. For a complete list of Immonium ion masses see Table 1-1 at the end of this chapter. Enter these in any order in the **Include residues** control. Similarly residues can be excluded from the program by entering their single letter codes in the **Exclude residues** control. **NOTE:** For modified residues such as Carboxymethylcysteine (CMC) use the chain editors PTM button to configure a specified modified residue and assign it to B or J etc. This letter can then be inserted in the include residues box and **C** can be entered in the exclude residues box.

12. Press **OK**.

A **status message box** is displayed to indicate processing of the selected data file. Once the hourglass disappears the operation can be halted by pressing the **Cancel** button. The interpretation of the MSMS9 data file (des-Arg bradykinin) takes less than 30 sec on a Celebris GL 6200.

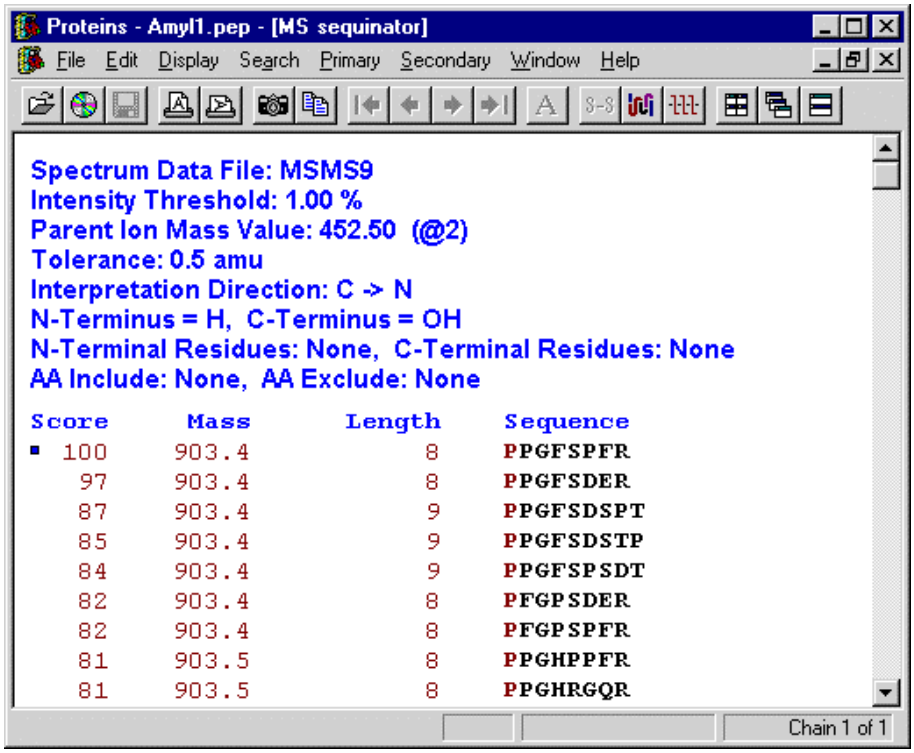


Figure 1.39 MS sequinator results of unknown peptide

The MS sequinator window displays all input parameters and a list of proposed sub-sequences. If any combinations of residues complete a sub-sequence they are displayed in **blue** text. Sub-sequences in **black** text are based on evidence from the original data. Similarly for printouts normal and bold text is used to indicate residue combinations completing sub-sequences.

The proposed sequences can similarly be matched against the original spectrum data by clicking on a sequence in the list. The matching is carried out if the **Match Spectrum** parameter is on (see the previous section **Mass Spectral Fragments**).

Hydropathic Profiles

A hydropathic profile is a plot of the mean hydropathy value versus residue position number. These plots assist in interpreting folding patterns of proteins and give an indication of *buried* residues as opposed to residues exposed to the aqueous solvent i.e., interior as opposed to exterior.

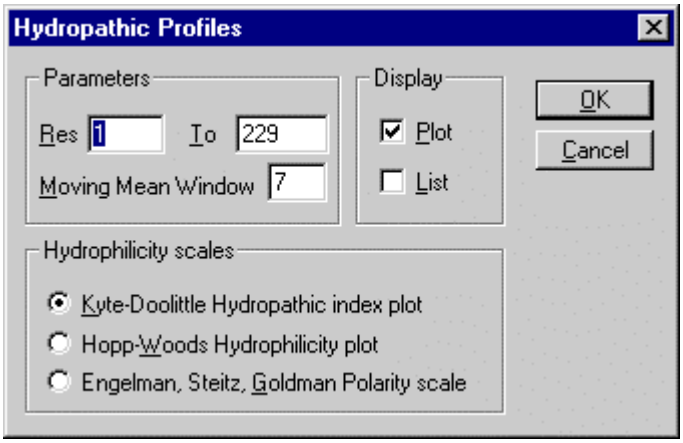


Figure 1.40 Hydropathic Profiles dialog

Carrying out a Hydropathic Profile

1. Select **Secondary, Hydropathic Profiles**.

2. Enter the **Res** and **To** numbers corresponding to the residue number in the sequence. This is by default the whole sequence. i.e., **From** is **1** and **to** is the last residue in the sequence.

3. Enter the **Moving Mean Window**. This is 7 by default since this value most accurately reflects the exterior as opposed to interior portions of a protein. Any odd numbers can be entered.

4. Click the **Plot** or **List** check boxes or **both** depending on whether you want a graphical plot and/or list of the hydropathic values.

5. Choose one of the three scales. **Kyte&Doolittle**; **Hopp&Woods** or **Engelman, Steitz&Goldman**.

6. Press **OK.**

The 3 scales are based on progressively evaluating the hydrophilicity and hydrophobicity of a protein along its amino acid sequence. The program uses a moving-segment approach that continuously determines the average hydropathy within a segment of pre-determined length as it advances through the sequence.

Kyte&Doolittle Hydropathic Index Plot

This scale highlights areas of interior as opposed to exterior regions of proteins. The more hydrophobic the area, i.e., the more negative, the more buried those corresponding residues.

Hopp&Woods Hydrophilicity Index Plot

This scale locates protein antigenic determinants by analysing amino acid sequences to find the point of greatest local hydrophilicity. The point of highest local average hydrophilicity is invariably located in, or immediately adjacent to, an antigenic determinant i.e., likelihood of an antibody binding to this specific area.

Engelman, Steitz & Goldman Polarity Scale

This scale as the name suggests highlights polar and non-polar regions. Acidic and basic amino acids such as (asp, glu) and (arg, lys) respectively tend to be polar while the neutral amino acids (gly, ala) are non-polar.

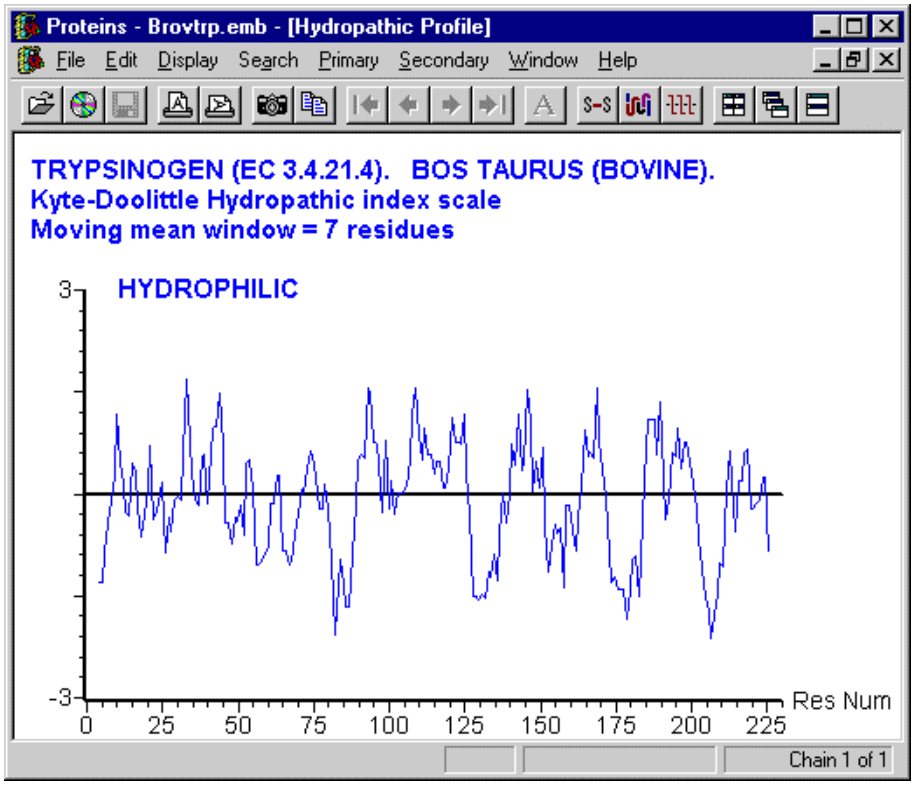


Figure 1.41 Hydrophobicity plot according to Kyte&Doolittle

If both **plot** and **list** are chosen two windows are displayed simultaneously. The relevant information is displayed at the top of the windows. The graphical plot cannot be scaled i.e., magnified or expanded. The scale can only be adjusted by changing the **Res** and **to** values to calculate values over a shorter sub-sequence. Negative values are hydrophobic and are relative to the specific plot.

Printer Support

Any displayed information can be translated into hard copy form. Hard copies can be obtained using any of the methods below.

Printing a Window or Contents of a Window

Use the keyboard Alt-Prn Scrn or Prn Scrn to copy a window to Windows Clipboard.

-or-

Print a Report by choosing File, Print Report or clicking one of the toolbar buttons. This invokes the Report Parameters dialog (Figure 1.42)

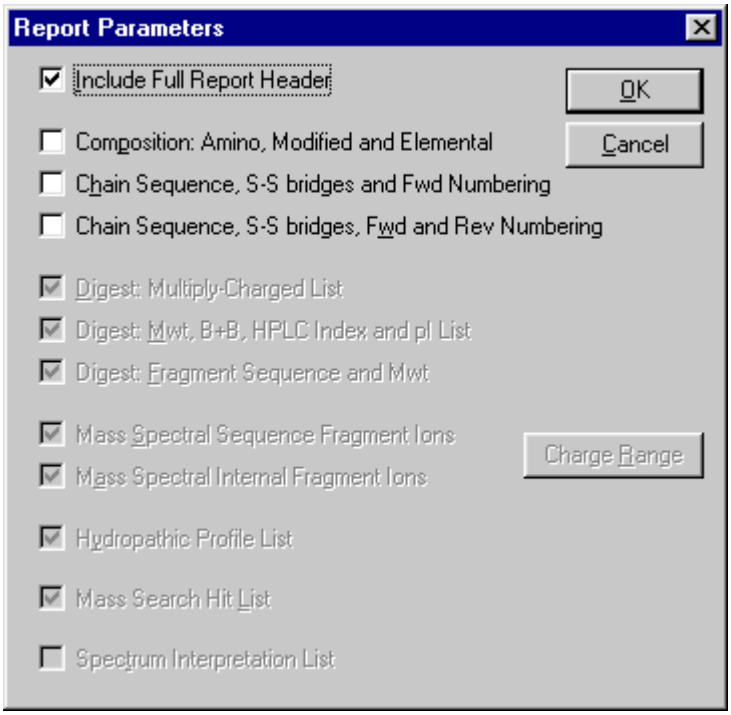


Figure 1.42 Print report dialog box used for obtaining hard copies of various analyses

Printing a Report

1. Click the required check box parameters. Grayed out options indicate that a particular analysis has not been carried out.

2. Press **OK**.

Include Full Report Header	Check this box to print a full report header attached to the printout. An alternative header, taking up fewer lines than the full report header, will be printed if this check box is unchecked.
Composition: Amino, Modified and Elemental	Check this box to print out the current chains contents i.e., number and composition of amino acids, elemental composition and any modified residues.
Chain Sequence, S-Bridges and Fwd/Rev Numbering	Check this box to print out the sequence of the current chain and a list of any disulfide bridges. The state of the disulfide bridges i.e., whether reduced or oxidized is indicated on the printout and reflected in the masses.
Digest: Multiply-Charged List and Mwt, B&B, HPLC & pI	Two alternative printouts of the same list of digest fragments.
Digest: Fragment Sequence and Mwt	Check this box to print out the sequence of residues and average and monoisotopic masses for each digest fragment
Mass Spectral Sequence Fragment Ions	Check this box to print out the sequence specific tabulated ions.

Mass Spectral Internal Fragment Ions	Check this box to print out the internal acyl fragment ions.
Hydropathic Profile List	Check this box to print a list of hydropathic values if a plot or list has been carried out.
Mass Search Hit List	Check this box to print out a list of mass search results
Spectrum Interpretation List	Check this box to print out a list of proposed sequences with their corresponding scores and masses.

Pressing the **Charge Range** button displays the **Charge State Parameters** dialog.

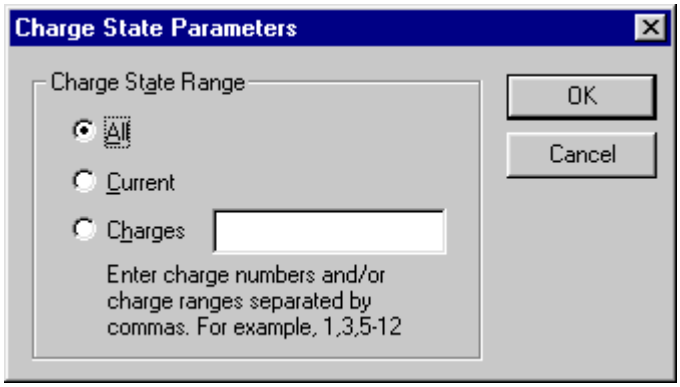


Figure 1.43 Charge State Parameters dialog

Select **All**, **Current** or enter specified charges in the **Charges** Box.

General Information

Amino Acid	Low Mass Ions	HPLC index	B&B	K&D hydro-value	H&W hydro-value	E,S&G polarity
Alanine	44	7.3	610	1.8	-0.5	-1.6
Arginine	129	-3.8	690	-4.5	3.0	12.3
Asparagine	87	-5.7	890	-3.5	0.2	4.8
Aspartic acid	88	-2.9	610	-3.5	3.0	9.2
Cysteine	76	-9.2	360	2.5	-1.0	-2.0
Glutamic acid	102	-7.1	510	-3.5	3.0	8.2
Glutamine	101	-0.3	970	-3.5	0.2	4.1
Glycine	-	-1.2	810	-0.4	0.0	-1.0
Histidine	110	-2.1	690	-3.2	-0.5	3.0
Isoleucine	86	6.6	-1450	4.5	-1.8	-3.1
Leucine	86	20.0	-1650	3.8	-1.8	-2.8
Lysine	101	-3.7	460	-3.9	3.0	8.8
Methionine	104	5.6	-660	1.9	-1.3	-3.4
Phenyl-alanine	120	19.2	-1520	2.8	-2.5	-3.7
Proline	70	5.1	-170	-1.6	0.0	0.2
Serine	60	-4.1	420	-0.8	0.3	-0.6
Threonine	74	0.8	290	-0.7	-0.4	-1.2
Tryptophan	159	16.3	-1200	-0.9	-3.4	-1.9
Tyrosine	136	5.9	-1430	-1.3	-2.3	0.7
Valine	72	3.5	-750	4.2	-1.5	-2.6
N-term (H)	-	4.2	-	-	-	-
C-term (OH)	-	2.4	-	-	-	-

Table 1-1 Values used in BioLynx for the most common amino acid residues.

Amino Acid Residue Mass Differences

Nominal Mass Difference	Amino Acid Change	Nominal Mass Difference	Amino Acid Change	Nominal Mass Difference	Amino Acid Change
0	Ile → Leu	18	Leu → Met	42	Gly → Val
0	Gln → Lys	18	Ile → Met	43	Leu → Arg
1	Lys → Glu	19	His → Arg	43	Ile → Arg
1	Ile → Asn	22	Asp → His	44	Ala → Asp
1	Asn → Asp	23	Asn → His	44	Cys → Phe
1	Gln → Glu	24	Leu → His	46	Gly → Cis
3	Lys → Met	25	Met → Arg	48	Val → Phe
4	Pro → Thr	26	His → Tyr	48	Asp →Tyr
9	Gln → His	26	Ala → Pro	49	Asn → Tyr
10	Ser → Pro	26	Ser → Leu	53	Cys → Arg
12	Thr → Ile	27	Ser → Asn	55	Thr → Arg
12	Thr → Leu	27	Thr → Lys	58	Gly → Asp
13	Thr → Asn	28	Lys → Arg	58	Ala → Glu
14	Gly → Ala	28	Ala → Val	59	Pro → Arg
14	Ser → Thr	28	Gln → Arg	60	Ser → Phe
14	Val → Leu	30	Val → Glu	60	Cys → Tyr
14	Asp → Glu	30	Arg → Trp	69	Ser → Arg
14	Asn → Lys	30	Thr → Met	72	Gly → Glu
15	Leu → Gln	30	Gly → Ser	73	Leu → Trp
15	Ile → Lys	30	Ala → Thr	76	Ser → Tyr
16	Val → Asp	31	Pro → Gln	83	Cys → Trp
16	Ser → Cys	32	Val → Met	99	Ser → Trp
16	Ala → Ser	34	Leu → Phe	99	Gly → Arg
16	Phe → Tyr	34	Ile → Phe	129	Gly → Trp
16	Pro → Leu	40	Pro → His		

Table 1-2 Mass difference between amino acid residues that are alterable through single-nucleotide mutation in coding triplet.

References

1. Stryer, L. in *Biochemistry* (1988) (3ed), Freeman, W.H and Company, New York.

2. Bull, H.B and Breese, K. (1974) *Arch. Biochem. Biophys* **161**, 665.

3. Browne, C.A., Bennet, H.P.J and Solomon, S. (1982) *Anal. Biochem.* **124**, 201.

4. Biemann, K. (1989) in *Protein Sequencing: a practical approach* (Chap.5). Findlay, J.B.C and Geisow,M.J (eds), IRL Press (Oxford).

5. Roepstorff, P and Fohlman, J. (1984) *Biomed. Mass Spectrom.* **11**, 601.

6. Kyte, J and Doolittle, R.F. (1982) *J. Mol.Biol.* **157**, 105.

7. Hopp, T.P and Woods, K.R. (1981) *Proc. Nat. Acad. Sci.* USA. **78**, 3824.

8. Engelman, D. , Steitz, T. and Goldman, A. (1986) *Ann. Rev. Biophys Chem.* **15**, 321.

9. Shackleton, C. H. L., Falick, A. M., Green, B. N. and Witowska, H. E. (1991) *J. Chromatogr.* **562**, 175.

10. Gilland and von Hippel. (1989) *Anal Biochm* **182**:319-326.

12. Sillero and Ribeiro, (1989) *Analytical Biochem.* **179**, 319-325.

Chapter 2 BioLynx - Peptide Sequencing

Contents

Overview 2-5
Peptide Fragmentation 2-5
The Scoring Algorithm 2-5
Creating a New Document 2-6
Specifying the Spectrum 2-6
Specifying Sequencing Parameters 2-7
Peptide Parameters Page 2-7
Sequence Parameters Page 2-9
Modifications Page 2-9
Digest Reagents Page 2-11
Analyzing the Spectrum 2-11
Troubleshooting 2-12
Isotopic Labelling 2-13
Peptide Sequencing Documents 2-14
The Main Window 2-14
Child Windows 2-15
The Sequence Pane 2-15
The Spectrum Pane 2-16
The Tree Views Pane 2-18
The Sequence Report Pane 2-19
The Toolbar 2-19
Menu Commands 2-21
File Menu 2-21
Report Options 2-21
View Menu 2-23
Options 2-24
Tools 2-24
Window Menu 2-25
References 2-26

Figures

Figure 2.1 Peak List Format 2-6
Figure 2.3 The Peptide Parameters Page 2-7
Figure 2.4 The Sequence Parameters Page 2-9
Figure 2.5 The Modifications page 2-9
Figure 2.6 Creating a new modification 2-10
Figure 2.7 The Digest Reagents Page 2-11
Figure 2.8 The MassSeq dialog 2-12
Figure 2.9 The calibration window 2-12
Figure 2.10 The Isotopic Labelling dialog 2-13
Figure 2.11 Peptide Sequencing Program Main Window 2-14
Figure 2.12 The Sequence Pane 2-15
Figure 2.13The Spectrum Pane 2-16
Figure 2.14 The Find Tag dialog 2-16
Figure 2.15 The Tree View Pane 2-18
Figure 2.16 Tree View Menu 2-18
Figure 2.17 The Sequence Report Pane, showing a Sequence Prediction Details Report 2-19
Figure 2.18 The Toolbar 2-19
Figure 2.19 The Report Options dialog, Reports page 2-22
Figure 2.20 Report Options dialog, Spectrum Plot page 2-22
Figure 2.21 The Spectrum View Options dialog 2-23
Figure 2.22 The View Options Font Sizes dialog 2-23
Figure 2.23 Blast Parameters dialog 2-24
Figure 2.24 The Amino Acid Combinations Lookup Configuration dialog 2-24

Overview

The Peptide Sequencing program assists in deducing the residue sequence of peptides that are the result of a digest, particularly tryptic peptides. There are two methods to sequencing – an automated "**MassSeq**" approach as well as a "**FindTag**" approach.

MassSeq This approach requires minimum parameters and involves loading the MS/MS data, specifying the peptide's molecular weight, any modifications and an indication of mass accuracy.

Note: This option is only available if purchased.

FindTag This approach guided by user-supplied parameters builds a set of sub-sequences whose series ions best match the experimental spectrum. Using step or leap-wise logic you can extend the sub-sequences until the molecular weight of the peptide matches that of the input parameters. Spectral data may be input directly from a MassLynx data file or as a list of masses and intensities.

The results of the analysis are displayed in graphical form as an annotated spectrum or tree view and in textual form as reports detailing the fragment series ions and spectral assignments based on the sequences.

Peptide Fragmentation

The rules for peptide fragmentation and the ions produced are explained in the Mass Spectral Fragments section in the Proteins chapter. The ions most frequently observed for low energy collision are the 'a', 'b', 'y' and 'z' ions. These ions are currently used in the PepSeq application for scoring and spectral annotation.

The Scoring Algorithm

FindTag The software finds sub-sequences that are consistent with the mass differences between peaks that represent amino acids. Because more than one possible sequence might be consistent with the spectral data and molecular mass, sequences are ranked in descending order according to the number of consecutive 'a', 'b', 'y' and 'z' ions matched against the spectrum using only singly-charged information and a weighted intensity of the matching peaks. For a more comprehensive guide to peptide mass spectral fragment ion scoring see reference 1.

The molecular weight of the sub-sequence is not used as a discriminating factor and an incorrect molecular weight does not preclude the sequence from being retained in the final set, because in many cases the full sequence cannot be determined from the series ions. Especially for longer peptides, residues toward the ends of the sequence may not be observed due to loss of fragmentation, and thus the sequence cannot be fully extended.

It is possible to exercise considerable control over the sequencing algorithm and scoring by appropriate parameter settings. For example, setting an intensity threshold will force the algorithm to ignore any spectral features of lower intensity. In this way, coincidental matches to minor ions can be minimized. Other adjustable parameters include the match tolerance between expected ions and spectral features.

The success of the algorithm at finding sequences is dependent on the quality of the input spectrum. Weak, noisy spectra are likely to yield poor sequences, because of the high probability of coincidental matches. The best results are achieved with stronger, higher signal to noise spectra and spectra that have previously been processed using the MaxEnt3 software.

Note for MassSeq users: Only MaxEnt3 processed data can be analysed using MassSeq. MassSeq only uses the singly charged molecular mass and the parameters defined in the MassSeq dialog, but you may define modifications for MassSeq to consider.

To Access Peptide Sequencing, select BioLynx, from the MassLynx Shortcut Bar and Peptide Sequencing from the MassLynx Bar.

Creating a New Document

Select **New** from the **File** menu or press the button on the Standard Toolbar. An empty Child Window will be created, named "Untitled1". There is no spectrum associated with this document, and all parameter settings are defaulted.

Specifying the Spectrum

Spectral data must be processed using MassLynx before the Peptide Sequencing program can analyze it. Only Massive Inference (MaxEnt3) processed data or multiply-charged, positive ion centered spectra can be analysed.

There are two ways to insert a centered spectrum into a document, copying a peak list from the clipboard or importing a spectrum from a MassLynx file.

Copying a Peak List from the Clipboard

A peak list is a set of m/Z, intensity pairs, listed one pair per line, as shown in Figure 2.1.

1. Open the listing file in Microsoft Notepad ™ or a similar program, select all the lines containing m/Z, intensity data and copy them to the clipboard using the **Edit, Copy** menu command or its equivalent.

2. Switch to the Peptide Sequencing program, and select **Paste Spectrum** from the **Edit** menu or press the button. The peak list will be copied in, replacing any existing spectrum in the document.

102.7 4.698e3

109.9 1.684e4

110.5 1.093e4

111.0 1.458e4

113.0 2.320e4

114.0 4.056e3

117.6 8.153e3

119.5 3.889e3

Figure 2.1 Peak List Format

Loading a Spectrum from a MassLynx File

Select the the **File**, **Load Spectrum** or press the toolbar button. This invokes the **MassLynx Spectrum Browser** dialog, from which a centered data file or MaxEnt3 spectrum can be chosen. See the following section for more information.

> The Peptide Sequencing program makes a copy of the centered spectrum and stores it in the Peptide Sequencing document. Once this copy has been made, subsequent changes to the MassLynx file will not be reflected in the copy. If you modify the centered spectrum using MassLynx, you must load the modified spectrum into the Peptide Sequencing document again.

Specifying Sequencing Parameters

Before a spectrum can be sequenced the Sequencer must be given a set of parameters, which describe known peptide information. The **Peptide Sequencing Parameters** (Figure 2.2) dialog is used for this purpose.

The **Parameters** dialog will open automatically when a spectrum is loaded, alternatively to open the dialog select **Process, Specify Parameters** or press the button.

The dialog has four pages:

- The Peptide page for providing molecular weight, charge and tolerance information.
- The Sequence page specifying composition related information
- The Modifications page for specifying applicable modifiers.
- The enzymatic/chemical digest page for providing information regarding sample treatment prior to mass spectral analysis.

Press the appropriate tab to view a different page. Parameters are not applied until you press **OK**.

Peptide Parameters Page

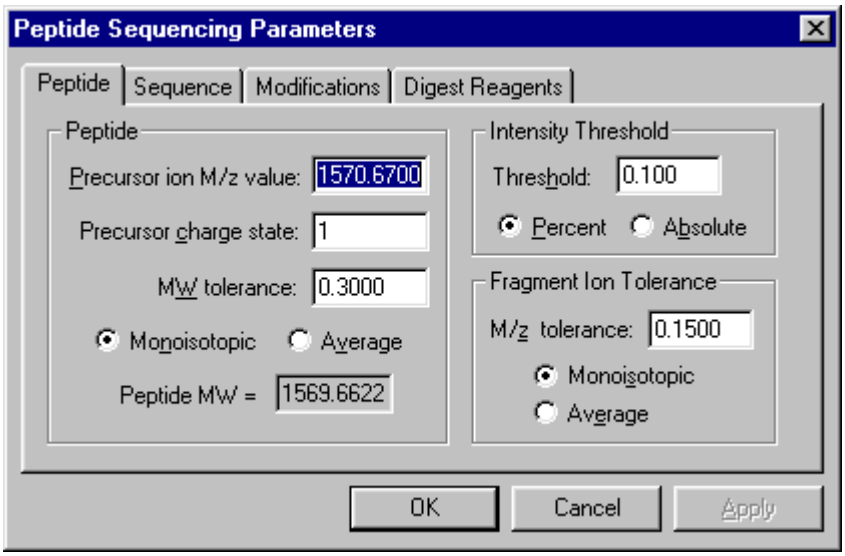


Figure 2.2 The Peptide Parameters Page

It is important to provide an accurate measurement of peptide molecular weight and charge state. This is usually determined from the parent spectrum. A narrow molecular weight tolerance helps eliminate sequences that are not correct.

Precursor ion M/z value Enter the precursor ion M/z value of interest. If the data has been processed by MaxEnt 3 then enter the M+H value, alternatively enter the multiply charged value.

Precursor charge state Charge state of precursor ion. If data has been processed by MaxEnt 3 then enter 1 for charge state.

The following parameters are not used in MassSeq processing:

MW tolerance A default tolerance of 0.3 is set on loading a spectrum. Values should not be higher than 1 Da.

Select either **Monoisotopic** or **Average** mass type depending on quality of data and molecular weight of peptide. In most cases monoisotopic will be adequate for peptides less than 3000 Da.

Threshold Enter a threshold value in either **Percent** or **Absolute**. Below this value peaks are not considered in the building of sub-sequences or in assigning peaks or annotation.

Fragment Ion Tolerance A default of 0.15 Da is set on loading a new spectrum. This value reflects the accuracy between two peaks belonging to the same ion series with the peak difference equivalent to the mass of an amino acid residue. Select Monoisotopic by default.

The threshold and fragment ion tolerance, are the most critical parameters for successful sequencing.

If the m/Z tolerance is too narrow the sequencer will not be able to match calculated series ions to the actual spectrum.

If it is too wide then incorrect matches will be made.

If the intensity threshold is too low the sequencer could inappropriately match calculated series ions to noise peaks.

If it is too high then legitimate but weak peaks will be ignored when matching.

A better intensity threshold can be determined by placing the spectrum pane into **Display Peaks Above Threshold** mode (press the toolbar button), and examining the effect of changing the threshold. In this display mode, the only peaks displayed in the spectrum plot are those which will be considered during series ion matching. Peaks below the threshold will not be shown. A better m/Z tolerance can be chosen by examining the spectrum or through trial and error.

Sequence Parameters Page

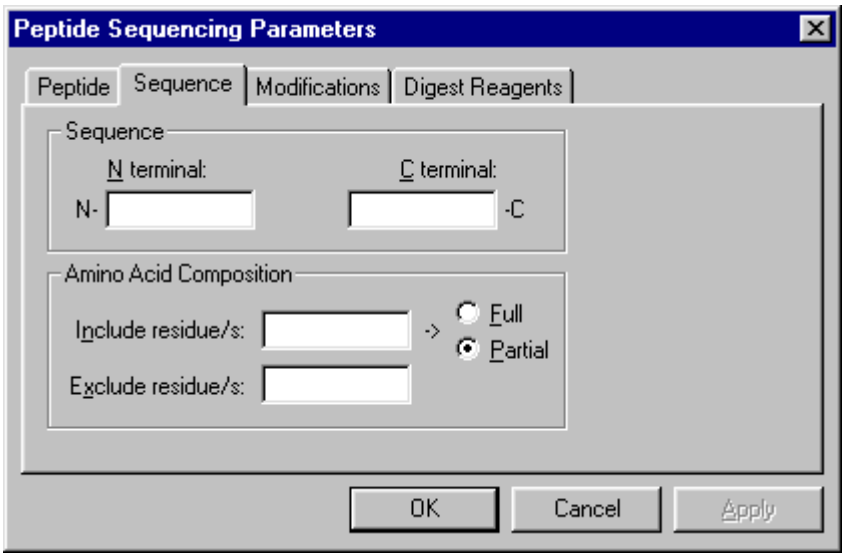


Figure 2.3 The Sequence Parameters Page

Note. The following parameters are not used in MassSeq processing

Enter any known residues in the **N terminal** and **C terminal** boxes, appropriate to their position in the sequence.

Enter any known residues, in any order, in the **Include residue/s** box and check the **Full** box if the complete composition has been entered, or the **Partial** box if a partial composition has been entered.

If you wish to exclude a peptide enter any known residues, in any order, in the **Exclude residue/s** box. You may wish to exclude peptides that are known to be contaminants or autolysis products.

Modifications Page

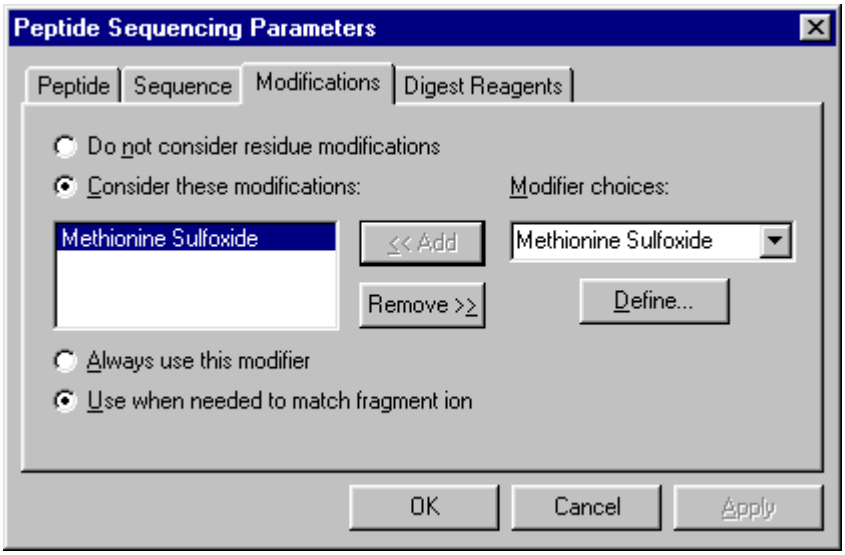


Figure 2.4 The Modifications page

Note. The following parameters are not used in MassSeq processing:

Do not consider residue modifications	Check this box if there are no modifications, if there are check the Consider these modifications box. Checking the latter will enable other fields on the screen

The box underneath the **Consider these modifications** box contains a list of modifications, which may have occurred and should be considered when searching.

To add new modifiers to this list click on the arrow at the end of the **Modifier choices** box, this will display a list of modifiers. Click on one of the modifiers and press the **<<Add** button.

To delete modifiers from this list click on one of the modifiers and press the **Remove>>** button.

Check the **Always use this modifier** box to use this modified residue in all calculations, or the **Use when needed to match fragment ion** box to only use if it results in finding a matching ion.

New modifiers can be defined, or existing ones edited, by selecting the **Define** button and entering the relevant information as in Figure 2.5

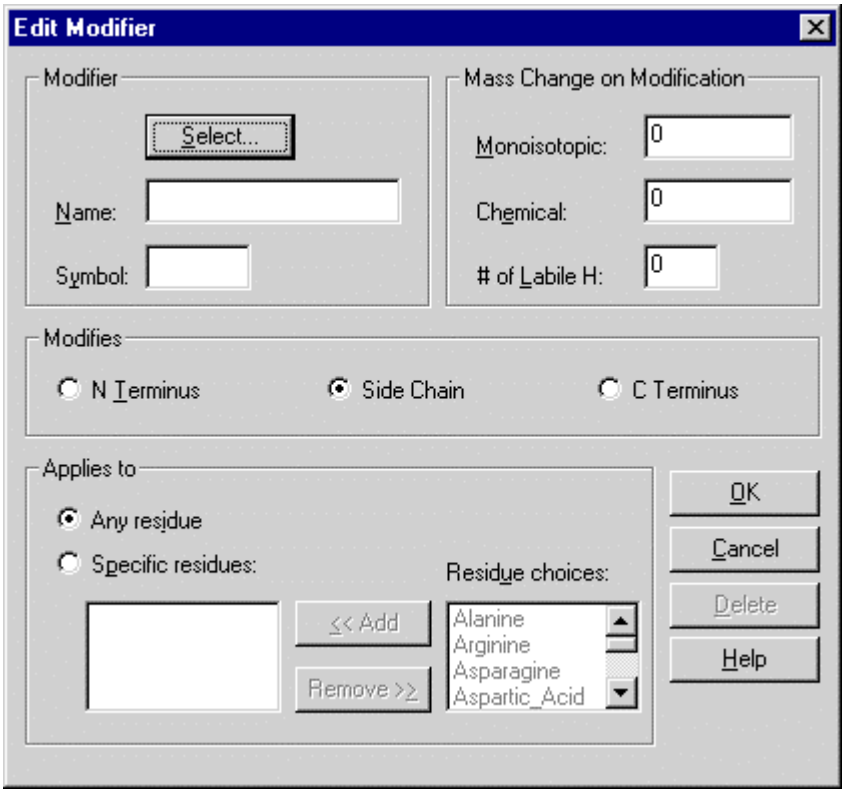


Figure 2.5 Creating a new modification

Enter details of a new modifier, or to edit an existing modifier press the **Select** button to invoke the **Select Modifier** dialog. Click on one of the modifiers in the list and press **OK**. Details of the selected modifier will be displayed in the **Edit Modifier** dialog.

Name	Enter a name to be displayed in the Modifier choices box.
Symbol	Enter a symbol to be displayed in the Modifier choices box.
Monoisotopic and/or **Chemical**	Enter a mass change due to modification.

# of Labile H	Enter the number of labile Hydrogens for deuteration in the box.
N Terminus, **Side Chain** or **C Terminus**	Select one of these options to show which group is modified.
Any residue	Select this option to apply the defined modifiers to any residue or select **Specific residues** to apply to the residues specified in the list box below this field. To add a residue to the Specific residues list box click on one of the residues in the **Residue choices** list and press the **<<Add** button. To delete a residue from the Specific residues list box click on one of the residues in the Specific residues list and press the **Remove>>**.
Delete	Pressing this button will clear the screen and result in the modifier being removed from the Modifier choices box.

Digest Reagents Page

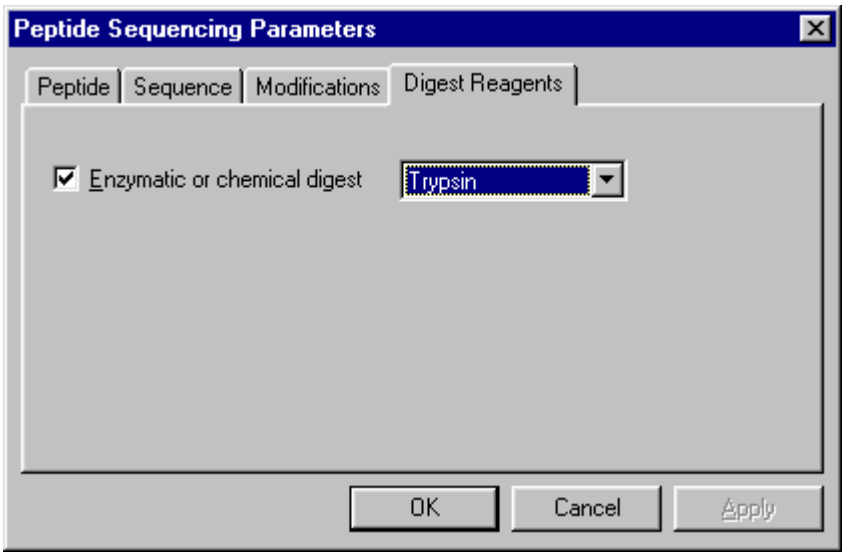


Figure 2.6 The Digest Reagents Page

If the peptide mass entered is the result of an enzymatic or chemical cleavage check the **Enzymatic or chemical digest** box and select a digest from the drop down list box.

Analyzing the Spectrum

Once a spectrum has been loaded and the parameters have been configured, analysis may proceed. PepSeq has been designed as an interactive program that allows the user to make decisions at each step in deducing the sequence of a peptide. The MassSeq option performs automatic peptide sequencing.

Three options are currently available:

- Entering a sequence manually to observe the expected theoretical fragment ions and spectrum assignment and annotation.

  To perform this operation click with the right mouse button on the Sequence pane and select **New Sequence** from the pop-up menu displayed. An edit box will appear that allows

the user to type in single letter codes for the amino acids. When a sequence has been entered press the **Enter** key.

To edit a sequence first select it by clicking on it, then click on it again. Change the required letter codes and press the **Enter** key. To cancel an edit operation press the **Esc** key. The maximum number of amino acid residues that can be entered is 50.

- Find tag – from the spectrum pane select a peak of interest by clicking on it with the right mouse button and select **Find tag** from the pop-up menu displayed. Enter the required parameters on the dialog displayed and press **OK**. The software will sequence the spectrum and suggest a possible list of sub-sequences. These sub-sequences can then be further acted on to extend or complete them in accordance with the user-input molecular weight.

- MassSeq – by pressing the toolbar button, de novo sequencing may be initiated. The MassSeq dialog is invoked, which prompts the user to enter the **minimum mass standard deviation**. This value is most easily obtained from the instrument calibration as shown in Figure 2.7.

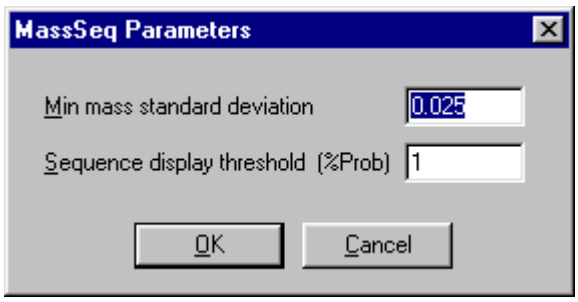


Figure 2.7 The MassSeq dialog

Min mass std dev Enter the value of the standard deviation of the residuals in the calibration window, Figure 2.8 (in this case, 0.009341) (which can be found on the Acquisition Control Panel of the main MassLynx interface).

Sequence display threshold Any peak with a % probability below the value entered in the box will not be displayed.

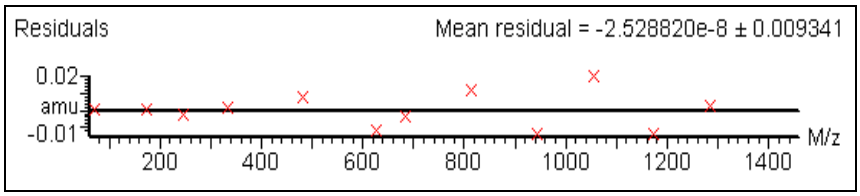


Figure 2.8 The calibration window

Troubleshooting

On occasion, the sequencer will be unable to find any sequences consistent with the spectrum and sequencing parameters, or may find sequences, which are known from other information to be incorrect. This could result from a number of problems:

- A poor quality spectrum was used. If the input spectrum has low signal-to-noise or incomplete fragmentation, it may not be possible for the sequencer to match sufficient expected series ions to spectral features to construct a sequence.

- The molecular weight is inaccurate. The more accurate the MW the more accurate will be the complementary ion calculations and hence the better scoring for the correct sequence

- The M/z tolerance is too low. If the mass accuracy or resolution of the spectrum is low the tolerance must generally be made wider, otherwise the computed series ion values will be outside the tolerance window and matching will fail. If the sequencer is unable to extend a candidate sequence because series ions for the next residue cannot be matched, the partial sequence is abandoned.

- The M/z tolerance is too wide. This will result in coincidental matches of series ions leading to incorrect sub-sequences. This problem will be most evident in the Sequence details textual report.

- The intensity threshold is too high. The sequencer ignores spectral peaks below the threshold when matching expected series ions. If the threshold is too high, legitimate peaks will be ignored and incorrect or no sequences will be found.

- The intensity threshold is too low. As in the case of too wide an M/z tolerance, too low an intensity threshold will result in coincidental matches and extension of the series sequences with inappropriate residues.

In many cases, determining the sequence of an unknown Peptide is a trial and error process. It is sometimes necessary to analyze several spectra of differing parent charge states and compare results, and to perform several analyses of each spectrum with differing parameter sets.

Isotopic Labelling

Select **Isotopic Labelling** from the Process menu or press the toolbar button to display the Isotopic Labelling Parameters dialog. This dialog can only be invoked when the loaded spectrum has been processed using Massive Inference (MaxEnt3).

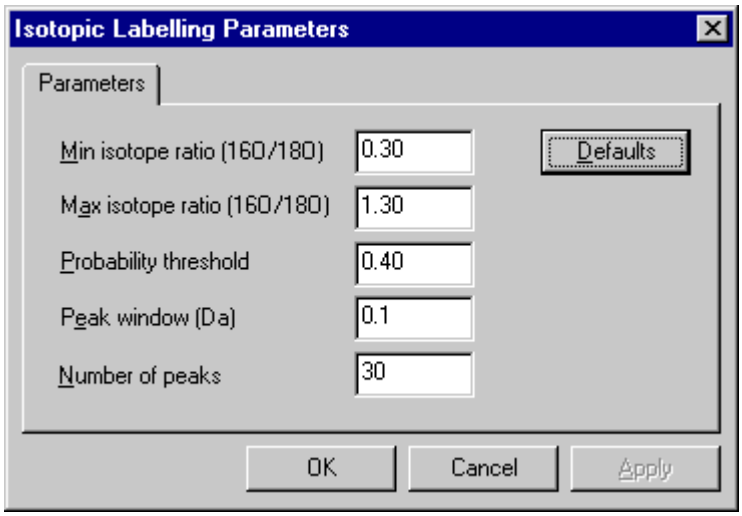


Figure 2.9 The Isotopic Labelling dialog

Data obtained from isotopically labelled digests allow unambiguous assignment of either ‘b’ or ‘y’ ions depending on the type of labelling. O^{18} labelling adds 2 Da to ‘y’ ions therefore these ions can be identified by their characteristic patterns. The MaxEnt3 result shows an O^{16} peak and an O^{18} peak for each ion separated by approximately 2Da.

The algorithm proceeds as follows:

First, all peaks with probability of detection less than the **Probability threshold** are removed. **Note:** The probability of detection is output from the MaxEnt3 program and is invisible to the user.

Next, all pairs of peaks separated by 2Da +/- **Peak window** are considered. The ratio O^{16} intensity / O^{18} intensity is calculated, and if this ratio falls within the limits **Min isotope ratio** < calculated ratio < **Max isotope ratio**, the pair is considered for the third stage.

Finally, all pairs of peaks from the stage above are ranked by a score based on the product of their intensities, and the observed divergence from the expected 2Da mass difference. The top-scoring **Number of peaks** pairs are output as y ions.

Pressing the **Defaults** button will reset all fields to the default values.

Peptide Sequencing Documents

Information in the Peptide Sequencing program is organized into *documents*. A document, when complete, contains all of the information produced during the analysis of a spectrum to predict Peptide sequences. This information includes a copy of the centroided spectrum and its descriptive comments, all parameters used in the analysis, and all analysis results.

The Main Window

The Peptide Sequencing program consists of a **Main Window** containing a **Menu** and an optional **Toolbar** and **Status Bar**. Within the Main Window, one or more **Child Window(s)** may be positioned, each containing a multiple-part view of a set of sequencing results. The Main Window is shown in Figure 2.10 below.

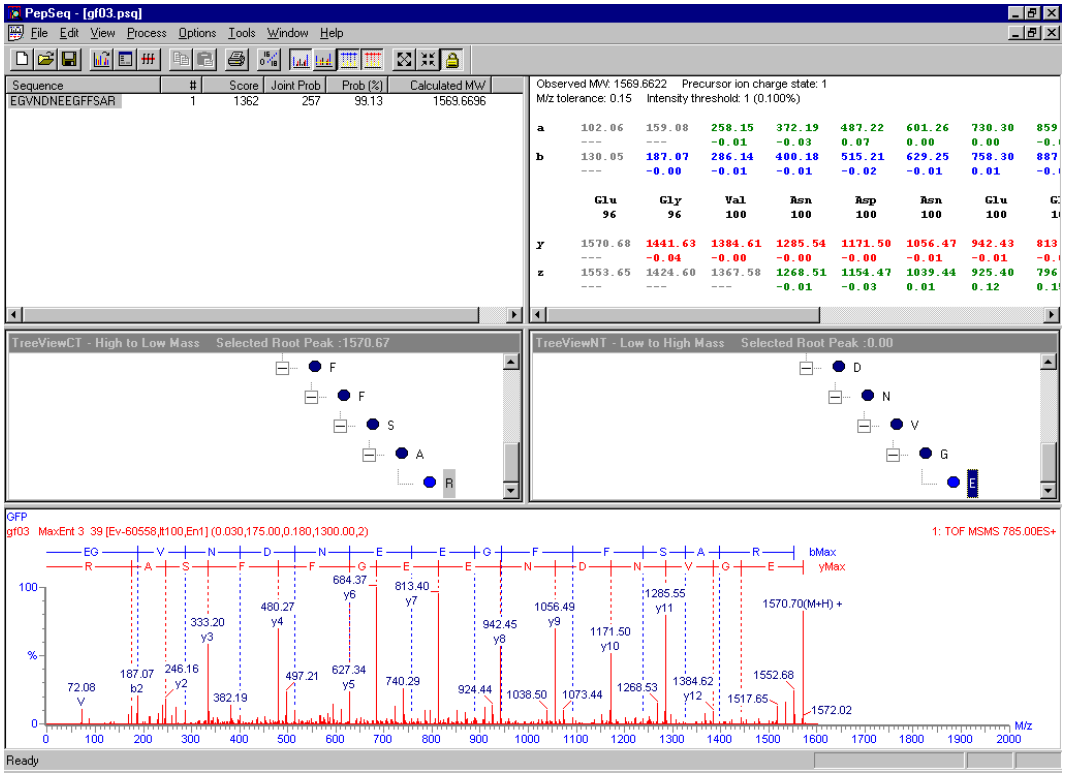


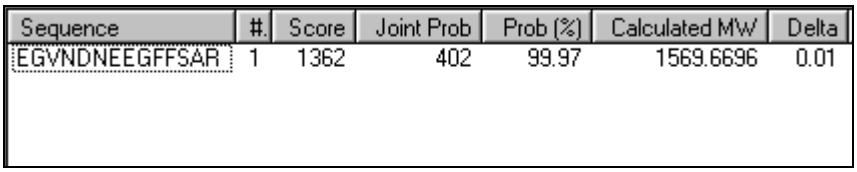
Figure 2.10 Peptide Sequencing Program Main Window

Child Windows

As described above, the Main Window can contain one or more Child Windows, each representing a different part of the Peptide Sequencing document. Using commands from the Menu, a Child Window may be cloned into two or more Child Windows, each giving an independent view of the same document. This allows results of two different candidate sequences to be studied side-by-side. Alternatively, two documents may be opened simultaneously (in two separate child windows), permitting side-by-side comparison of sequences derived from analyses with different parameter settings or daughter spectra from two different precursor ions of the same parent spectrum.

Each Child Window is divided into five **Panes**, containing the sequences, detailed reports, tree views and spectrum respectively.

The Sequence Pane



Sequence	#.	Score	Joint Prob	Prob (%)	Calculated MW	Delta
EGVNDNEEGFFSAR	1	1362	402	99.97	1569.6696	0.01

Figure 2.11 The Sequence Pane

The sequence pane shows the predicted sequences from an analysis of the spectrum. These sequences can be sorted by clicking on the required column heading. **Joint Prob** and **Prob(%)** are only used if **MassSeq** is installed.

Each sequence has a score, calculated MW and delta mass. The # column indicates the position of the sequence relative to the best scoring sequence.

To Select a Sequence

A sequence is selected by clicking on it. The currently selected sequence is highlighted. When a new sequence is selected, the contents of the other two panes will change to reflect analysis details for that sequence.

To Modify a Sequence

To edit a sequence double click on it. Change the required letter codes and press the **Enter** key. To cancel an edit operation, press the **Esc** key. The maximum number of amino acid residues that can currently be entered is 50. Only sequences that have been manually entered can be edited. Sequences predicted via the **FindTag** procedure cannot be edited manually.

To Insert a Sequence

To insert a new sequence right click on the Sequence pane and select **New Sequence** from the pop-up menu displayed. An edit box will appear that allows the user to type in single letter codes for the amino acids. Once a sequence has been entered press the **Enter** key.

To Delete a Sequence

To delete a sequence right click on the Sequence pane and select **Clear Sequence** from the pop-up menu displayed. To delete all sequences select the **Clear All Sequences** option. Each sequence has a score, calculated MW and delta mass. The # column indicates the position of the sequence relative to the best scoring sequence. Clicking on the Score or Delta column heading will sort the sequences in the corresponding order.

The Spectrum Pane

The spectrum pane (Figure 2.12) contains a plot of the spectrum, annotated with details from the currently selected sequence. The annotations which appear on the spectrum can be specified using the toolbar or menu commands described below. Currently only ‘b’ and ‘y’ ion assignments are indicated above the spectrum. The **bMax** and **yMax** indicators serve to illustrate the maximum mass for these ions based on the user entered molecular mass.

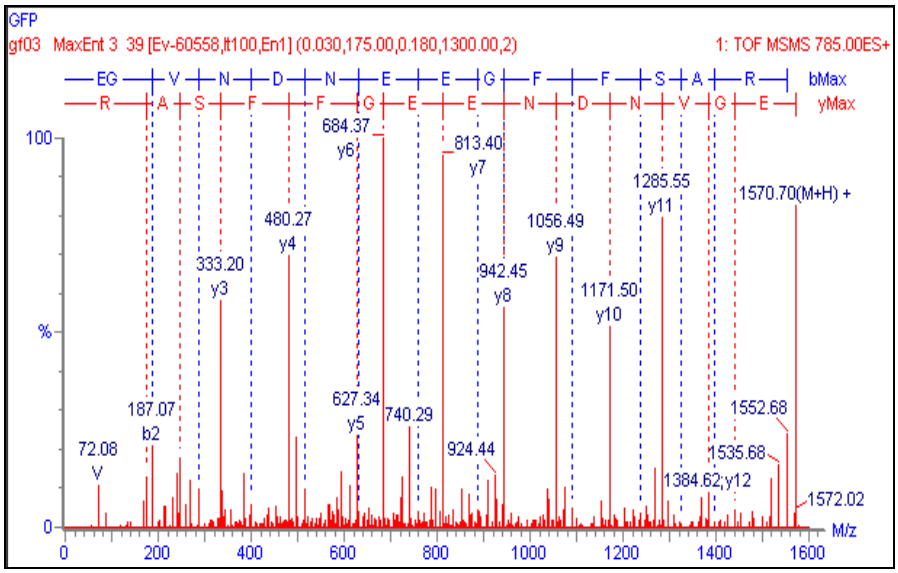


Figure 2.12The Spectrum Pane

Find Tag

The Find Tag dialog will find all the possible sub-sequences which match the input parameters and the spectrum, starting at a particular peak. To carry out this operation right on a peak of interest. If the data is from a tryptic digest select an intense peak at the high end of the mass scale. It is also recommended that a MaxEnt 3 processed spectrum, which would have deconvoluted the multiply charged data into its singly charged equivalent, is used. Once a peak is selected click on it with the right mouse button and select **Find tag** from the pop up menu displayed. The Find Tag Options dialog is invoked, enter the required parameters and press **OK** to start the sequencing operation.

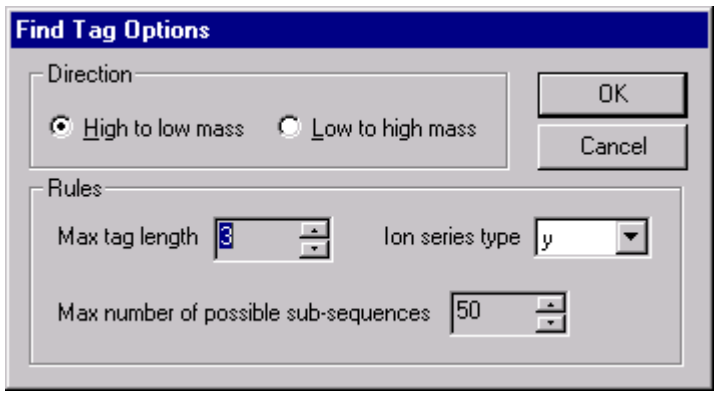


Figure 2.13 The Find Tag dialog

High to Low mass Starting at the selected peak the sub-sequence tracking will proceed in the indicated direction. If a high mass peak has been selected then choose the High to Low mass direction.

Low to High mass If a low mass peak has been selected choose this option.

Max Tag Length Enter the maximum number of amino acids to appear in the sub-sequence.

Note: FindTag currently operates on a step by step basis so missing peaks which would account for a missing residue will not be found using this approach. See below for information on jumping gaps in the peptide sequence. It is also important to find small tags i.e. 3 to 4 residues long rather than 6 to 7 residues long as the computational time increases exponentially.

Ion Series Type The ion series must be indicated. Currently only 'a', 'b', 'y' and 'z' ions are considered. For tryptic-like peptides use 'y' ions starting at high mass and move down to low mass.

Max number of possible sub-sequences Enter the maximum number of sub-sequences to display in the Sequence pane. Use 50 as default. Lowering this number may result in the correct sub-sequence not appearing on the final sequence list.

When the **OK** button is pressed the sequence processing is started and a progress dialog is displayed. Processing can be stopped by pressing the **Stop** button. All sub-sequences found will be displayed in the sequence view window in order of score.

Repeat Find Tag

This option is only active if the Find Tag operation has already been performed. It extends the tag by the number of residues specified in the Find Tag dialog. E.g. if 3 was specified then Repeat Find Tag find the next 3 sub-sequences for each tag found in the Find Tag operation.

To carry out this operation click with the right mouse button on either the lowest or highest assigned peak for the sequence of interest and select **Repeat find tag** from the pop up menu displayed. Repeat Find Tag repeats the original Find Tag operation using the previous Find Tag parameters.

Extend Tag

A right click on the first peak in the series also allows the following options to be selected from the pop up menu displayed.

Step Similar To Repeat Find Tag except that the sub-sequence is only extended by 1 residue from the selected peak of interest.

Leap Allows the program to leap or jump a possible missing peak giving up to 2 residues from the selected peak of interest.

MolWt Allows the program to move from the selected peak of interest to either the maximum or minimum allowed mass for that ion series i.e. **max** or **min**. The mass difference between the selected peak of interest and the min or max determines the possible amino acid combinations that can fill this mass difference.

Undo Tag

To undo a the previous Find tag, Repeat find tag or Extend tag operation, click with the right mouse button anywhere on the spectrum pane and select **Undo tag** from the pop up menu displayed.

Copy Tag

Right click anywhere on the spectrum pane and select **Copy tag** to copy the current sub-sequence tag to the clipboard. The format is *m1 tag m3* separated by tabs, where *m1* is the first assigned peak from the low mass end, *tag* is the residue sub-sequence read from low to high mass and *m3* is the last assigned peak at high mass. This tag can be pasted into the database search engine – ProteinProbe for peptide retrieval.

The Tree Views Pane

There are two tree views, **TreeViewCT** for High to Low mass (or towards the C-terminal end of the spectrum)and **TreeViewNT** (or towards the N-terminal end of the spectrum) for Low to High mass.

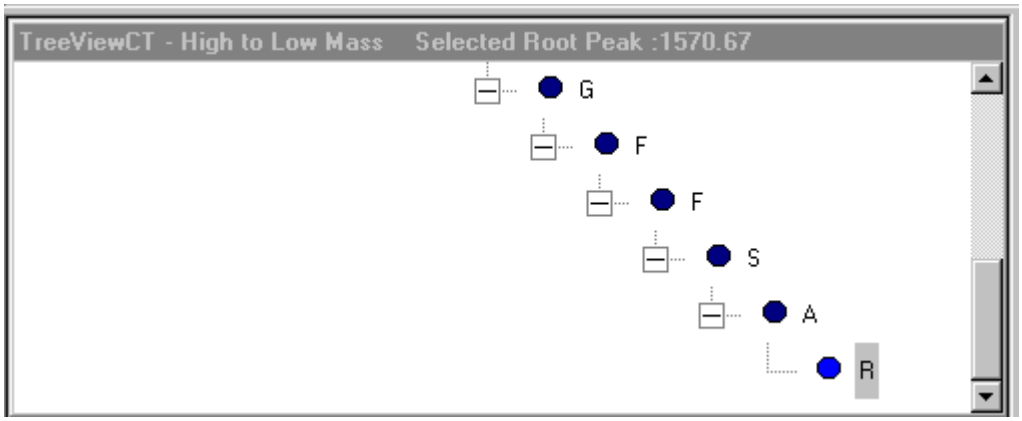


Figure 2.14 The Tree View Pane

The tree view allows for tag sequences to be extended without loosing previous possibilities. Previous sequences can be revisited and new possibilities can be explored. Each node in the tree is collapsible, by checking the box, thus allowing the whole view to be simplified.

A particular sequence can be extended by:-

- Selecting a particular residue from the tree view. When a new sequence is selected the other three panes, will change to reflect analysis details for that sequence.

- Right click in the pane to bring up the Tree View menu (Figure 2.15) and choose extend tag (step or leap). The new sequences are added to the tree and shown in the sequence pane and spectrum pane.

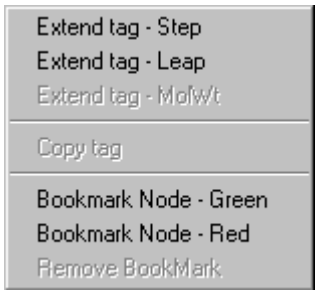


Figure 2.15 Tree View Menu

These new sequences can be further explored, and interesting nodes on the tree can be bookmarked in either green or red. Revisited nodes, further back on the tree structure can be

double clicked for all the sequence possibilities to be re-shown in the sequence pane and extended again.

The Sequence Report Pane

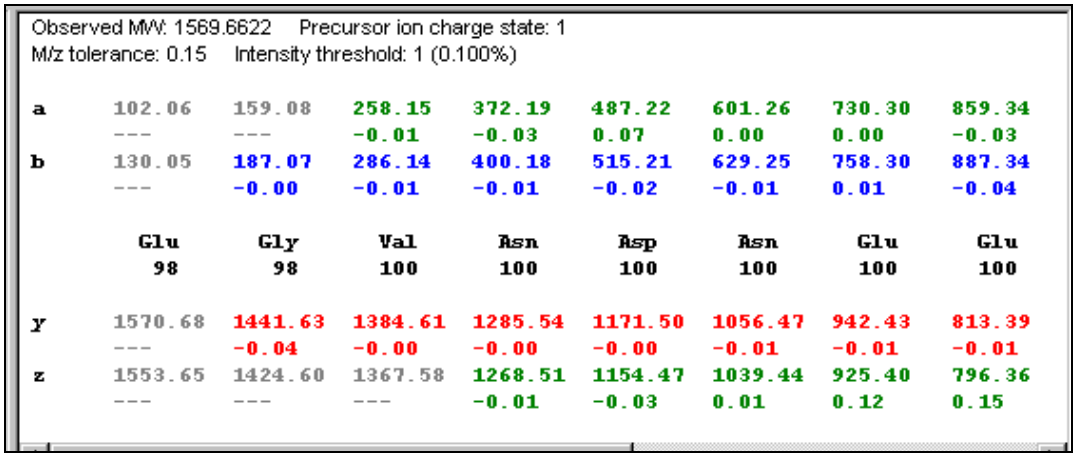


Figure 2.16 The Sequence Report Pane, showing a Sequence Prediction Details Report

The sequence report pane (Figure 2.16) contains the text-based report describing analysis details for the selected sequence. It shows the matching details and mass deltas from the expected and observed fragment ions. Currently only the 'a', 'b', 'y' and 'z' ions are detailed for the sequence. Matching 'b' ions are displayed in blue, y' ions in red and x' and z' in bold.

If **MassSeq** has been used to sequence the spectrum then residue probabilities are displayed below the three-letter code of the amino acids.

The Toolbar

Figure 2.17 The Toolbar

The **Standard Toolbar** provides shortcuts for commonly used **File** and **Edit** menu commands.

Press the **New File** button to create a new, empty Peptide Sequencing document. This document must be configured with a centered spectrum and parameters before sequencing can proceed. (Menu equivalent: **File,New**).

Press the **Open File** button to invoke the File Open dialog. This retrieves a previously saved Peptide Sequencing document and opens a new child window containing it. By convention, Peptide Sequencing documents use the file extension **.psq**, so the File Open dialog is pre-loaded with all of the existing **.psq** files. (Menu equivalent: **File,Open**).

Press the **Save File** button to save the currently selected Peptide Sequencing document. If the document has not previously been saved the **File Save As** dialog will be invoked to prompt for a name. While any name may be chosen, by convention, the extension **.psq** should be used. (Menu equivalent: **File, Save** or **File, Save As**, as appropriate).

Press the **Load Spectrum** button to invoke the Spectrum Data Browser to select a MassLynx spectrum. If the Peptide Sequencing document already contains a spectrum, you will be asked if you wish to replace this data. If you press the **Yes** button, the Spectrum Browser will be displayed to allow a new spectrum to be selected.

Only centered spectra may be selected; the Peptide Sequencer program cannot process continuum data. When a new spectrum has been selected sequencing results will be deleted since they were based on the previous spectrum and are no longer consistent.

Press the **Parameters** button to invoke the **Peptide Sequencing Parameters** dialog and specify sequencing parameters and conditions. (Menu equivalent: **Process, Parameters**).

Press the MassSeq button to invoke the automatic sequencing program. (Menu equivalent: Process**, Predict Sequence**).

Press the **Copy** button to make a copy of the currently selected pane in the Windows Clipboard. For the sequence pane the text is copied, for the Sequence Report and Spectrum panes a picture is copied. E.g. to copy the list of sequences to the Clipboard, click in the **Sequence Pane** to select it, then press the **Copy** button. The contents of the pane are copied into the Clipboard as a Windows metafile and may be pasted into other applications, which accept metafile formats. (Menu equivalent: **Edit, Copy**. Note that the text of the menu item will change to reflect the currently selected pane i.e. **Edit, Copy Spectrum** if the spectrum pane is selected).

Press the **Paste Spectrum** button to retrieve a centered mass spectrum from the Clipboard and replace the current spectrum. The Clipboard spectrum must be in the form of a peak list of m/Z, intensity pairs, one pair per line and in increasing m/Z order. Load the text file into Notepad or some other text editor, select all the peaks, and copy them to the Clipboard. The **Paste Spectrum** button will then be enabled, and the spectrum can be pasted into the Peptide Sequencing document. (Menu equivalent: **Edit, Paste Spectrum**).

Press the **Print** button to print a report according to the currently selected reporting options (see **Report Options**, below). (Menu equivalent: **File, Print**).

Press the **O^{16}/O^{18}** button to predict sequences for a spectrum representing data obtained from a digest carried out in O^{18} water using the current parameters. (Menu equivalent: **Process, Isotopic Labelling)**.

Press the **Display All Peaks** button to show every peak in the spectrum, regardless of intensity.

Press the **Display Peaks AboveThreshold** button to show only those peaks with intensity exceeding the threshold set in the sequencing parameters. The peaks displayed are those which will be used in sequencing; any peaks of lower intensity will be ignored.

Press the **Display 'b' ion sequence** button once to display the 'b' ion sequence connection lines and assignments to the data. Press it again to turn the display off.

Press the **Display 'y' ion sequence** button once to display the 'y' ion sequence connection lines and assignments to the data. Press it again to turn the display off.

Press the **Zoom Out** button to restore the spectrum display to full M/z range

The **Zoom In** button decreases the displayed range by half, around the center of the current range. For example, if the current range is 100 < m/Z < 500 (a 400 Da range, centered at 300 Da) the new range will be 200 < m/Z < 400 (a 200 Da range, centered at 300 Da). This is a quick way to zoom in to the currently displayed range.

The **Display Tree Views** button. Press the button to reveal the tree views and again to turn the display of (Menu equivalent**: Options, Show/Hide Trees**)

Menu Commands

The toolbar provides shortcuts for most menu items. There are a few additional menu items that have no toolbar shortcuts:

File Menu

Close — Closes the currently selected Peptide Sequencing document and removes any Child Windows viewing it. If the document has not previously been saved, you will be prompted to save it and to provide a name. If the document has been previously saved, you will be prompted to save the changes.

Load Peaklist file — Allows a mass/intensity peak list to be loaded into the spectrum view. Two file formats are available.

***.pkl** (Micromass proprietary format) which is exported using the ProteinLynx software.

***.dta** (SEQUEST format file) which is exported from the Spectrum utility within MassLynx or from ProteinLynx automation.

Print Setup — Invokes the **Print Setup** dialog to allow selection of the printer and printer properties.

Print Preview — Displays a preview of the printed report for the current document.

Load on Startup — Allows the last saved file to be opened on running the PepSeq application. If this option is selected a tick mark will appear next to it.

Most Recent Files — This list contains the names of the last four Peptide Sequencing documents viewed. To open any of them in a new window simply select it from the list.

Exit — Exit the Peptide program. Any open documents will be closed (you will be prompted to save them where appropriate).

Report Options

Invokes the **Report Options** dialog (Figure 2.18). This dialog is used to select the types of printed reports and their appearance. It is a *tab dialog* with two pages. You may specify the report types on the first page, shown in Figure 2.18.

Reports

Sequences with scores — Check this box to print a report listing the sequences and their scores.

Selected sequence — Check this box to print reports for the currently selected sequence, or **All Sequences** to print reports for all sequences.

Tabulated fragment ion details Check this box to print the expected and observed fragment ion details report. (The Sequence pane details).

Spectrum plot with assignments Check the box to print an annotated spectrum. Any annotation options or magnification ranges specified for the **Spectrum** pane are also applied to the printed report.

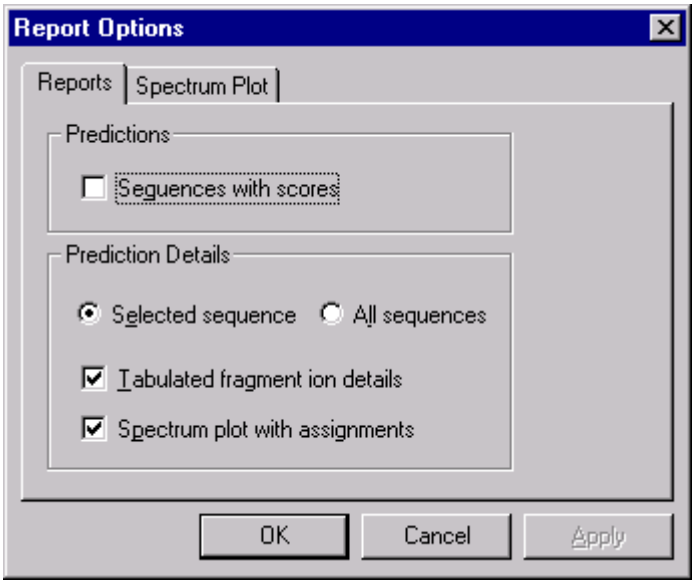


Figure 2.18 The Report Options dialog, Reports page

Spectrum Plot

The appearance and spectral range may be specified independently for the printed report without affecting the **Spectrum** pane using the second page, shown below.

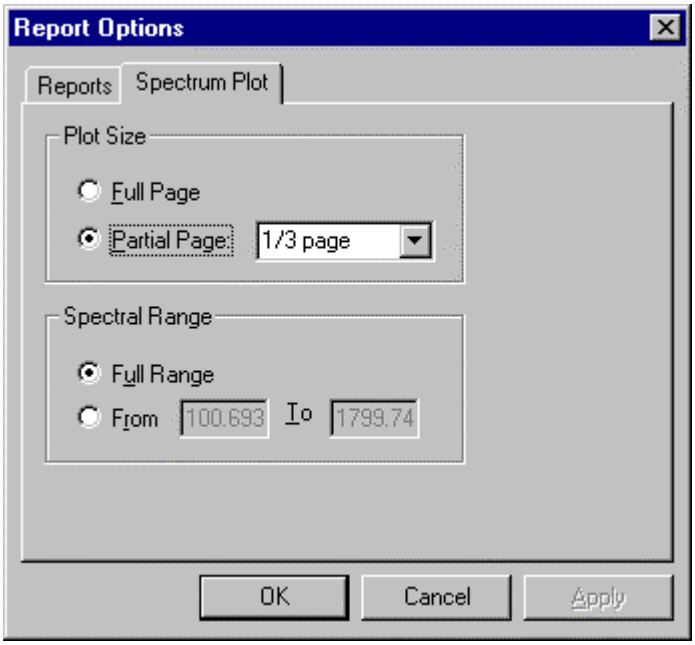


Figure 2.19 Report Options dialog, Spectrum Plot page

Plot Size Select **Full Page** or **Partial Page**. If **Partial Page** is selected, you can select one of Full Page, ½ page, ⅓ page or ¼ page.

Spectral Range Select **Full Range** or **From** and **To**. If **From** and **To** are selected enter the M/z range, in Da, of the range required.

View Menu

Options Invokes the **View Options** dialog. It is a *tab dialog* with two pages. You may specify the spectrum view options on the first page, shown below.

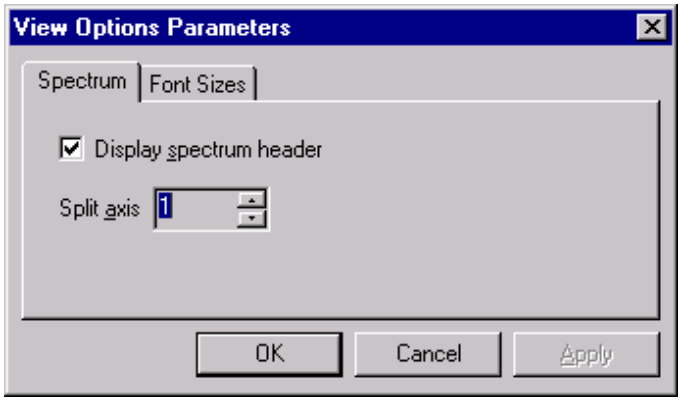


Figure 2.20 The Spectrum View Options dialog

Display spectrum header Check this box to display the spectrum header information at the top of the spectrum.

Split Axis Use the buttons to view finer details. This will display the same mass range spectrum over the number of axis specified.

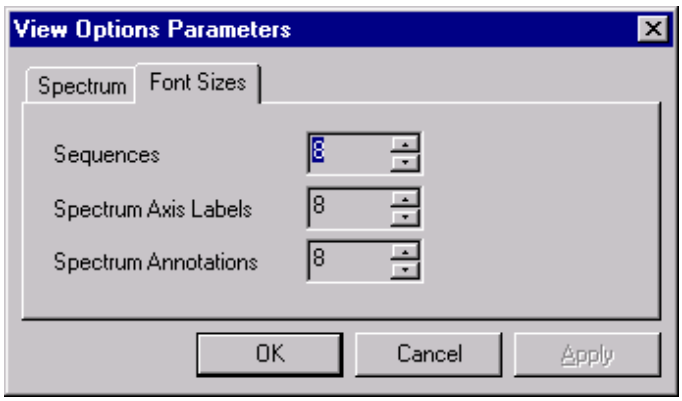


Figure 2.21 The View Options Font Sizes dialog

Font Sizes Can be altered for the report pane, spectrum axis labels and spectrum annotations.

Toolbars When ticked, the toolbar will be made visible. If not ticked, the toolbar will not appear on screen.

Status Bar When ticked, the status bar at the bottom of the Main Window is displayed. It is used to display the program state and to provide more information about menu commands and toolbar buttons. To see more information about a menu item, pull down the menu and highlight the desired item; to see toolbar button information, position the cursor over the button. Descriptive information will then appear in the status bar.

Options

Modifiers Invokes the Modifiers dialog box. See Modifications Page on Page 2-9 for details.

BLAST Invokes the Blast Parameters dialog (Figure 2.22).

Show Tree / Hide Tree Shows and Hides the tree views.

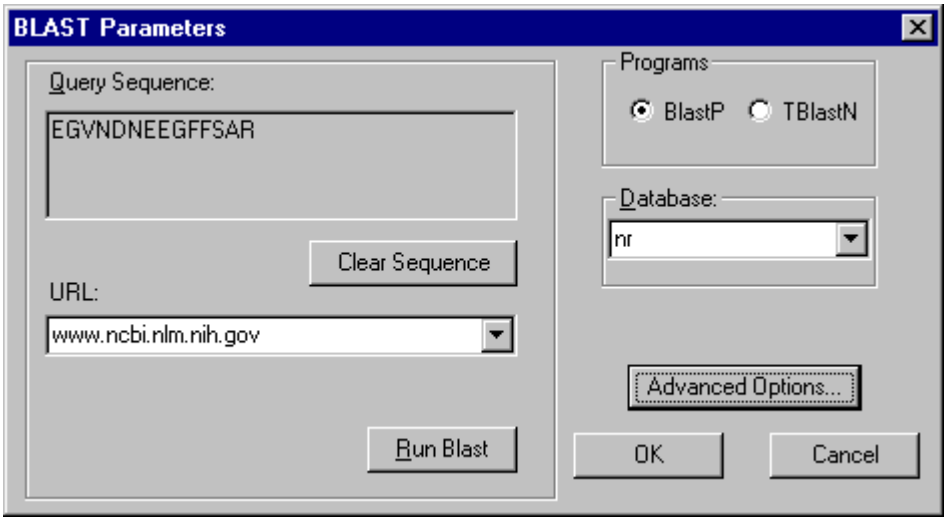


Figure 2.22 Blast Parameters dialog

Tools

The **lookup.dat** file contains the molecular weights of all amino acid combinations and the user-specified modified residues. It is used as a quick lookup table for amino acid combinations representing mass differences between peaks. Selecting **Build Lookup File** displays the following dialog

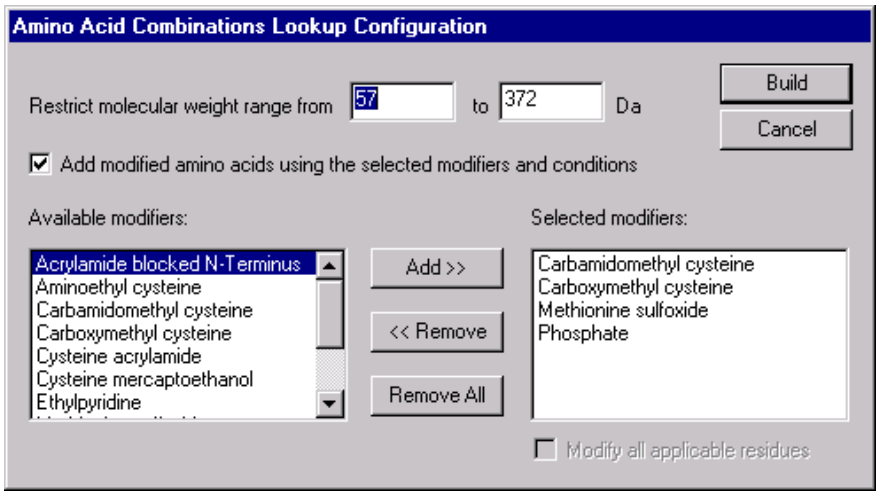


Figure 2.23 The Amino Acid Combinations Lookup Configuration dialog

The *lookup.dat* file installed with MassLynx is configured as above.

Restrict molecular weight range	The range of amino acid combinations can be restricted by mass. The **from** mass should be 57 or greater. Masses greater than 57 will result in peak differences of 57 not being represented by any amino acids (e.g. Glycine = G). The upper mass **to** has no limit but computations will take longer the higher the mass.
Add modified amino acids	Modifications can be taken into account in finding residues that are modified by checking this box. The list of possible modifications will be enabled if this box is checked. To add new modifiers to the **Selected Modifiers** list click on one of the modifiers in the **Available modifiers** list and press the **Add>>** button. To delete modifiers from the **Selected Modifiers** list, click on one of the modifiers and press the <<**Remove** button. To **Remove All** modifiers from the **Selected Modifiers** list press the **Remove All** button. For each individual modification in the **Selected Modifiers** list check the **Modify all applicable residues** checkbox if the modification is mandatory i.e. always modified. If unchecked the residue mass will be stored as well as the mass of the modified residue i.e. optional modification.
Build	Pressing the Build button displays the progress dialog to indicate the status of the residue combination calculations. This process can be stopped by pressing the **Stop** button. If stopped then all combinations already calculated will be stored and the *lookup.dat* file will be updated
Cancel	Choosing cancel closes the dialog box without affecting the *lookup.dat* file.

Window Menu

New Window	Creates a new Child Window containing a second view of the currently selected Peptide Sequencing document. Results for a different selected sequence may be viewed in the second window; however, changes to the *document* contents (such as parameters or re-sequencing) will affect the contents of both windows.
Cascade	If more than one Child Window is open, this command will resize and arrange them in an overlapping cascade within the Main Window.
Tile Horizontally	If more than one Child Window is open, this command will resize and arrange them so that they are stacked one above the next within the Main Window.
Tile Vertically	If more than one Child Window is open, this command will resize and arrange them side-by-side in the Main Window.
Arrange Icons	If any Child Windows are minimized, this command will arrange their icons along the bottom of the Main Window.
Open Documents List	The list of Child Windows containing open Peptide Sequencing documents appears at the end of the Window menu. The currently active Child Window is ticked. Select a different Child Window to make it active and bring it to the top of the stack of overlapping Child Windows.

References

1. Eng, J.K., McCormack, A.L. and Yates, J.R, *J Am Soc Mass Spectrom.* **1994**, 5, 976-989.

Chapter 3 BioLynx - Database Searching

Contents

Figures 3-2
Overview 3-3
Using the Database Searching Program 3-3
Creating a New Document 3-3
Loading a Spectrum 3-4
Database Setup 3-4
    Exclude from search dialog 3-6
    Saving/Loading Query parameters 3-6
    Carrying out a Search 3-7
The ProteinProbe™ Interface 3-7
    ProteinProbe Documents 3-7
    The Main Window 3-7
    Child Windows 3-7
Toolbar 3-10
    The Toolbar 3-11
Menu Commands 3-13
    File Menu 3-13
    View Menu 3-14
    Options Menu 3-15
    Window Menu 3-15
References 3-16

Figures

Figure 3.2 ProteinProbe digest query 3-3
Figure 3.3 Database Setup page 3-4
Figure 3.4 Exclude from search dialog 3-6
Figure 3.5 The ProteinProbe program Main Window 3-8
Figure 3.6 The Search Results pane showing hit list and search parameters 3-8
Figure 3.7 The Query Pane showing a summary of the query details 3-9
Figure 3.8 The Entry Text Report pane showing a report of a FASTA entry as stored in the database 3-9
Figure 3.9 The Spectrum Pane 3-10
Figure 3.10 The Add Peak Mass pop up menu 3-10
Figure 3.11 The Toolbar 3-11
Figure 3.12 The Entry Text Report pane showing a report of the actual entry as stored in the database 3-12
Figure 3.13 The View Match Details Report Pane 3-12
Figure 3.14 The Peptide Map Report pane showing highlighted sub-sequences 3-13
Figure 3.15 Report Options page 3-14
Figure 3.16 Font Sizes dialog 3-15
Figure 3.17 The Web Search Server Setup dialog 3-15

Overview

The Database searching program (ProteinProbe™) allows user defined peptide queries to be built from MS and MS/MS spectra. These queries are then used to search FASTA format databases.

Briefly, the search program uses a query file generated by the user, which can be saved to disk. The search results include the list of hits, the original query and any peptide matching details (if applicable). These results can be saved to disk as a document. The ability to load previous queries or documents allows the user to re-search a query against an updated or newer database.

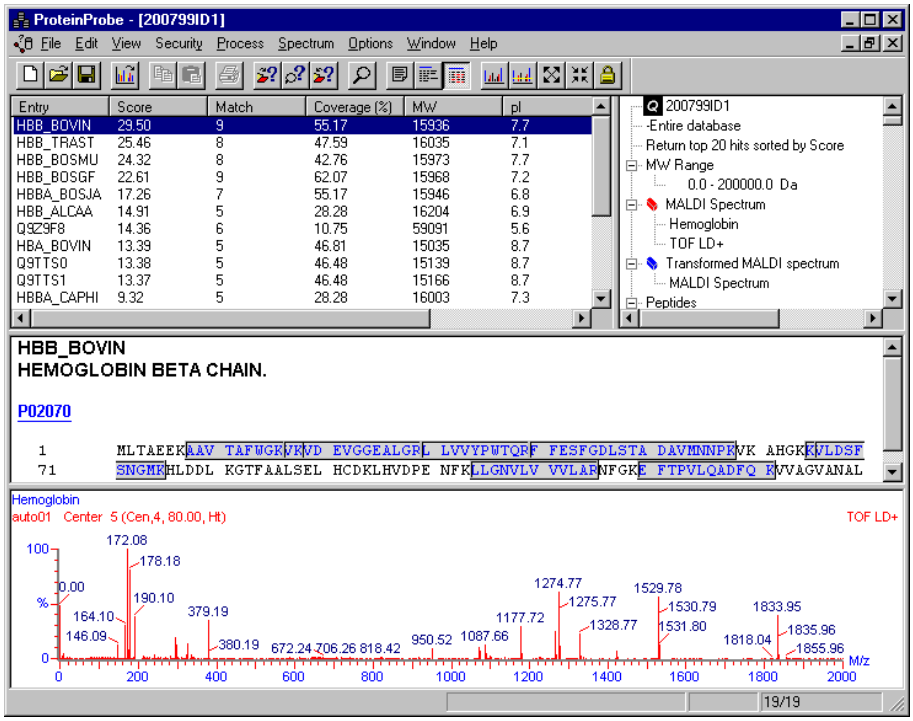


Figure 3.1 ProteinProbe digest query

Using the Database Searching Program

Select **BioLynx - Database Search** from the MassLynx Shortcut Bar**.** This brings up a new document entitled “PtProb1”. Several configuration files are required for the program to execute properly and are installed during MassLynx installation in the MassLynx\Pepembl directory. These files are **aminos.dat**, **modifier.dat**, **reagents.dat** and **scores.dat**. If for some reason the registry has become corrupted, a message box will appear on running the program, asking the user to specify the location of these files.

Creating a New Document

Select **New** from the **File** menu or press the button on the toolbar. An empty Child Window will be created named “PtProb2”. This new window can be displayed along side the existing document to compare results of similar searches.

Loading a Spectrum

To load a spectrum press the toolbar button or select **Load Spectrum** from the ProteinProbe **File** menu. Select a centered or MaxEnt 3 spectrum from the dialog displayed. Only M@LDI and MS/MS spectra can be loaded.

Database Setup

Before a search can be carried out a search query has to be set up. To open the dialog, select **Setup Search Query** from the **Process** menu or press the button.

Database Connection

If ProteinProbe and MassLynx are installed on the same PC, the default local host Web Search Server URL: 127.0.0.1 is applicable.

If ProteinProbe and MassLynx are not installed on the same PC, you may need to amend the URL of the Web Search Server. From the BioLynx Database Search dialog, select **Web Server** from the **Options** menu and enter the correct **Web Search Server URL**.

Note: To view the Database Setup page, you must have an Internet connection and Internet Explorer version 5, or higher, must be installed.

Database Setup Page

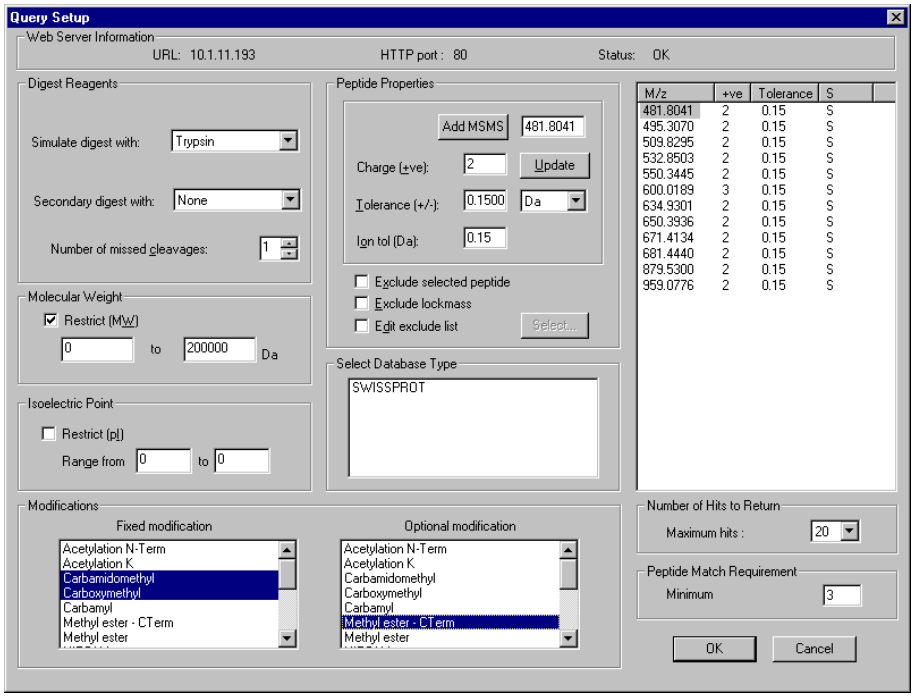


Figure 3.2 Database Setup page

Digest Reagents

Simulate digest with If the peptide masses entered are the result of an enzymatic or chemical cleavage, then select an enzyme from the drop down list box.

Secondary digest with	Select a secondary digest reagent from the drop down list box, if applicable.
Number of missed cleavages	Press the arrows to take into consideration missed cleavage sites. A maximum of 2 missed cleavage sites is currently allowed. If you do not want to consider missed cleavage sites then set this to 0. Computing all possible nearest neighbor partial fragments adds significantly to the number of peptides entered in the database (by a factor of two). The major effect of this is to increase the background score by increasing the number of random molecular weight matches, which can significantly reduce discrimination.

Peptide Properties

Add MSMS	You can build an MSMS Query by adding further MSMS spectra to the list. Press the Add MSMS button and use the Spectrum Data Browser dialog to load a Mass Measured or Maxent 3 spectrum.
Charge (+ve)	Enter a charge state and press the **Update** button to apply the charge state to a peptide mass peak.
Tolerance (+/-)	Enter an allowed mass error and select **%**, **ppm** or **Da** from the drop down list box.
Ion tol (Da)	Enter the tolerance of the ions in a fragmented peptide daughter spectrum.
Exclude selected peptide	Check this box to exclude from the search the peptides that are defined in the Exclude from search dialog (see Figure 3.3).
Exclude lockmass	Check this box to exclude from the search any lockmass peaks found in the data
Edit exclude list	Check this box and press the **Select** button to display the **Exclude from search** dialog.

Molecular Weight

Restrict (MW)	Check this box if there are restrictions on the molecular weight of the peptide and enter a **Range from** and **to** in the relevant boxes.

Isoelectric Point

If there are restrictions on the Isoelectric Point, check this box and enter values in the **Range from** and **to** boxes.

Select Database Type

Select Database Type	A list of the databases available on the server is displayed. Select the database that you want to search.

Modifications

Modifications can be set up for individual or groups of residues. For example, you can specify that the serine, threonine and tyrosine residues are phosphorylated. This modification will then be taken into account during the search.

Fixed modification	If there are modifications to the residues, select the modifications that you would like to be considered during the search.
Optional modification	If there are suspected modifications to the residues, select the modifications that you would like to be considered during the search, in addition to searching for the normal mass.

Number of Hits to Return

Maximum hits	From the drop down list box, select the maximum number of hits to display.
Peptide Match Requirement	Enter the minimum number of peptides that a protein must contain in order to be returned as a 'hit'.

Exclude from search dialog

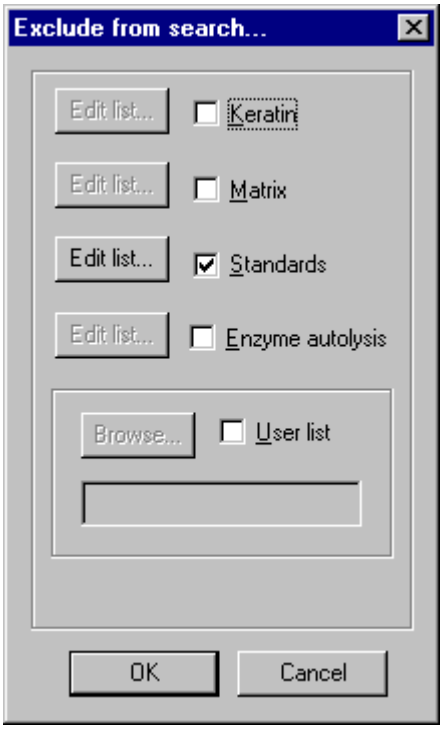


Figure 3.3 Exclude from search dialog

Check the box next to the relevant peptide group and press the associated **Edit list** button. Check the box(es) on the dialogs displayed to exclude specific masses from the search.

Saving/Loading Query parameters

Query parameters setup, according to the Database Setup page can be saved as a query file. Select **Save Query** from the **File** menu or click the toolbar button. Enter a name in the **File**, **Save As** dialog and press the **Save** button. Query files should only be saved as a precaution against the

possibility of losing the query parameters due to unforeseen circumstances. **Note: If you perform a search and save the results the query parameters are saved as part of the document.**

To load a previously defined query file, select **File, Load Query** or click the toolbar button. Select a query file from the File Open dialog and press the **Open** button.

Carrying out a Search

To carry out a search first ensure that a query has been loaded from disk or entered using the query pages. Select **Search** from the **Process** menu or click on the toolbar button. A progress status dialog box indicates the total number of entries, the number of hits found and an approximation of how long the search will take. A **Stop** button allows the search to be terminated in which case only hits already found will be displayed in the results and hit list window.

The ProteinProbe™ Interface

ProteinProbe Documents

Information in the program is organized into documents. A document, when complete, contains all of the information produced as a result of a search. This information includes all the query parameters, hit list, search results and scores.

The program provides an interactive, graphical way to build and view these documents.

The Main Window

The program consists of a **Main Window** containing a **Menu**, optional **Tool Bars**, and a **Status Bar**. Within the Main Window, one or more **Child Windows** may be positioned, each containing a multiple-part view of a set of search results. The parts of the Main Window and a Child Window are shown in Figure 3.4 below.

Child Windows

As described above, the Main Window could contain one or more Child Window. Each Child Window represents a view into part of a document. Using commands from the Menu, a Child Window may be cloned into two or more Child Windows, each giving an independent view of the same document. This allows results of two different search hits to be studied side-by-side. Alternatively, two documents may be opened simultaneously, permitting side-by-side comparison of search results derived from different parameter settings.

Each Child Window is divided into four panes, the search results and the query at the top of the screen, several detailed reports in the middle and a spectrum at the bottom.

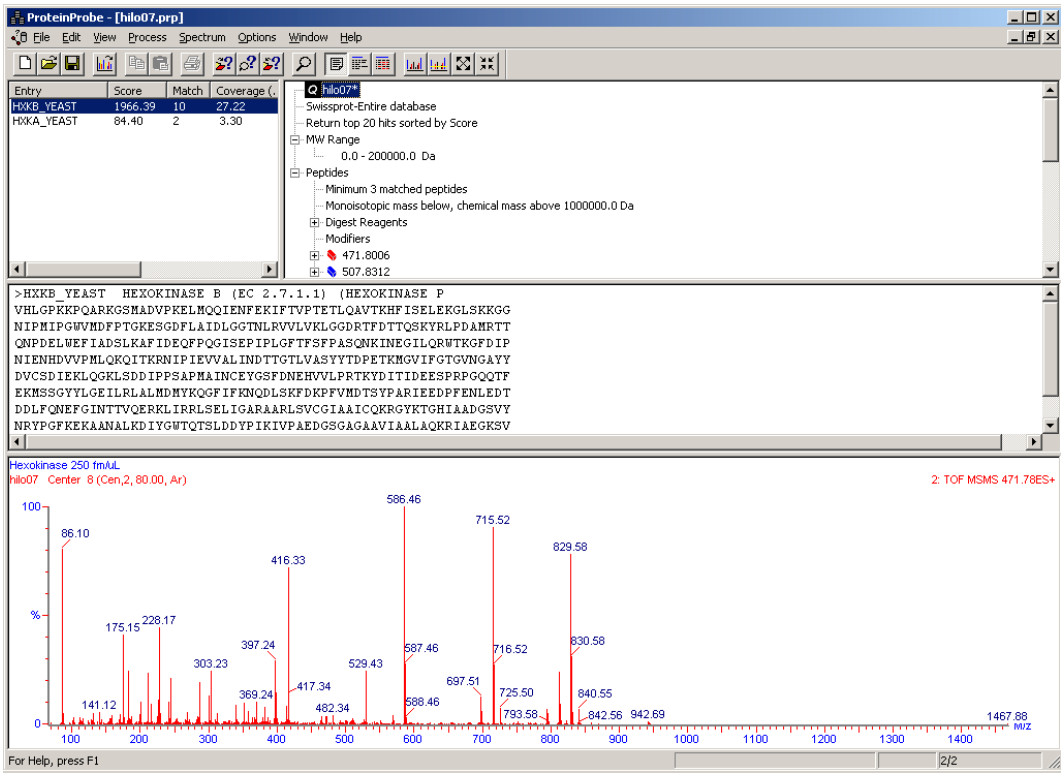


Figure 3.4 The ProteinProbe program Main Window

The Search Results Pane

Entry	Score	Match	Coverage (...	MW	pI	Description	Reading Frame
BGAL_ECOLI	929.60	6	6.35	116333	5.4	BETA-G...	0
BGAL_KLEPN	220.85	2	1.64	117499	5.8	BETA-G...	0
BGAL_PSBAT	145.10	1	0.77	118050	5.2	BETA-G...	0
BGAL_ENTCL	75.75	1	0.88	116660	6.1	BETA-G...	0

Figure 3.5 The Search Results pane showing hit list and search parameters

The Search Results pane displays the hit list showing the ID and name of the database **Entry**.

The hits may be sorted in the following ways:

- By **Entry** i.e., entries displayed as extracted from sequence database.
- The sequence that is the most likely to match based on the number of matching peptide fragment masses, percent coverage and mass difference. (**Score**).
- The number of matching peptide fragment masses (**Match**).
- The % **Coverage** of a particular entry (this becomes an important indicator when a high number of peptide masses are searched). This option compensates for high molecular weight entries.
- By **pI** (calculated Isoelectric Point).

To sort entries by a method click on the column heading.

An entry is selected by clicking on it. The currently selected entry is highlighted and the contents of the other panes will display analysis details for that entry.

The Query Pane

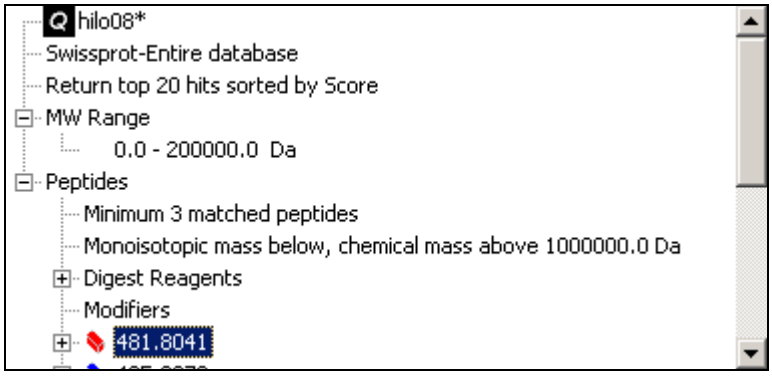


Figure 3.6 The Query Pane showing a summary of the query details

The **Current Query** pane displays the search query parameters used in the search.

Select **Partially Expanded Query** from the **View** menu to display details for each page of the query. Select **Fully Expanded Query** from the **View** menu to show page details and peptide details.

Click on a ⊞ box to display more details for the page or parameter, or a ⊟ box to show less.

For MS data the query will contain one spectrum. For MS/MS data there will be one spectrum for each mass. The icon of the spectrum displayed in the spectrum pane is shown in red in the query pane, the icons for other spectra are shown in blue.

The Entry Report Pane

Figure 3.7 The Entry Text Report pane showing a report of a FASTA entry as stored in the database

This pane contains one of three text-based reports describing analysis details for the currently selected entry. These reports are

- The Entry Text report.
- The View Match Details Report showing peptide fragment match details.
- The Peptide Map Report displaying a coverage map of sequence.

The report type displayed can be changed using the toolbar buttons or **View** menu commands.

The Spectrum Pane

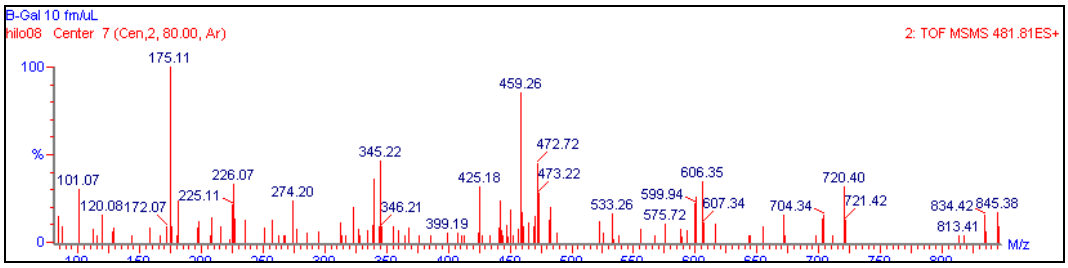


Figure 3.8 The Spectrum Pane

The bottom pane contains a plot of the spectrum, annotated with details from the currently selected consensus sequence. The annotations which appear on the spectrum can be specified using the toolbar or menu commands described on page 3-13

For MS Spectra a right mouse click on the spectrum window will display a pop up menu, allowing the user to add or delete a peak mass to or from the current query.

Figure 3.9 The Add Peak Mass pop up menu

A peak will be colored black if its mass is listed in the current query, or blue if a match is found after a search and the user chooses **View**, **Match Details**.

To Add a Peak Mass

1. Right click on a peak in the spectrum.
2. Select **Add peak mass** from the pop up menu.
3. Select the required **Charge**.

To Delete a Peak Mass

1. Right click on a peak in the spectrum.
2. Select **Delete peak mass** from the pop up menu.

Toolbar

To display the toolbar, select **Toolbar** from the **View** menu. The toolbar can be docked against one of the sides of the Main Window, or can be left floating.

To dock a floating toolbar or to change the location of a docked toolbar, simply click on the bar in the background area (between or surrounding the buttons) and drag it. An outline of the bar will appear, to aid in positioning it. When a floating toolbar is brought near the edge of the Main Window, the outline will change to show its orientation when docked. Releasing the mouse will dock the toolbar.

To float a docked toolbar, click and drag it off the toolbar. Move it away from the edge of the Main Window and release the mouse. The toolbar will float in a “mini-window” frame, as shown in the figures below.

The Toolbar

Figure 3.10 The Toolbar

The **Toolbar** provides shortcuts for commonly used menu commands.

Press the **New Document** button to create a new, empty document. This document must be configured with query parameters before searching can proceed. (Menu equivalent: **File... New**).

Press the **Open Document** button to invoke the File Open Dialog to retrieve a previously saved ProteinProbe document and open a new child window containing it. By convention the documents use the file name extension ***.prp**, so the Open an existing document dialog displays the existing ***.prp** files. (Menu equivalent: **File, Open**).

Press the **Save Document** button to save the currently selected document to a file. If the document has not previously been saved, the Save the active document dialog will be invoked to prompt for a name. While any name may be chosen, by convention, the extension ***.prp** should be used. (Menu equivalents: **File, Save** or **File, Save As**, as appropriate).

Press the **Load Spectrum** button to invoke the Spectrum Data Browser dialog. This allows a spectrum to be loaded into the ProteinProbe document. (Menu equivalents: **File, Load Spectrum**).

Press the **Copy** button to make a copy of the currently selected pane in the document window on the Windows Clipboard. For example, to copy the list of hits to the Clipboard, click first in the **Search Results Pane** to select it, then click the **Copy** button. The contents of the pane are copied into the Clipboard as a Windows metafile, and may be pasted into other applications that accept metafile formats. (Menu equivalent: **Edit, Copy**).

Allows for a spectrum from MassLynx to be pasted into ProteinProbe. Select **Edit,Copy Spectrum List** in MassLynx Spectrum select the spectrum pane and click this button.

Press the **Print** button to print a report according to the currently selected reporting options (see **Report Options**, below). (Menu equivalent: **File, Print**).

Press this button to load a previously defined query. (Menu equivalent: **File, Load Query**).

Press this button to invoke the Specific Query dialog. (Menu equivalent: **Process, Specify Query**).

Press this button to invoke the Save Query Template dialog. (Menu equivalent: **File, Save Query**).

Press this button to perform a Search. (Menu equivalent: **Process, Search**).

Text Reports

The **Text Report** toolbar buttons change the details displayed in the lower, Entry Report pane. The Entry Report pane always shows a report for the currently selected entry. The report may be one of three types.

Pressing the **Entry Text Report** button displays the actual entry text details. (Menu equivalent: **View, Entry Text**) – see **Figure 3.11**.

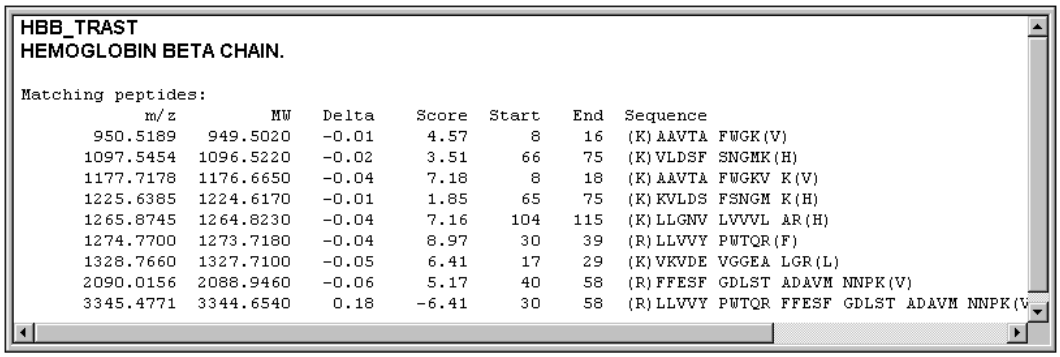
This report displays the information as it is held on the database.

Figure 3.11 The Entry Text Report pane showing a report of the actual entry as stored in the database

Pressing the **View Match Details** button displays the peptide fragment masses that match theoretical peptide fragments from the specified entry. (Menu equivalent: **View, Match Details**) – see Figure 3.12.

This report displays the Entry name and ID and a list of matching fragment masses with their molecular weight (**MW**), mass error (**Delta**), the position of the peptide in the sequence (**start** and **end**) and the **sequence**. If a * appears to the left of one of these entries then it is a single partial fragment (i.e. one missed cleavage site), if ** appears it is a double partial fragment (i.e. two missed cleavage sites).

The matching peptide list is followed by a list of unmatched peptides.



HBB_TRAST
HEMOGLOBIN BETA CHAIN.

Matching peptides:

m/z	MW	Delta	Score	Start	End	Sequence
950.5189	949.5020	-0.01	4.57	8	16	(K)AAVTA FWGK(V)
1097.5454	1096.5220	-0.02	3.51	66	75	(K)VLDSF SNGMK(H)
1177.7178	1176.6650	-0.04	7.18	8	18	(K)AAVTA FWGKV K(V)
1225.6385	1224.6170	-0.01	1.85	65	75	(K)KVLDS FSNGM K(H)
1265.8745	1264.8230	-0.04	7.16	104	115	(K)LLGNV LVVVL AR(H)
1274.7700	1273.7180	-0.04	8.97	30	39	(R)LLVVY PWTQR(F)
1328.7660	1327.7100	-0.05	6.41	17	29	(K)VKVDE VGGEA LGR(L)
2090.0156	2088.9460	-0.06	5.17	40	58	(R)FFESF GDLST ADAVM NNPK(V)
3345.4771	3344.6540	0.18	-6.41	30	58	(R)LLVVY PWTQR FFESF GDLST ADAVM NNPK(V

Figure 3.12 The View Match Details Report Pane

Pressing the **Peptide Map** button displays the sequence specific to that entry. Currently this display is only used for highlighting residues from a global sub-sequence search but in future releases peptide fragments will be underlined or highlighted (**Figure 3.13**). (Menu equivalent: **View, Peptide Map**).

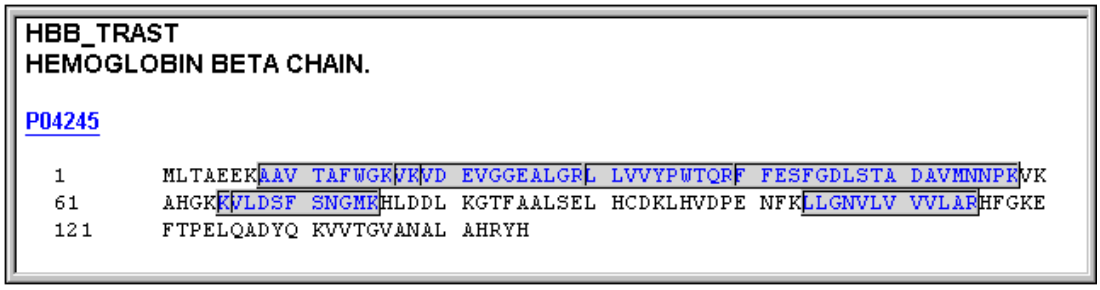


Figure 3.13 The Peptide Map Report pane showing highlighted sub-sequences

Spectrum Manipulation

The **Spectrum Manipulation** toolbar buttons provide shortcuts to control the threshold and range of the spectrum display.

Press the **Display All Peaks** button to show every peak in the spectrum, regardless of intensity. (Menu equivalent: **Spectrum, Display Peaks, All Peaks**).

Note: This button is only enabled when MS/MS spectra are displayed.

Press the **Display Peaks Above Threshold** button to show only those peaks with intensity exceeding the threshold. (Menu equivalent: **Spectrum, Display Peaks, Peaks Above Threshold**).

Note: This button is only enabled when MS/MS spectra are displayed.

Press the **Zoom Out** button to restore the spectrum display to full m/Z range. (Menu equivalent: **Spectrum, Zoom Out**).

The **Zoom In** button decreases the displayed range by half, around the center of the current range. For example, if the current range is 100 < m/Z < 500 (a 400 Da range, centered at 300 Da) the new range will be 200 < m/Z < 400 (a 200 Da range, centered at 300 Da). This is a quick way to zoom in to the currently displayed range. (Menu equivalent: **Spectrum, Zoom In**).

Menu Commands

The toolbar provides shortcuts for most menu items, and these have been described above. There are a few additional menu items that have no toolbar shortcuts:

File Menu

Close — Closes the currently selected document and removes any Child Windows viewing it. If the document has not previously been saved, you will be asked if you wish to save it and to provide a name. If the document has been previously saved, you will be asked if you wish to save the changes.

Open Workspace — This invokes the **Open Workspace** dialog, from which a Report file can be selected and viewed.

See the ProteinLynx Results chapter for a description of the workspace.

Close Workspace — This closes the Report file.

Export Entry Saves the currently selected entry, which is highlighted in the hit list, to a named file with default ***.emb** or ***.fas** extension, according to the format of the current database. This file can then be imported into the BioLynx Protein editor for further analysis or modification.

Print Preview Displays a preview of the printed report for the current document, according to the printing options. The display is as close as possible to what will actually appear on paper.

Print Setup Invokes the **Print Setup** dialog to permit selection of the printer and its properties.

Report Options Invokes the **Report Options** dialog. This dialog is used to select the types of printed reports and their appearance.

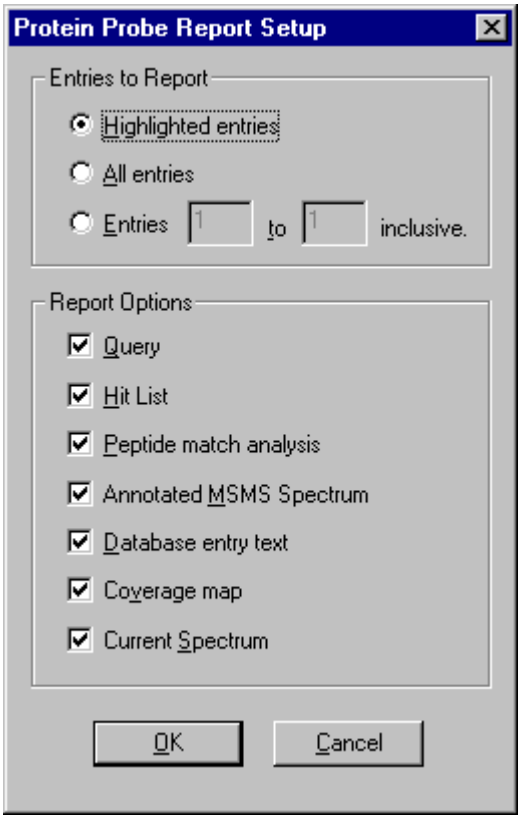


Figure 3.14 Report Options page

Most Recent Files This list contains the names of the last four documents you viewed. To open any of them in a new Child Window, simply select it from the list.

Exit Exit the program. Any open documents will be closed (you will be prompted to save them where appropriate). The layout and appearance of the Main Window, Toolbar, and Child Window will be saved to the Windows Registry.

View Menu

Font Sizes Invokes the Font Sizes dialog. Enter the font sizes required for the spectrum axis and annotation and press **OK**.

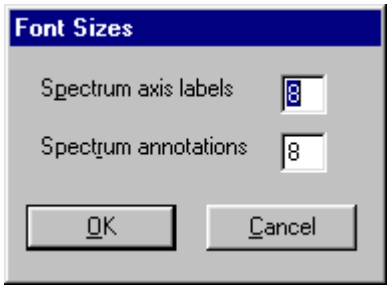


Figure 3.15 Font Sizes dialog

Toolbar Toggles the toolbar display. The menu item will be ticked if the toolbar is visible. The size and position of the toolbar is independent of visibility.

Status Bar When ticked, the status bar at the bottom of the Main Window is displayed. The status bar is used to display the program state and to provide more information about menu commands and toolbar buttons. To see more information about a menu item, pull down the menu and highlight the desired item; to see toolbar button information, position the cursor over the button. Descriptive information will then appear in the status bar.

Options Menu

From the Options menu, select Web Server. The Web Search Server Setup dialog is displayed.

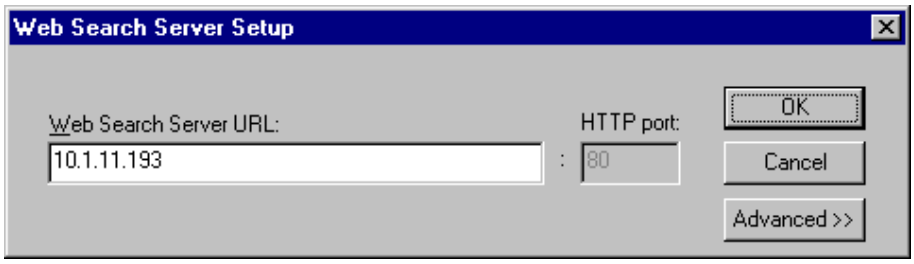


Figure 3.16 The Web Search Server Setup dialog

If ProteinProbe and MassLynx are installed on the same PC, enter the default local host Web Search Server URL: 127.0.0.1.

If ProteinProbe and MassLynx are not installed on the same PC, enter the URL of the Web Search Server.

It is recommended that you do not amend the **Advanced** options.

Window Menu

New Window Creates a new Child Window containing a second view of the currently selected document. Results for a different selected entry may be viewed in the second window; however, changes to the *document* contents (such as parameters or re-searching) will affect the contents of both windows.

Cascade If more than one Child Window is open, this command will resize and arrange them in an overlapping cascade within the Main Window.

Tile Horizontally	If more than one Child Window is open, this command will resize and arrange them so that they are stacked one on top of the next within the Main Window.
Tile Vertically	If more than one Child Window is open, this command will resize and arrange them side-by-side in the Main Window
Arrange Icons	If any Child Windows are minimized, this command will arrange their icons along the bottom of the Main Window.
Open Documents List	The list of Child Windows containing open documents appears at the end of the Window menu. The currently active Child Window is ticked. Select a different Child Window to make it active and bring it to the top of the stack of overlapping Child Windows

References

1. Bairoch, A and Apweiler, R., The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res.* 24:21-25 (1996).

2. Pappin, D.J.C., Hojrup, P and Bleasby, A.J., *Current Biology.* **1993**, 3, 327-332.

Chapter 4 BioLynx - CarboTools

Contents

Overview 4-3
Toolbars 4-3
Menu Commands 4-4
Creating a New Document 4-5
Configuration 4-5
Loading a Spectrum 4-6
Display Settings 4-6
Calculating Compositions 4-7
User Defined Modifiers and Terminal Groups 4-8
Manually Identifying Residues 4-9
Series Assignments 4-9
Configuring the Results Display 4-10
Deleting Results 4-10
Saving/Loading Parameters and Results 4-11

Figures

Figure 4.1 Carbohydrate Main Window 4-3
Figure 4.2 Configuration dialog 4-5
Figure 4.3 Spectrum Settings dialog 4-6
Figure 4.4 Carbo Composer dialog 4-7
Figure 4.5 The Edit User Defined Modifiers and Terminal Groups dialog 4-8
Figure 4.6 Add New Monsaccharide Edit box 4-8
Figure 4.7 Labelled Ion Series 4-10
Figure 4.8 Display Columns dialog 4-10

Overview

CarboTools is a program used in the interpretation of carbohydrate mass spectra. It allows the peaks of a spectrum to be identified as coming from a combination of monosaccharide residues. The software can be configured to assign composition values to the spectrum and display a table of the m/z values assigned. Compositional assignments can be printed, along with the associated spectrum, and copied to the clipboard.

To access the CarboTools program select **Biolynx - CarboTools** from the MassLynx Shortcut Bar. This displays the last CarboTools document viewed.

The main window contains a menu, optional toolbars and a status bar. It is split into two panes, the upper pane contains a list of results of compositional assignments and the lower pane contains the spectrum.

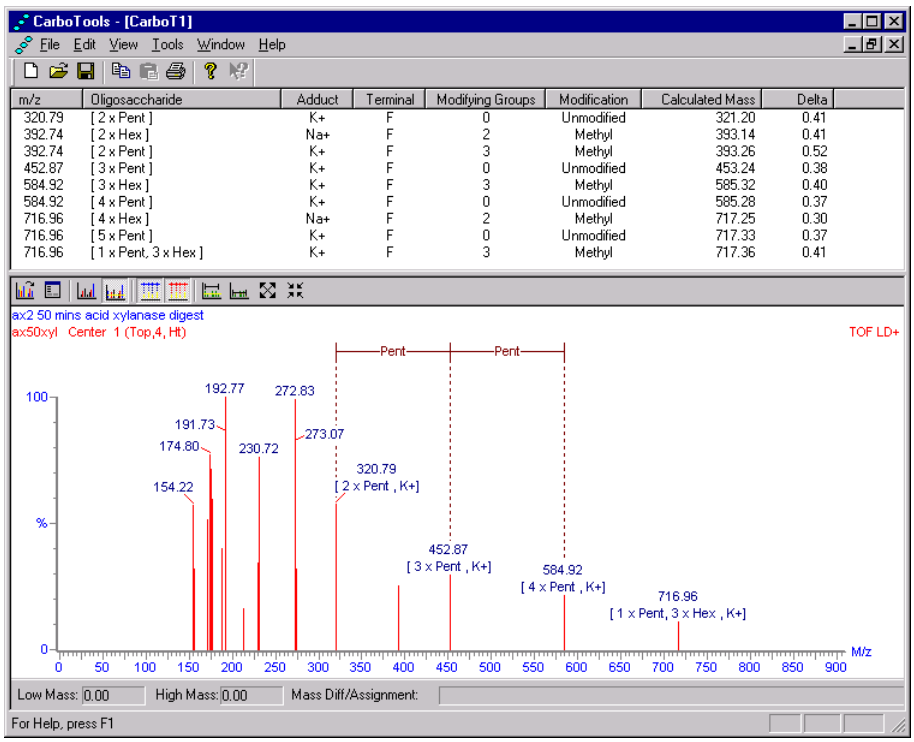


Figure 4.1 Carbohydrate Main Window

Toolbars

The CarboTools program has two toolbars that provide shortcuts for commonly used menu commands. To display the main toolbar select **Toolbar** from the **View** menu. The Spectrum toolbar is always displayed.

Main Toolbar

Toolbar button	Menu equivalent	Purpose
	File… New	Create a new Carbohydrate analysis file.
	File… Open	Open an existing Carbohydrate analysis file.

	File… Save or File… Save As	Save the current Carbohydrate analysis file.
	Edit…Copy	Copy the assignments to the clipboard.
	File… Print	Print the current Carbohydrate analysis file.
	Help… About Carbo Tools	Display CarboTools software version information.
	Help	Display MassLynx help for the CarboTools software.

Spectrum Toolbar

Toolbar button	Purpose
	Press the **Load Spectrum** button to invoke the Spectrum Data Browser dialog. This allows a spectrum to be loaded into the CarboTools document. Continuum data must be pre-processed to produce a centred spectrum before loading
	Press the **Display Settings** button to display the Spectrum Settings dialog, used for setting threshold and other spectrum display parameters.
	Press the **Display All Peaks** button to show every peak in the spectrum, regardless of intensity.
	Press the **Display Peaks Above Threshold** button to show only those peaks with intensity above the defined threshold.
	Press the **Ion Series1** button to display the first user defined ion series.
	Press the **Ion Series 2** button to display the second user defined ion series.
	Press the **Snap Current Threshold** button to enable cursor location on peaks above the current display threshold.
	Press the **Snap Start** button to enable cursor location on peaks with an intensity equal to or greater than the selected peak.
	Press the **Zoom Out** button to restore the spectrum display to full m/z range.
	Press the **Zoom In** button to decrease the displayed range by half, around the center of the current range. For example, if the current range is m/z 100 to 500 (a 400 Da range, centered at 300 Da) the new range will be 200 to 400 (a 200 Da range, centered at 300 Da).

Menu Commands

The toolbar provides shortcuts for some of the menu items, and these have been described above. There are also windows standard menu items that have no toolbar shortcuts. On the **File** menu

Print Setup, **Print Preview**, **Most Recent Files** and **Exit** and on the **Window** menu **New Window, Cascade, Tile, Arrange Icons** and **Split**.

Two items on the **Tools** menu do not have toolbar buttons **Configuration** and **Carbo Composer**, these are described in detail later in this chapter.

Creating a New Document

To create a new CarboTools document, select **New** from the **File** menu or press the toolbar button.

Configuration

The configuration dialog allows the monosaccharide groups to search for to be defined, along with any derivatives modifications and adducts to consider. To access select **Tools, Configuration.**

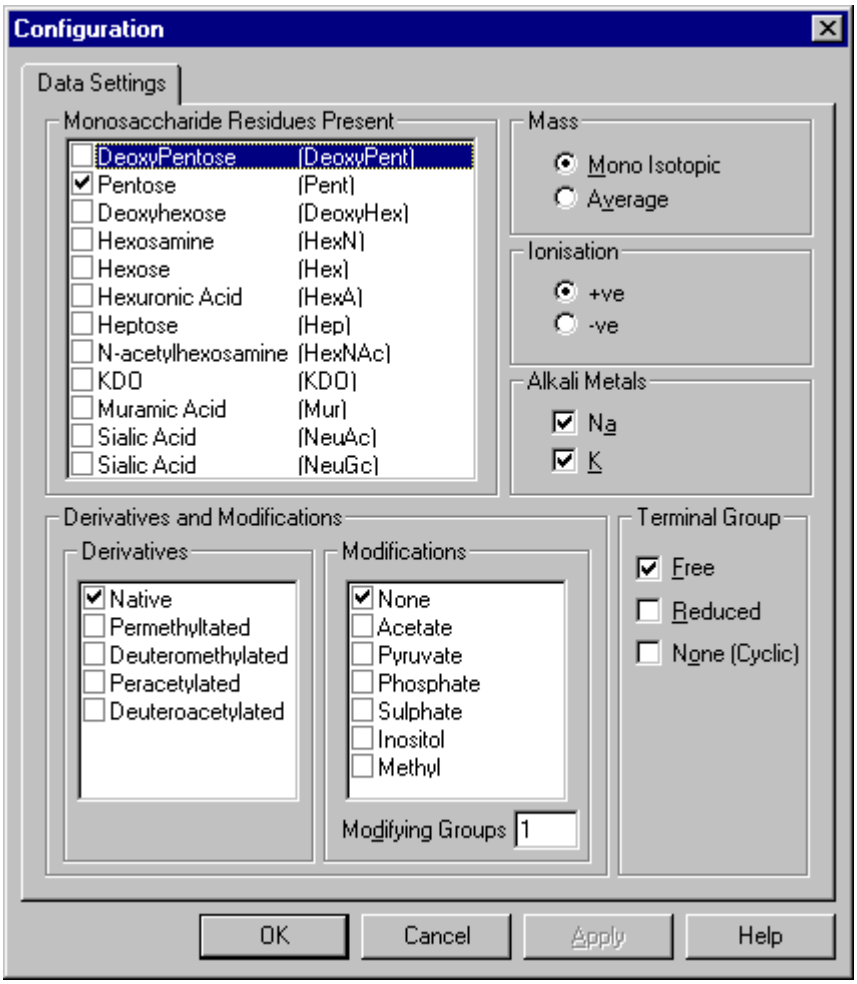


Figure 4.2 Configuration dialog

Monosaccharide Residues Present A list of residues is displayed showing both a long name and a shorter name which will be displayed on the Spectrum. Check the box(es) for the residues that may be present in the sample. **Note:** The more residues selected the more likely multiple assignments will be made to a peak.

Derivatives	Check the box for the derivative to consider. **Note:** Only one derivative may be selected.
Modifications	Check the box(es) for the modifications to be considered. When making assignments, the mass of an oligosaccharide is adjusted to account for the modifying groups selected. **Note:** This can be very expensive computationally so the type and number of groups should be limited.
Modifying Groups	Enter the maximum number of modifying groups to consider. When making assignments, the mass of an oligosaccharide is adjusted (up to the number entered) to account for the number of modifying groups selected. **Note:** This can be very expensive computationally so the type and number of groups should be limited.
Mass	Select whether to use monoisotopic or average mass values. Normally monoisotopic values will be used.
Ionization	Select the ionization mode appropriate to the conditions under which the data was acquired.
Alkali Metals	Sodium and Potassium adducts can be selected, check the boxes as required. If neither is selected then only the H+ adduct is considered. If negative ionization is selected then these adducts will not be used in the calculation of the oligosaccharide mass.
Terminal Group	Check the box(es) for the terminal group according to the arrangement of the reducing end. More than one arrangement can be selected. None means there is no reducing end or in other words a cyclic arrangement.

Loading a Spectrum

Load an appropriate spectrum by pressing the toolbar button. The Spectrum Data Browser will be displayed. Locate the required file and select a centroided spectrum. You may need to press the History button and select a centered (or MaxEnt) spectrum from the Process History list.

Display Settings

Press the toolbar button to display the Spectrum Settings dialog.

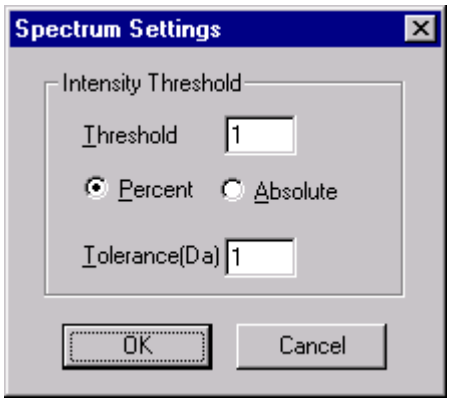


Figure 4.3 Spectrum Settings dialog

Threshold Select **Percent** or **Absolute** and enter the value above which peaks must be to be displayed on the spectrum when the **Display Peaks Above Threshold** button is pressed ( ).

Tolerance Enter a tolerance value. A peak must be at the defined mass +/- the tolerance to be assigned a composition.

Calculating Compositions

All the peaks in the loaded spectrum can be give a compositional assignment. The program takes each peak mass in the spectrum and calculates the mass of all possible combinations of monosaccharides selected. Each combination where the mass is equal, to within the set tolerance of the current peak, is added to the list of possible assignments. Selecting more residues (and also modifiers) will lengthen the calculation and produce more matches.

To calculate compositions select **Tools, Carbo Composer** to display the Carbo Composer dialog.

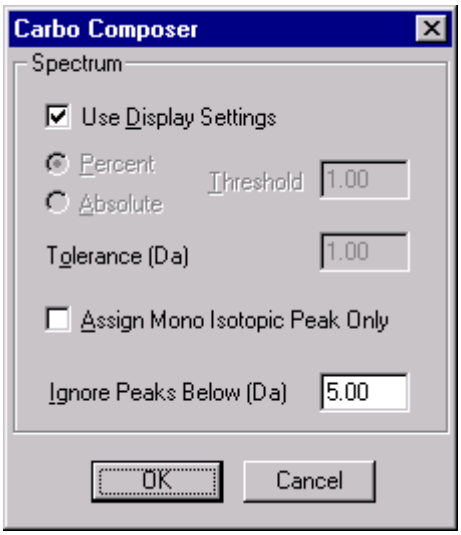


Figure 4.4 Carbo Composer dialog

Use Display Settings Check this box to use the Threshold and Tolerance values defined in the Spectrum Settings dialog. See Display Settings on page 4-6. To use different Threshold and Tolerance values uncheck this box and enter new values.

Assign Mono Isotopic Peak Only Check this box to assign only the main peak of an isotope cluster. Charge state deconvolution is carried out so that multiply charged peaks can also be processed. If a spectrum has already been processed by MaxEnt 3 then this option is ignored.

Ignore Peaks Below Enter a value that a peak must be above to be assigned a composition.

Press the **OK** button to begin the calculation. The software will scan the spectrum and evaluate carbohydrate compositions for the peaks in the spectrum. **Note:** Calculation will only be performed if the Configuration dialog has been setup and a spectrum has been loaded. Results will be displayed in the Results List pane.

During calculation a progress bar is displayed at the bottom of the Results List pane. To stop calculation press the abort button on the progress bar.

User Defined Modifiers and Terminal Groups

User Defined monsaccharide masses can be created and saved to a new file in the User Defined Modifiers and Terminal Groups dialog. This is invoked by selecting **Tools, Edit Mods and Terms**.

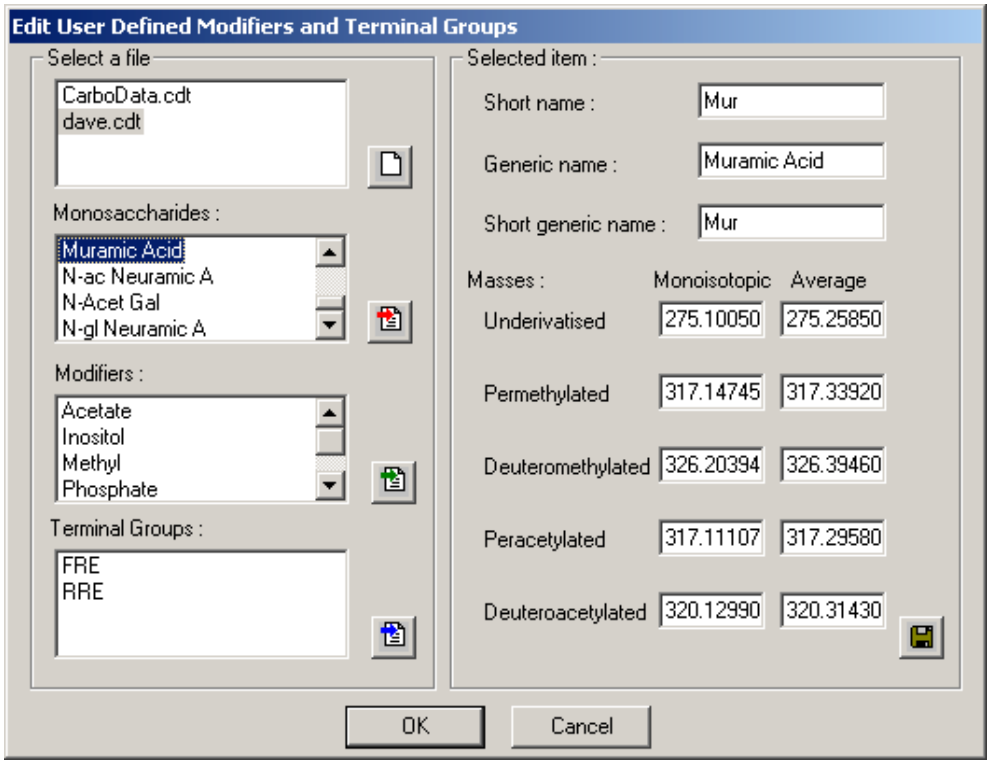


Figure 4.5 The Edit User Defined Modifiers and Terminal Groups dialog

Before any new masses can be added a new ***.cdt** file has to be created. To create a new file press the button and add the name of the new file when prompted.

The default **CarboData.cdt** file cannot be edited and a warning dialog is invoked if this attempted

On selecting an item, its name and properties are displayed on the right hand side of the dialog. These can then be edited and any changes saved by clicking on the button.

Add a New Monosaccharide

Press the button and add the new name to the New Monsaccharide edit box (Figure 4.6).

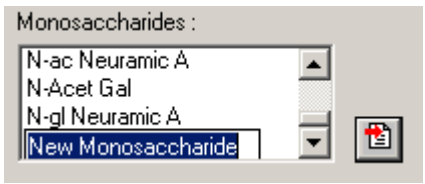


Figure 4.6 Add New Monsaccharide Edit box

Its properties can now be entered in the **Selected Item** box on the right of the dialog.

Add a New Modifier

A new modifier can be added in a similar way to Add a New Monosaccharide. Press the button to invoke the New Modifier edit box.

Add a New Terminal Group

A new modifier can be added in a similar way to Add a New Monosaccharide. Press the button to invoke the New Terminal Group edit box.

Manually Identifying Residues

Residues can be identified manually. Hold down the shift key and click on a peak and drag along the spectrum. The mass of the first peak will be displayed in the **Low Mass** field, below the spectrum, the **High Mass** field will display the mass at the current cursor position and the difference will be displayed in the **Mass Diff/Assignment** field. If the mass difference corresponds to one of the monosaccharide residues selected in the Configuration dialog the mass difference will be followed by the monosaccharide short name (e.g. 132.04 Pent). **Note:** The mass difference must be within the tolerance window for the description to be displayed.

Snap to Current Threshold

Press the Snap to Current Threshold button ( ) to consider only peaks above the Threshold value defined on the Spectrum Settings or Carbo Composer dialog. When manually identifying residues a value is only displayed in the Mass Diff/Assignment field when a peak has been located and the peak intensity is above the defined threshold value.

Snap to Start

Press the Snap to Start button ( ) to consider only peaks with an intensity equal to or above the intensity of the Low Mass peak. When manually identifying residues a value is only displayed in the Mass Diff/Assignment field when a peak has been located and the peak intensity is equal to or above the Low Mass peak intensity.

Series Assignments

Up to two ion series can be displayed on a spectrum (usually corresponding to a sodium adduct and a potassium adduct series).

1. Press the (Ion Series 1) and (Ion Series 2) buttons on the spectrum toolbar, to display the first and second ions series.

2. Click on a peak and hold down the mouse button and the **Shift** key. Drag the mouse to the second peak and then release the mouse button. If this distance between the peaks corresponds to a monosaccharide residue then this difference will be labelled as in Figure 4.7.

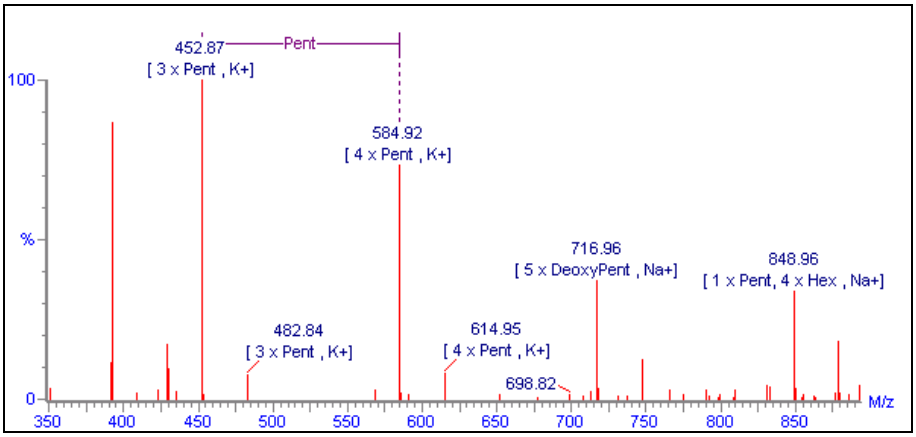


Figure 4.7 Labelled Ion Series

To build a second series repeat the above procedure press both the **Ctrl** and **Shift** keys.

To hide the ion series assignments press the and buttons a second time.

To remove a series assignment label click, with the right mouse button, on the label. A popup menu will be displayed, select Delete Assignment to delete the current assignment or Delete All Assignments to delete all assignments.

Configuring the Results Display

To define which columns are displayed in the Results List pane click, with the right mouse button, on the pane and select **Setup Columns** from the popup menu displayed. The **Display Columns** dialog is displayed.

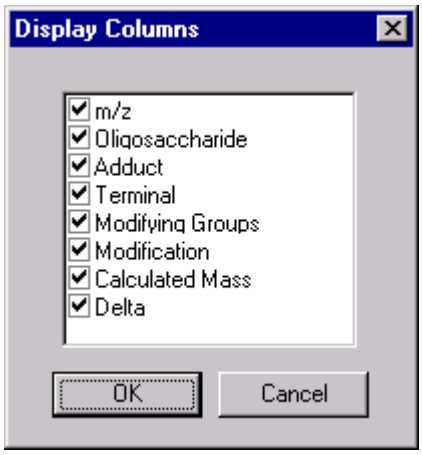


Figure 4.8 Display Columns dialog

Check the boxes for the columns to display, and uncheck the boxes for the columns to hide.

Deleting Results

Compositions can be deleted from the Results List pane and the Spectrum pane. To delete a composition from the Results List, press the **Delete** key. To reflect the change in the Spectrum pane click, with the right mouse button, on the Results List pane and select **Update Spectrum** from the popup menu displayed.

This may be useful for peaks that have multiple assignments; those that are less likely can be deleted.

Saving/Loading Parameters and Results

The configuration and results of an analysis can be saved to and loaded from a file. This includes any series assignments that have been made.

Parameters and results can be saved by selecting **File, Save** or **SaveAs** or press the toolbar button. Enter a name in the File Save As dialog and press the **Save** button.

To load a parameter file, select **File, Open** or press the toolbar button. Select the required *.cbt file from the Open dialog box and press the **Open** button.

Chapter 5 BioLynx - Nucleic Acids

Contents

Contents 5-2
Figures 5-2
Overview 5-2
The Sequence Window 5-3
    Display Options 5-3
The Nucleic Acids Toolbar 5-3
The Sequence Editor 5-4
    Editing Text 5-6
    Terminal/Linkage Group Modifications 5-7
    Modified Bases 5-9
    To Exit the sequence editor 5-11
Miscellaneous Editing 5-11
    Sequence Name and Source 5-11
    Clipboard 5-11
Calculations 5-12
    Base Composition 5-12
Printer Support 5-13
References 5-14

Figures

Figure 5.2 Sequence window showing the nucleotide base sequence of a tRNA molecule 5-3
Figure 5.3 Sequence Editor showing text editor, masses and composition 5-5
Figure 5.4 Elemental Composition dialog 5-5
Figure 5.5 Find and Replace dialog editor 5-7
Figure 5.6 RNA molecule showing bases held together by phosphodiester bonds between ribosyl moieties attached to the nucleobases by N-glycosidic bonds 5-7
Figure 5.7 Terminal group editor dialog 5-8
Figure 5.8 Create new terminal/linkage group dialog 5-8
Figure 5.9 Modification dialog editor 5-9
Figure 5.10 Create new modified base dialog editor 5-9
Figure 5.11 Sequence name and Database source information 5-11
Figure 5.12 Result of base composition calculation 5-12
Figure 5.13 Print report dialog box used for obtaining hard copies of various analyses 5-13

Overview

The Nucleic Acids program is a biological software toolset used to analyse and manipulate nucleotide sequences being either DNA or RNA so as to facilitate the interpretation of data obtained from mass spectrometry. The program currently only supports mass calculations, modifications and base composition calculations.

The Sequence Window

Selecting **BioLynx - Nucleic Acid Editor** from the MassLynx Shortcut menu, brings up a sequence window displaying an empty chain with 5' OH and 3' OH groups. The sequence window is one of several windows within BioLynx which use a Multiple Document Interface (MDI) display which allows multiple windows (called documents) to be displayed simultaneously. The sequence window also features a toolbar located at the top of the window for easy access to common commands and a status bar to display useful information.

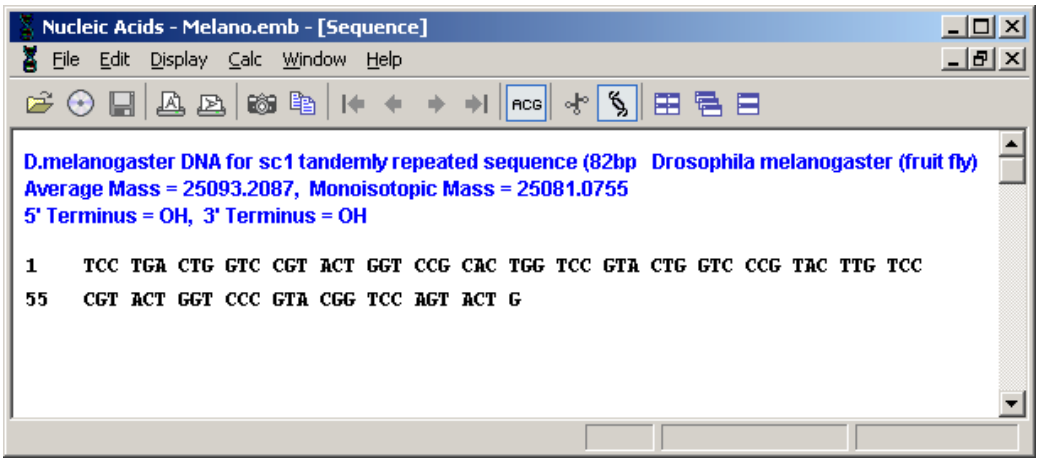


Figure 5.1 Sequence window showing the nucleotide base sequence of a tRNA molecule

Display Options

The font type (fixed or proportional) and size can be customized for display purposes. This would be useful when wanting to capture for e.g. Figure 5.1and paste the picture into a word processing package. The use of a proportional font allows the picture to be suitably scaled.

To customize the font type and size, select **Tools, Colors and Fonts** from the MassLynx top level menu.

The color of the **header text** in Figure 5.1is derived from the color of the **header text** in the fonts and color dialog box. All other text on display i.e. mass lists etc. are derived from the **list text** in the fonts and colors dialog box.

The Nucleic Acids Toolbar

The Toolbar is displayed at the top of the window and allows you to perform some common operations with a single click of the appropriate Toolbar button.

Press to open a nucleotide file.

Press to open an EMBL file.

Press to save file in nucleotide format.

Press to print report of selected analysis in portrait format.

Press to print report of selected analysis in landscape format.

Press to send a picture of current window to the clipboard.

Press to copy contents of the current window to the clipboard.

Press to display the first chain of a set of chains or press the **HOME** key on the keyboard.

Press to display the previous chain of a set of chains or press the ← key on the keyboard.

Press display the next chain of a set of chains or press the → key on the keyboard.

Press to display the last chain of a set of chains or press the **END** key on the keyboard.

Press to toggle between displaying bases in blocks of 3 or 10 bases.

The display of bases in blocks of 3 or 10 and the RNA/DNA display can also be accessed from the Nucleic acids **Display** menu.

Press to display sequence as RNA.

Press to display sequence as DNA.

Press to arrange the windows in a tiled view.

Press to arrange the windows in a cascaded view

Press to arrange the windows in a stacked view.

The Sequence Editor

The sequence editor allows basic text editing of the nucleotide sequence with informative displays.

To Access the Sequence Editor

Select **Sequence** from the **Edit** menu

-or-

Double click on the Sequence Display window.

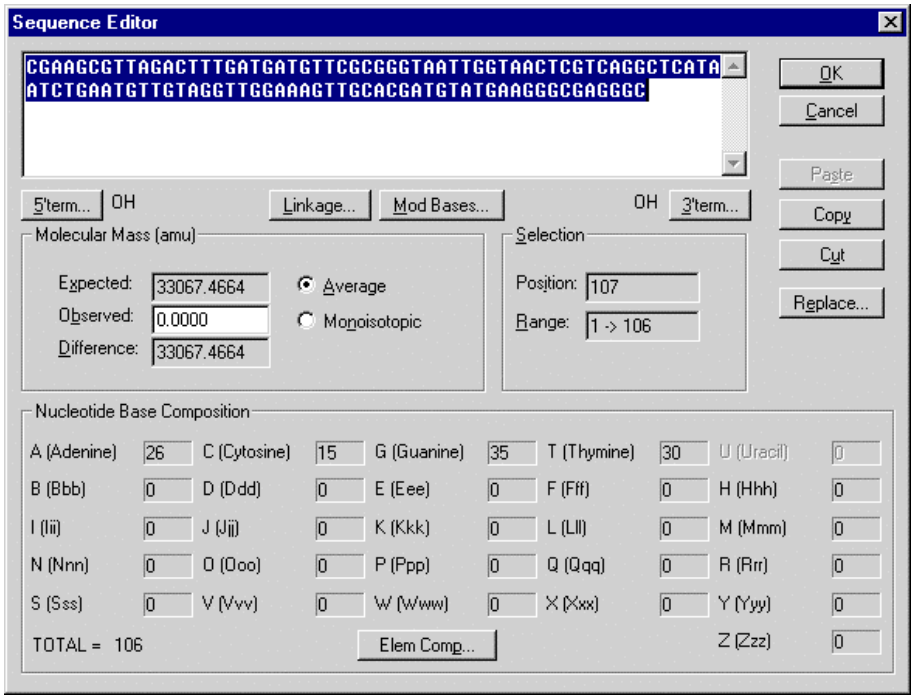


Figure 5.2 Sequence Editor showing text editor, masses and composition

Molecular mass The displayed **Expected** mass is the expected mass for the portion of text highlighted. An observed mass can be entered, using the keyboard, in the **Observed** edit control. The **Difference** mass reflects the difference between the observed and expected masses.

Selection The position of the cursor in the text edit window is indicated in the **Position** control. If there is no text in the text editor this position will be 2. If text is selected i.e., highlighted, the range of the selection will be indicated in the **Range** control

Nucleotide Base Composition Shows the number of each base present in the current sequence, including modified bases.

Elem Comp Displays of the elemental composition of the sequence. Ten of the most common elements are displayed showing the number of carbons, hydrogens etc. These calculations are based on zero charge species.

The displayed formula can be copied into the **Isotope Model** program in **Spectrum** by highlighting the formula and selecting **Copy**.

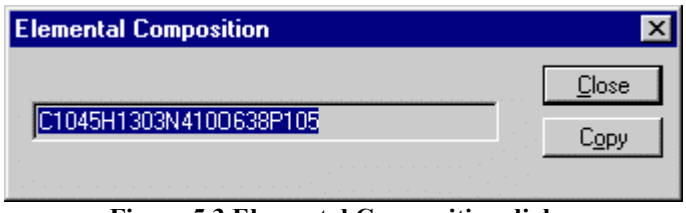


Figure 5.3 Elemental Composition dialog

The **Cut, Copy, Paste and Replace** buttons are used for editing text and are described in the next section.

The **5'Term, 3'Term, Linkages** and **Mod Bases** buttons load the 5' terminal, 3' terminal, linkages and modification group editors.

Editing Text

Editing text in the sequence editor involves the same keyboard and mouse input as in most word processing packages. The focus is set to the edit window and all residues in the chain are highlighted. If text is accidentally deleted, press cancel and re-enter the sequence editor dialog box.

Keyboard Operation

- Text entry using the single-letter nucleotide base codes.
- **BACKSPACE** deletes the character to the left of the cursor.
- **DELETE** deletes the character to the right of the cursor.
- **CURSOR** keys move the cursor to the left or right of text.
- Simultaneous **SHIFT** and **CURSOR** key operation allows text to be highlighted.

Mouse Operation

- New nucleotide bases can be added or inserted into the current sequence by clicking on the relevant base in the **Nucleotide Base Composition** part of the Sequence Editor window.
- A click in the text editor positions the cursor.
- Simultaneous click and drag highlights a block of text.
- Double click highlights all text.

  Only alphabetical text can be entered in the edit window and the unassigned letters are represented as zero mass until assigned a user defined base. All displays are constantly updated to reflect the current status of input - average or monoisotopic masses (zero charge); the position of the cursor in the sequence; and the nucleotide base composition. When bases are highlighted the range is displayed as well as the appropriate mass and linkage groups, or in the case of 5' or 3' termini bases the appropriate terminal group.

Clipboard Operation

Text can be cut, copied and pasted to and from the clipboard using the **Cut**, **Copy** and **Paste** buttons in the chain editor.

- **CUT** removes the highlighted text from the editor and inserts it into the clipboard.
- **COPY** copies the highlighted text to the clipboard without removing it from the editor.

Both these commands are enabled when text in the editor is highlighted.

- **PASTE** copies text from the clipboard into the editor inserting the text at the position of the cursor. Highlighted text is replaced with the inserted text.

This command is enabled when text has been copied to the clipboard.

Find and Replace operation

Bases can easily be replaced by modified bases or highlighted to show their occurrence in a particular sequence. To do click **Replace** to invoke Figure 5.4.

- **Find Next** highlights the next occurring specified base.
- **Replace** replaces the specified base with the new base.
- **Replace All** replaces all occurrences of a specified base with the new base.

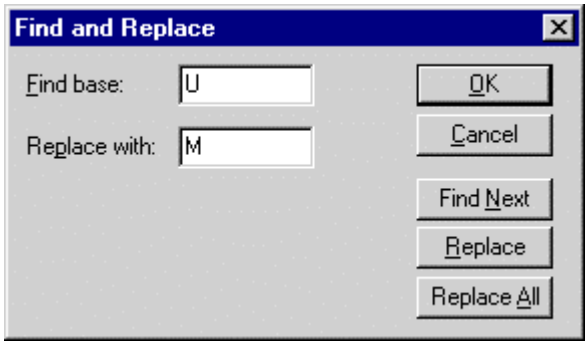


Figure 5.4 Find and Replace dialog editor

The 5', 3' terminal and linkage groups can be modified by editing and creating new groups. The default termini groups are OH and OH respectively. The default linkage group is phosphate.

Terminal/Linkage Group Modifications

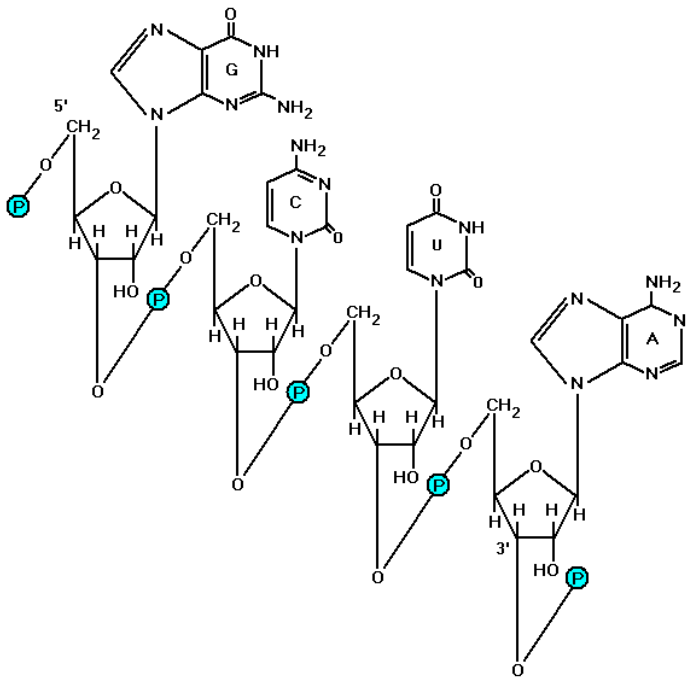


Figure 5.5 RNA molecule showing bases held together by phosphodiester bonds between ribosyl moieties attached to the nucleobases by N-glycosidic bonds

To modify either the 5' or 3' terminus press the **5' term** or **3'term** button in the sequence editor. To modify the linkage between bases press the **Linkage** button.

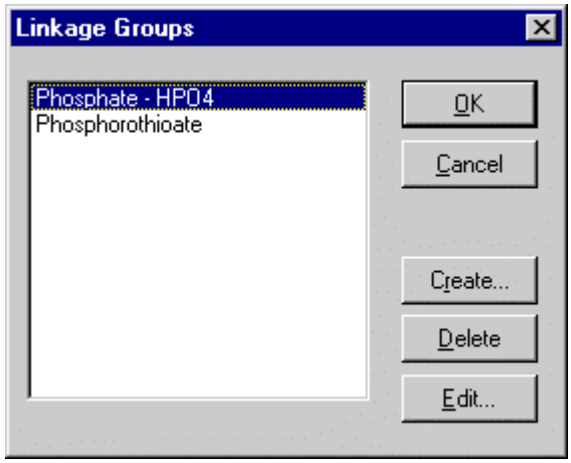


Figure 5.6 Terminal group editor dialog

To Select a Different Terminal/Linkage Group

To select a new or edited group into the sequence editor, click on the group of interest in the Linkage Groups list box (Figure 5.6), and press **OK**. The chain editor mass displays are automatically updated.

To Create a New Terminal/Linkage Group

1. Press the **Create** button this invokes Figure 5.7.
2. Enter a unique name for the group (maximum 50 characters).
3. Enter a unique symbol e.g. OH for Free Acid (maximum 50 characters).
4. Enter a formula using IUPAC nomenclature e.g. Cl for Chlorine.
5. Press **OK** to enter group in database. The new group is displayed in the list box.

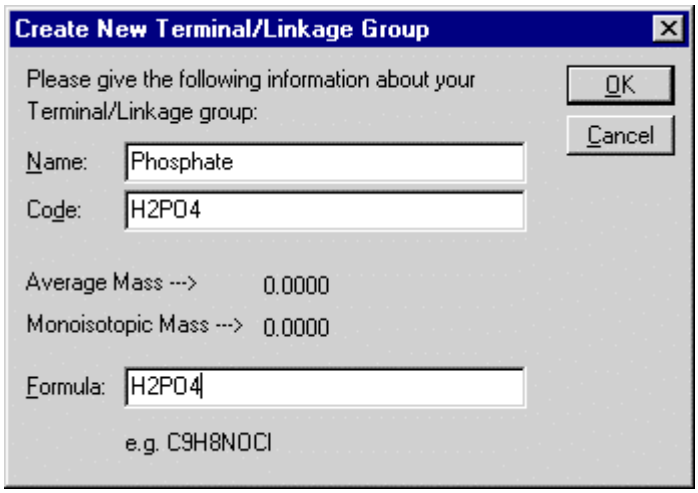


Figure 5.7 Create new terminal/linkage group dialog

To Edit a Terminal/Linkage Group

1. Click on the linkage in the list box (Figure 5.6).
2. Press the **Edit** button.
3. Change parameters and press **OK** to enter group in database, or step 4.

4. View parameters and if no changes press **Cancel**.

To Delete a Terminal/Linkage Group

1. Click on the linkage in the list box (Figure 5.6).
2. Press the **Delete** button.
3. You will be prompted to confirm deletion, press **Yes**.

Modified Bases

To modify a base press the **Mod Bases** button in the sequence editor (Figure 5.2).

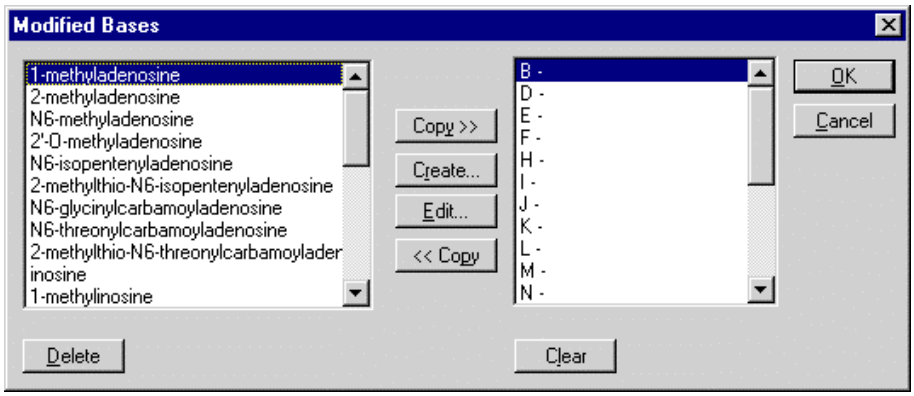


Figure 5.8 Modification dialog editor

Modified bases are listed in the list box (Figure 5.8) with the input focus on the currently selected modified base.

To Create a Modified Base

1. Press the **Create** button this invokes Figure 5.9.
2. Enter a unique name for the group (maximum 50 characters).

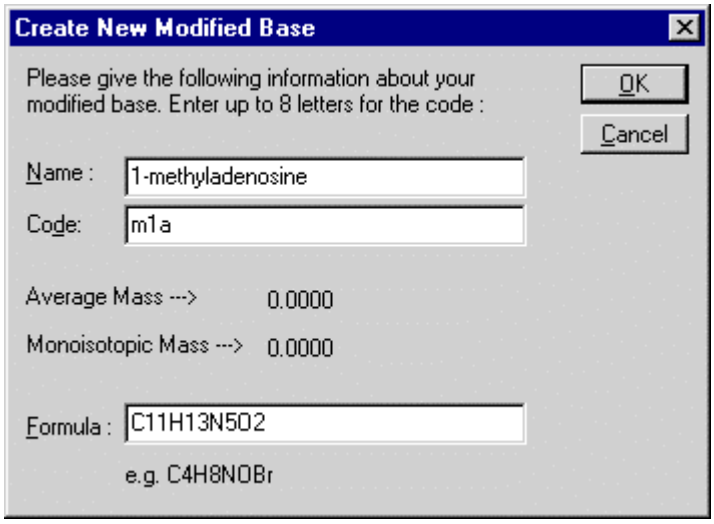


Figure 5.9 Create new modified base dialog editor

3. Enter a unique **three** letter symbol.

4. Enter a formula using IUPAC nomenclature.

5. Press **OK**, the new modified base will appear in the modified bases list box.

To Edit a Modified Base

1. Click on the modified base in the list box.

2. Press the **Edit** button.

3. Change parameters and press **OK**.

To Delete a Modified Base

1. Click on the modified base in the list box.

2. Press the **Delete** button.

3. You will be prompted to confirm deletion, press **Yes**.

To Copy a Modified Base to Modified Base List

1. Click on the required modified base in the list box.

2. Click on the required unused letter in the list box.

3. Press **Copy >>**.

4. Press **OK** to make the changes permanent.

To Copy a Modified Base from Modified Base List to Database

This allows modified nucleotide bases created on an alternative data system to be copied into the database.

1. Copy the required file into the Nucdata folder using NT Explorer.

2. Choose **Open** from the Nucleic Acids Sequence **File** menu, or press the toolbar button, and select the required file. The filename will appear next to the first unused letter.

3. Select the modified base from modified base list.

4. Press **Copy <<** and the modified base will appear in the base list box.

5. Press **OK** to confirm the copy.

To Clear a Modified Base/s

This the clears the base associated with a letter.

1. Click on the required modified base in the list box.

2. Press **Clear**.

3. Press **OK** to enter the modified groups into the database.

4. The chain editor mass displays are automatically updated.

To Exit the sequence editor

Press **OK** to display the edited chain in the Sequence Display window.

-or-

Press **Cancel** to disregard any changes made in the sequence editor.

Miscellaneous Editing

Sequence Name and Source

The name of the current sequence can be edited by choosing **Edit, Sequence Name** and is used for display purposes in all windows. The maximum allowed length for a name is 100 characters and, although it can be the same, should not be confused with the **File, Save As** name given to the sequence for disk storage. Files imported from the EMBL will have information displayed as in Figure 5.10. Any modifications to sequences can be entered in the **Sequence Modifications** edit control. This is displayed on hard copy formats.

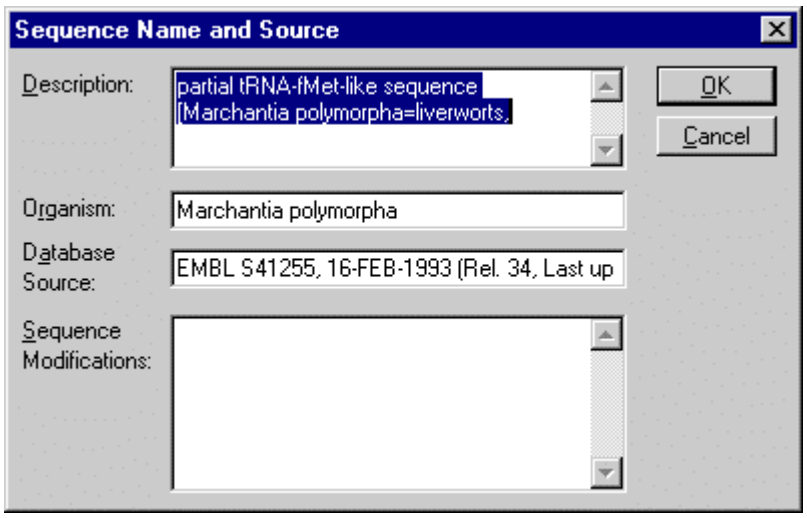


Figure 5.10 Sequence name and Database source information

Clipboard

A picture of the currently active window can be copied to the clipboard by choosing **Edit, Copy** from the menu or pressing the toolbar button.

Edit, Copy List or the toolbar button copies a list of the contents of the currently active window to the clipboard i.e. sequences can be copied to the clipboard.

Calculations

Base Composition

Determining oligonucleotide composition from molecular mass is an extremely useful tool, especially when masses are obtained by electrospray with errors less than 1 Da. Combined with selective cleavage and hence constraining the identity of one base allows the unambiguous assignment of composition up to at least 14 mers (1,2).

To carry out the base mass combination calculation choose **Calc, Base Composition** from the BioLynx menu. For a given mass this routine displays all possible combinations of the common bases with user selected termini. The maximum number of bases is 100 so the maximum mass is approximately 30000.

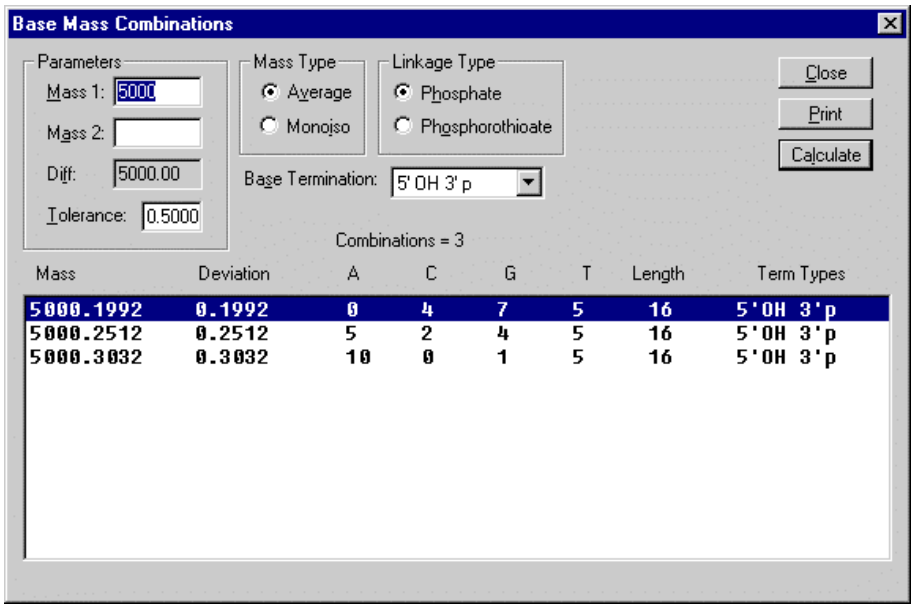


Figure 5.11 Result of base composition calculation

Base Combination Parameters

1. Parameters to be entered are a mass (maximum 30000), a tolerance (default 0.5) and either average or monoisotopic mass.

2. Select the type of **Base Termination**, **OH** and **OH**, **OH** and **P**, **cyclic** or **Consider All**.

3. Specify the **Linkage Type** as phosphate or phosphorothioate.

To Display the Combinations

There are two methods of displaying the combinations. You can either press the enter key when the cursor is in the mass 1 or mass 2 edit control or click on the **Calculate** button.

A maximum of 100 combinations are sorted in the list with the smallest deviation at the top of the list.

Printer Support

Any displayed information can be translated into hard copy form. Hard copies can be obtained using any of the methods below.

Printing a Window or Contents of a Window

Use the keyboard **Alt-Prn Scrn** or **Prn Scrn** to copy a window to **Windows Clipboard.**

-or-

Print a **Report** by choosing **Print Report** from the **File** menu or pressing one of the toolbar buttons.

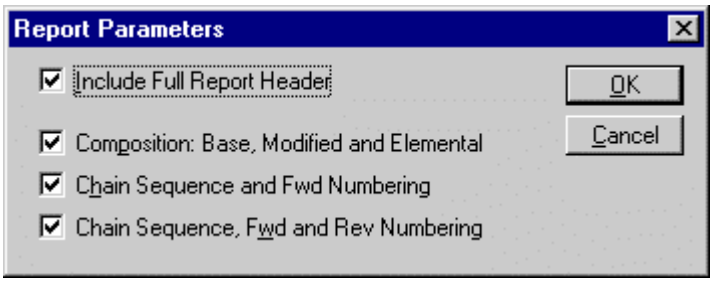


Figure 5.12 Print report dialog box used for obtaining hard copies of various analyses

Printing a Report

1. Click the required check box parameters. Grayed out options indicate that a particular analysis has not been carried out.

2. Press **OK**.

Include Full Report Header	Check this box to attach a full report header to the printout. An alternative header, taking up fewer lines than the full report header, will be printed if this check box is unchecked.
Composition: Base, Modified and Elemental	Check this box to print out the current chains contents i.e., number and composition of bases, elemental composition and also any modified bases.
Chain Sequence and Fwd/Rev Numbering	Check this box to print out the sequence with appropriate numbering.

References

1. Kowalak, J.A., Pomerantz, S.C., Crain, P.F. and McCloskey, J.A. (1993) *Nucleic Acids Res.,* **21**, 4577-4585.

2. Pomerantz, S.C., Kowalak, J.A. and McCloskey, J.A. (1993) *J. Am. Soc.Mass Spectrom.,***4**, 204-209.

Chapter 6 BioLynx - Oligonucleotide Sequencing

Contents

Overview 6-3
Oligonucleotide Fragmentation 6-3
The Sequencing Algorithm 6-3
Using the Oligonucleotide Sequencing Program 6-5
    Accessing the Oligonucleotide Sequencer 6-5
    Creating a New Document 6-5
    Specifying the Spectrum 6-5
    Specifying Sequencing Parameters 6-6
    Analyzing the Spectrum 6-9
    Troubleshooting 6-9
The Oligonucleotide Sequencing Interface 6-11
    Oligonucleotide Sequencing Documents 6-11
    The Main Window 6-11
    Child Windows 6-11
    Toolbar 6-13
    Menu Commands 6-19
References 6-24

Figures

Figure 6.2 Oligonucleotide product ion nomenclature 6-3
Figure 6.3 Peak List Format 6-5
Figure 6.4 The Sequence Parameters Page 6-6
Figure 6.5 The Oligonucleotide Parameters Page 6-7
Figure 6.6 The Series Ion Parameters Page 6-8
Figure 6.7 The Spectrum Parameters Page 6-9
Figure 6.8 Oligonucleotide Sequencing Program Main Window 6-11
Figure 6.9 The Consensus Sequence Pane 6-12
Figure 6.10 The Spectrum Pane 6-12
Figure 6.11 The Sequence Report Pane, showing a Sequence Prediction 6-13
Figure 6.12 The Standard Toolbar 6-14
Figure 6.13 The Sequence Toolbar 6-15
Figure 6.14 The Spectrum Manipulation Toolbar 6-15
Figure 6.15 The Text Report Toolbar 6-16
Figure 6.16 The Sequence Prediction Details Report 6-16
Figure 6.17 The Analysis Details Report 6-18
Figure 6.18 The Series Ion Assignment Report 6-19
Figure 6.19 Report Options, Reports page 6-20
Figure 6.20 Report Options, Spectrum Plot page 6-21
Figure 6.21 The Font Sizes dialog 6-22
Figure 6.22 The View Toolbars dialog 6-22
Figure 6.23 The Spectrum Magnify dialog 6-23
Figure 6.24 Oligonucleotide Sequencing document icon 6-23

Overview

The Oligonucleotide Sequencing program assists in deducing the residue sequence of a DNA- or RNA-based oligonucleotide from its multiply charged negative ion, centered daughter spectrum. The program, guided by user-supplied parameters, builds a set of consensus sequences whose series ions best match the experimental spectrum. The results of the analysis are displayed in both graphical form as an annotated spectrum and in textual form as reports detailing the series ions searched for and actually matched to the spectrum, and spectral assignments based on the consensus sequences.

The user supplies a number of parameters that serve to guide the interpretation of the spectrum, including the spectral data, known oligonucleotide information, and ion matching tolerances. Spectral data may be input directly from a MassLynx data file or from a listing of masses and intensities.

Oligonucleotide Fragmentation

Oligonucleotides, like peptides, exhibit characteristic daughter ion spectra due to fragmentation along the phosphate backbone with charge retention both to the 3' and 5' termini (1-3). These series are labeled similarly to those of peptides, the labels **a, b, c,** and **d** used for ions resulting from fragmentation with charge retained toward the 5' end, while the labels **w, x, y,** and **z** are used where charge is retained toward the 3' end. This nomenclature is illustrated schematically in Figure 6.1 below:

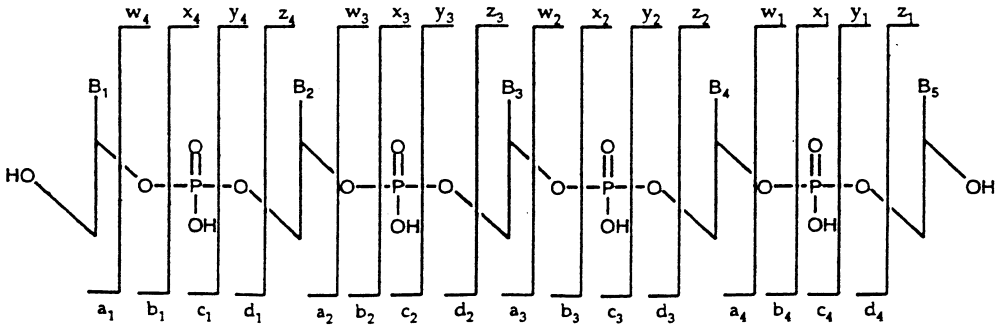


Figure 6.1 Oligonucleotide product ion nomenclature

Product ions for any of the possible series may be observed in a spectrum. Two series, **w** and **a-B** (**a** series fragment, with loss of the base) can provide unambiguous sequence information from the 3' and 5' ends, respectively (1). The first ion in the **w** series is the singly-charged $\mathbf{w_1}$, followed by $\mathbf{w_2}$, which may be present in both doubly and singly charged forms, and so forth, each successive series ion existing in up to as many charge states as $z \leq$ the charge of the precursor ion. That is, if the parent ion is triply charged (3-), then no series ion may be more than triply charged. The first **a-B** series ion is $\mathbf{a_2\text{-}B_2}$, corresponding to the first 5' nucleotide since the base $\mathbf{B_2}$ has been lost. The same charge state restrictions apply as for **w** series ions.

The Sequencing Algorithm

The Oligonucleotide Sequencing program finds sequences that are a consensus of sequences derived from the w and a-B series ions. The algorithm works by extending the sequences from both the 5' (**a-B** ions) and 3' (**w**) ends. The first residue is determined by finding the best match between the expected ion for each of the four possible residues (A, C, G or T in the case of DNA, or A, C, G, or U for RNA) and the spectral data. When more than one matching ion is found, the matches are ranked in order of decreasing intensity. The next residue is determined by again

adding each of the possible residues to the sequence and searching the spectrum for the expected ion(s).

Because more than one possible sequence might be consistent with the spectral data and molecular mass (4), sequences are ranked in descending order according to the number of charge states matched and the total intensity of the ions matched. For example, if two residues are found to be consistent for $\mathbf{w_6}$, the one which matches the most expected charge states (*e.g.* both $\mathbf{w_6^{3-}}$ and $\mathbf{w_6^{4-}}$, *vs.* only $\mathbf{w_6^{2-}}$), or for equal number of charge state matches, the greater intensity is chosen as the preferred residue to extend the sequence. At any time during the derivation of the series ion sequences, a set of the topmost ranked sequences is retained as candidates for the next sequence extension step. The size of this set (which is a user-configurable parameter) determines in part the amount of time and memory required to search for consistent sequence matches, the larger the size of the set, the more possibilities must be considered and the more memory required to hold the results.

Once the two sets of highest ranked series ion sequences (**w** and **a-B**) have been derived, a set of consensus sequences is derived by overlapping each of the **w** (3' $\rightarrow$ 5') and **a-B** (5'$\rightarrow$ 3') series in pair-wise fashion. If the consensus sequence matches the molecular weight (within the specified MW tolerance), the sequence is added to the result set, otherwise it is discarded. (This is also configurable - see the section **Sequencing Parameters**, below).

The consensus sequences in the result set are ranked by a score computed using nine ion series: **a-B, b, b-H_2O, d, d-H_2O, w, w-H_2O, y,** and **y-H_2O.** For each nucleotide, the partial score is the sum of matches of these expected ions against the spectrum, and the sequence score is the sum of the partial scores. The final set of consensus sequences is ranked in descending order of the sequence score. The number of consensus sequences retained is determined by a user-configurable parameter.

The molecular weight of the oligonucleotide is used as a check on the final sequences; sequences with a composition inconsistent with the molecular weight are obviously incorrect. However, it is often useful to retain these incorrect sequences in the final set, because in many cases the full sequence cannot be determined from the series ions. Especially for longer oligonucleotides, ions for nucleotides toward the center of the sequence may not be observed, and thus the sequence cannot be fully extended. Matches of expected ions to minor spectral features for these internal residues will necessarily be simply coincidental. However, the algorithm will still select these as possible extensions, and will use them in deriving consensus sequences even though they cannot be correct.

It is possible to exercise considerable control over the algorithm by appropriate parameter settings. For example, setting an intensity threshold will force the algorithm to ignore any spectral features of lower intensity. In this way, coincidental matches to minor ions can be minimized. Other adjustable parameters include the match tolerance (between expected ions and spectral features), and the maximum charge state used.

The success of the algorithm at finding sequences is very dependent on the quality of the input spectrum. Weak, noisy spectra are likely to yield poor sequences, because of the high probability of coincidental matches, and because the intensity of ions deriving from internal residues is generally weaker than those from residues closer to the termini. Best success can be achieved with stronger, higher S/N spectra. In addition to spectral quality, the charge state of the precursor ion can influence the sequencing. Because it bases the ranking of candidates on both number of charge states and intensity matched, a higher charge state precursor may be better. The trade-off is that the total signal due to any particular residue is thus distributed over more charge state (more ions), resulting in higher competition from coincidental matches.

Using the Oligonucleotide Sequencing Program

Accessing the Oligonucleotide Sequencer

To Access the Oligonucleotide Sequencer, select **BioLynx - Oligonucleotide Sequencing** from the MassLynx Shortcut bar.

Creating a New Document

Select **New**, **File** menu or press the button on the Standard Toolbar. An empty Child Window will be created, named "Untitled1". There is no spectrum associated with this document, and all parameter settings are defaulted.

Specifying the Spectrum

> Spectral data must be processed using MassLynx before it can be analyzed by the Oligonucleotide Sequencing program. Only multiply-charged, negative ion centered spectra can be analyzed.

There are two ways to insert a centered spectrum into a document, copying a peak list from the clipboard or importing a spectrum from a MassLynx file.

Copying a Peak List from the Clipboard

A peak list is a set of m/Z, intensity pairs, listed one pair per line, as shown below.

102.7 4.698e3

109.9 1.684e4

110.5 1.093e4

111.0 1.458e4

113.0 2.320e4

114.0 4.056e3

117.6 8.153e3

119.5 3.889e3

Figure 6.2 Peak List Format

1. Open the listing file in Microsoft Notepad ™ or a similar program, select all the lines containing m/Z, intensity data and copy them to the clipboard using the **Edit, Copy** menu command or its equivalent.

2. Switch to the Oligonucleotide Sequencing program, and select **Edit, Paste Spectrum** or press the button. The peak list will be copied in, replacing any existing spectrum in the document.

3. Edit the Sample Description and Spectrum Description, if desired.

Importing a Spectrum from a MassLynx™ File

There are two ways to import MassLynx data:

Select the **Import Spectrum…** command from the **File** menu. This invokes the **MassLynx Spectrum Browser** dialog, from which a data file and centered spectrum can be chosen.

-or-

Specify a spectrum using the **Spectrum** page of the **Oligonucleotide Sequencing Parameters** dialog. See the following section for more information.

Specifying Sequencing Parameters

Before a spectrum can be sequenced the Sequencer must be given a set of parameters, which describe known information about the oligonucleotide and spectrum, and determine details of the analysis process. The **Oligonucleotide Sequencing Parameters** dialog is used for this purpose.

To open the dialog, select **Process, Parameters** press the button. The dialog has four pages:

- The Sequence page for specifying composition-related information.
- The Oligonucleotide page for providing molecular weight, charge and tolerances.
- The Series Ion page for specifying ion match tolerances.
- The Spectrum page for access to the MassLynx Spectrum Browser.

Press the appropriate tab to view a different page. Parameters are not applied until you press **OK**.

The Sequence Parameters Page

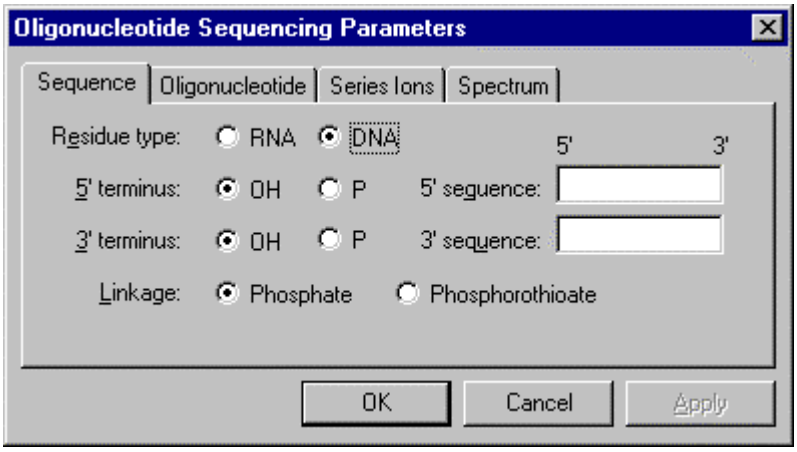


Figure 6.3 The Sequence Parameters Page

Residue Type Class of oligonucleotide, select RNA or DNA.

5' and 3' terminus Select OH for hydoxyl terminus or P for phosphate.

Linkage Choose either phosphate or phosphorothioate linkage. Note all linkages must be the same.

5' and 3' sequence Enter any 5' or 3' sequence known in 5'→3' order.

If termini and linkage type is not known, you can try sequencing with various choices; incorrect choices will usually result in no matching consensus sequences.

The Oligonucleotide Parameters Page

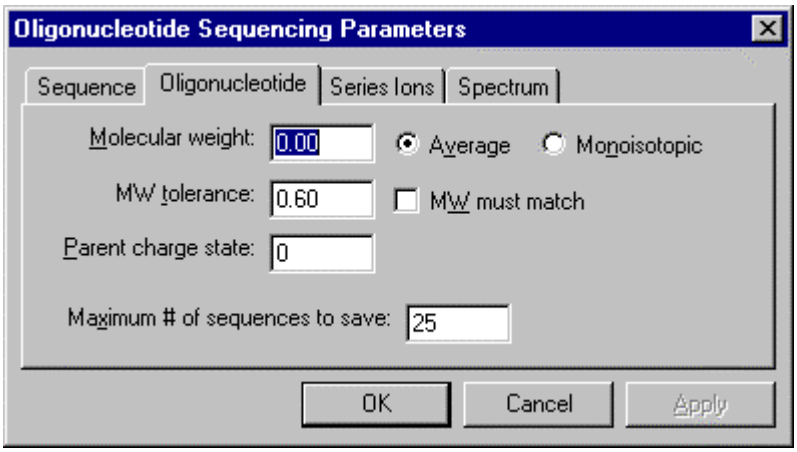


Figure 6.4 The Oligonucleotide Parameters Page

Molecular Weight If the molecular weight is known, enter it in the molecular weight box and check the **MW must match** box. If it is not known enter an estimated molecular weight and make sure the **MW must match** box is not checked.

Average or Monoisotopic Select the appropriate option for the check the MW mode.

MW tolerance Specify the tolerance for matching the calculated MW of consensus sequences.

MW must match As mentioned above this box should be checked if the known molecular weight is entered. If it is ticked only consensus sequences which match the MW within tolerance will be saved. If it is not ticked all sequences will be saved, many of these may be partial sequences.

Parent charge state If known, specify the charge state of the parent ion, if not enter an estimate.

Maximum # of sequences to save This specifies the number of sequences saved during series ion evaluation and in the final consensus set. A reasonable setting is 25. Increasing this value will result in much slower analysis while decreasing it could cause lower scoring, but otherwise correct, sequences to be rejected.

It is important to provide an accurate measurement of oligonucleotide molecular weight and charge state. This is usually determined from the parent spectrum. A narrow molecular weight tolerance helps eliminate sequences which are not correct but which are accepted by the sequencer because their MW falls within the tolerance.

If the spectrum was loaded from a MassLynx file, the MW tolerance will be estimated from the spectral details and the charge state will be extracted from the experiment information. These values can be changed. It is sometimes useful to set the charge state to a lower value than the

experimental value to reduce coincidental matches of high charge state expected series ions to the spectrum.

The Series Ion Parameters Page

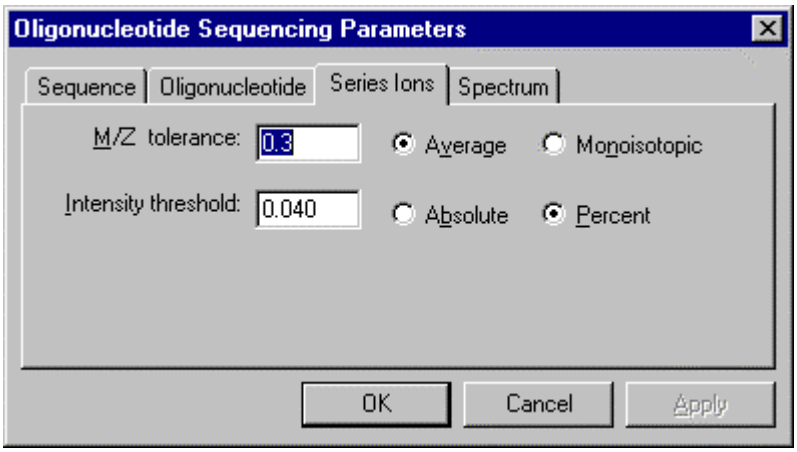


Figure 6.5 The Series Ion Parameters Page

M/Z tolerance Specifies the m/Z tolerance and **Average** or **Monoisotopic** mass for matching series ions to the spectrum.

Intensity threshold Specifies the lower bound on peak intensities as **Absolute** counts or **Percent** of base peak.

This page specifies the most critical parameters for successful sequencing. If the m/Z tolerance is too narrow, the sequencer will not be able to match calculated series ions to the actual spectrum. If it is too wide, then incorrect matches will be made. If the intensity threshold is too low, the sequencer could inappropriately match calculated series ions to noise peaks. If it is too high, then legitimate but weak peaks will be ignored when matching.

Both the m/Z tolerance and intensity threshold are estimated when the spectrum is first loaded. The m/Z tolerance is estimated as twice the minimum m/Z difference between peaks, while the intensity threshold is estimated as twice the minimum peak intensity.

If sequencing is unsuccessful with the estimated parameters, either or both these parameters should be adjusted. A better intensity threshold can be determined by placing the spectrum pane into **Display Peaks Above Threshold** mode (select the **Spectrum, Display Peaks, Peaks Above Threshold** menu item or press the button on the Spectrum Manipulation Toolbar), and examining the effect of changing the threshold. In this display mode, the only peaks displayed in the spectrum plot are those which will be considered during series ion matching. Peaks below the threshold will not be shown. A better m/Z tolerance can be chosen by examining the spectrum or through trial and error.

The Spectrum Parameters Page

This page gives access to the MassLynx Spectrum Browser dialog to permit selection of a centered spectrum from a MassLynx data set. You must have previously processed and saved the centered spectrum using MassLynx before it can be imported into the Oligonucleotide Sequencing program.

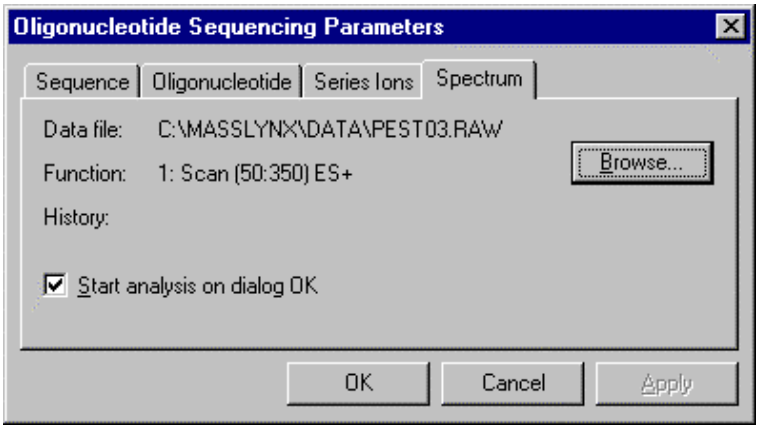


Figure 6.6 The Spectrum Parameters Page

Data file, Function and History	Show details of the selected spectrum.
Browse	Press this button to invoke the MassLynx Spectrum Browser.
Start analysis on dialog OK	If this box is checked the sequencer will start analysing the spectrum when the **OK** button is pressed. If is not checked pressing **OK** will save the parameters but not start the sequencer. It is more useful to leave this box unticked so that the spectrum, intensity threshold, and other parameters might be examined before analysis begins.

The Oligonucleotide Sequencing program makes a copy of the centered spectrum and stores it in the Oligonucleotide Sequencing document. Once this copy has been made, subsequent changes to the MassLynx file will not be reflected in the copy. If you modify the centered spectrum using MassLynx, you must import the modified spectrum into the Oligonucleotide Sequencing document again.

Analyzing the Spectrum

Once a spectrum has been loaded and the parameters have been configured, analysis may proceed.

Select **Predict Sequences** from the **Process** menu or press the button. The cursor will change to the hourglass "wait cursor", and a progress indicator will be displayed at the top of the Consensus Sequences pane. When analysis is complete, the list of consensus sequences will be displayed and the normal cursor will be restored. The highest-ranked sequence will be selected by default. To see results for a different sequence, select it by clicking on the list box.

The time required to analyze a spectrum depends primarily on two parameters: the length of the oligonucleotide (which is determined by the molecular weight) and the number of sequences saved. In both cases, the time required increases approximately with the square of the size. In other words, a spectrum of a 12-mer will take four times as long to sequence as a spectrum of a 6-mer; saving 50 sequences will take four times as long as saving only 25 sequences. This is because of the number of possibilities that must be considered and the fact that the series ion sequences (**w** and **a-B**) are combined by pairs, a nucleotide at a time, to generate the consensus sequences.

Troubleshooting

On occasion, the sequencer will be unable to find any sequences consistent with the spectrum and sequencing parameters, or may find sequences which are known from other information to be incorrect. This could result from a number of problems:

- A poor quality spectrum was used. If the input spectrum has low signal-to-noise, or a low number of multiply-charged series ions due to incomplete fragmentation, or a parent ion of low charge state, it may not be possible for the sequencer to match sufficient expected series ions to spectral features to construct consensus sequences.

- The molecular weight is inaccurate. There are typically only a small number of nucleotide compositions consistent with a given MW and, of these, even fewer that will produce a consistent set of series ions. Thus, the more accurate the MW, the fewer the possibilities.

- The m/Z tolerance is too low. If the mass accuracy or resolution of the spectrum is low the tolerance must generally be made wider, otherwise the computed series ion values will be outside the tolerance window and matching will fail. If the sequencer is unable to extend a candidate sequence because series ions for the next residue cannot be matched, the partial sequence is abandoned.

- The m/Z tolerance is too wide. This will result in coincidental matches of series ions for incorrect nucleotides, leading to incorrect consensus sequences. This problem will be most evident in the Analysis Details report - there will be many matches to low-intensity peaks, or inappropriate matches of high charge state ions.

- The intensity threshold is too high. The sequencer ignores spectral peaks below the threshold when matching expected series ions. If the threshold is too high, legitimate peaks will be ignored and incorrect or no sequences will be found.

- The intensity threshold is too low. As in the case of too wide an m/Z tolerance, too low an intensity threshold will result in coincidental matches and extension of the series sequences with inappropriate residues.

- The maximum charge state is set too high. This causes, coincidental matches of higher charge state expected series ions to high intensity, low m/Z spectral features. While the charge state entered in the parameters dialog may be correct, for the parent ion, there may not be many fragments in the daughter spectrum with charge state that high. Because the sequencer will consider all charge states possible for a given series length, expected series ions may be coincidentally matched. Matches to peaks of high intensity will skew the scoring in favor of these incorrect residues. This problem can be diagnosed by examining the Analysis Details report for inconsistencies in series ion matches and intensities. Using a maximum charge state setting one or two units lower than the parent ion charge state can correct the problem.

- The spectrum does not contain sufficient series ions to fully derive the oligonucleotide sequence. Series ion intensity tends to decrease for residues further into the sequence from either end. (*E.g.* $\mathbf{w_1 > w_2 > w_3 \ldots}$) Incomplete sequences will not match the oligonucleotide MW, so if the "MW must match" condition is set, no sequences will be reported back. In practice, ES-MS/MS will produce useful spectra for oligonucleotides of 15 residues in length or less. Under very favorable conditions, longer oligonucleotides can be successfully sequenced.

- The 5' or 3' termini or linkage were incorrectly chosen. This will cause incorrect computation of expected m/Z for series ions and incorrect MW calculation for the resulting oligonucleotide. As a result, no sequences will be found.

- The 5' nucleotide is actually cytosine, but a different nucleotide was chosen. Series ions arising from cytosine residues are generally weaker than those arising from other residues. Since the choice of 5' residue is principally based upon the intensity of the matching peak, monomer peaks corresponding to other nucleotides may be preferentially chosen. The series thus starts off wrong, leading to incorrect consensus sequences. This problem can be corrected by forcing the 5' residue to be C by preassigning it in the Oligonucleotide Sequencing parameters dialog.

In many cases, determining the sequence of an unknown oligonucleotide is a trial and error process. It is sometimes necessary to analyze several spectra of differing parent charge states and compare results, and to perform several analyses of each spectrum with differing parameter sets.

The Oligonucleotide Sequencing Interface

Oligonucleotide Sequencing Documents

Information in the Oligonucleotide Sequencing program is organized into *documents*. A document, when complete, contains all of the information produced during the analysis of a spectrum to predict oligonucleotide sequences. This information includes a copy of the centroided spectrum and its descriptive comments, all parameters used in the analysis, and all analysis results.

The process of analyzing a spectrum to predict sequences builds this Oligonucleotide Sequencing document. The program provides an interactive, graphical way to build and view these documents.

The Main Window

The Oligonucleotide Sequencing program consists of a **Main Window** containing a **Menu**, optional **Tool Bars**, and a **Status Bar**. Within the Main Window, one or more **Child Window(s)** may be positioned, each containing a multiple-part view of a set of sequencing results. The parts of the Main Window and a Child Window are shown in Figure 6.7.

Child Windows

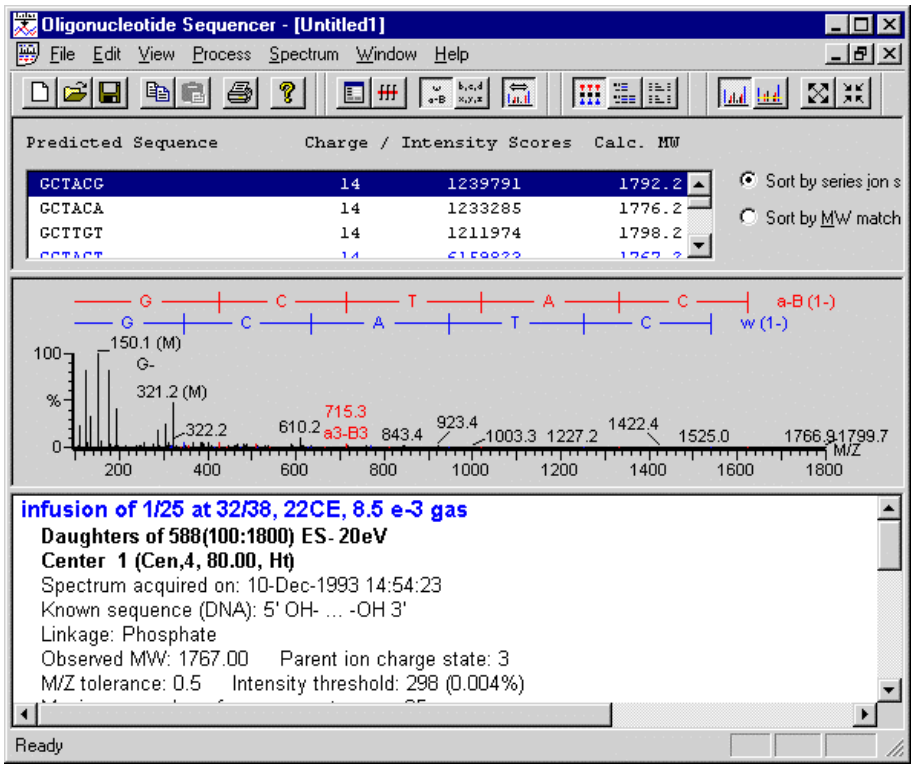


Figure 6.7 Oligonucleotide Sequencing Program Main Window

As described above, the Main Window can contain one or more Child Windows, each representing a different part of the Oligonucleotide Sequencing document. Using commands from the Menu, a Child Window may be cloned into two or more Child Windows, each giving an independent view of the same document. This allows results of two different candidate sequences

to be studied side-by-side. Alternatively, two documents may be opened simultaneously (in two separate child windows), permitting side-by-side comparison of sequences derived from analyses with different parameter settings or daughter spectra from two different precursor ions of the same parent spectrum.

Each Child Window is divided into three **Panes**, containing the consensus sequences, spectrum, and detailed reports, respectively.

The Consensus Sequence Pane

The topmost pane (Figure 6.8) holds a list box containing predicted consensus sequences from an analysis of the spectrum. These consensus sequences may be sorted in one of two ways: by the score derived from series ion matches, or (if the restriction on MW matching is relaxed), by increasing difference between the observed and calculated molecular weight. Sequences that match the observed molecular weight (within tolerance) are printed in blue; non-matching sequences are printed in black. The currently selected sequence is highlighted.

When a new sequence is selected, the contents of the other two panes will change to reflect analysis details for that sequence.

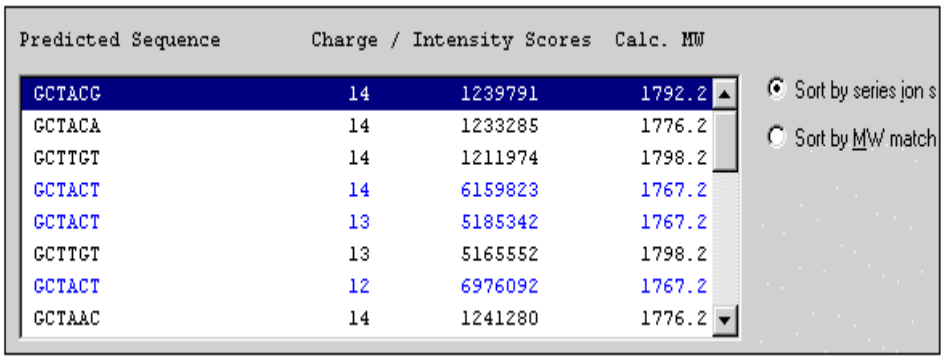


Figure 6.8 The Consensus Sequence Pane

The Spectrum Pane

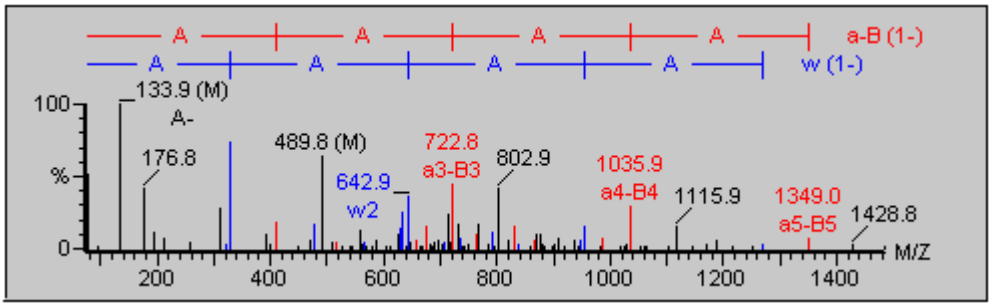


Figure 6.9 The Spectrum Pane

The middle pane contains a plot of the spectrum, annotated with details from the currently selected consensus sequence. The annotations which appear on the spectrum can be specified using the toolbar or menu commands described below.

The Sequence Report Pane

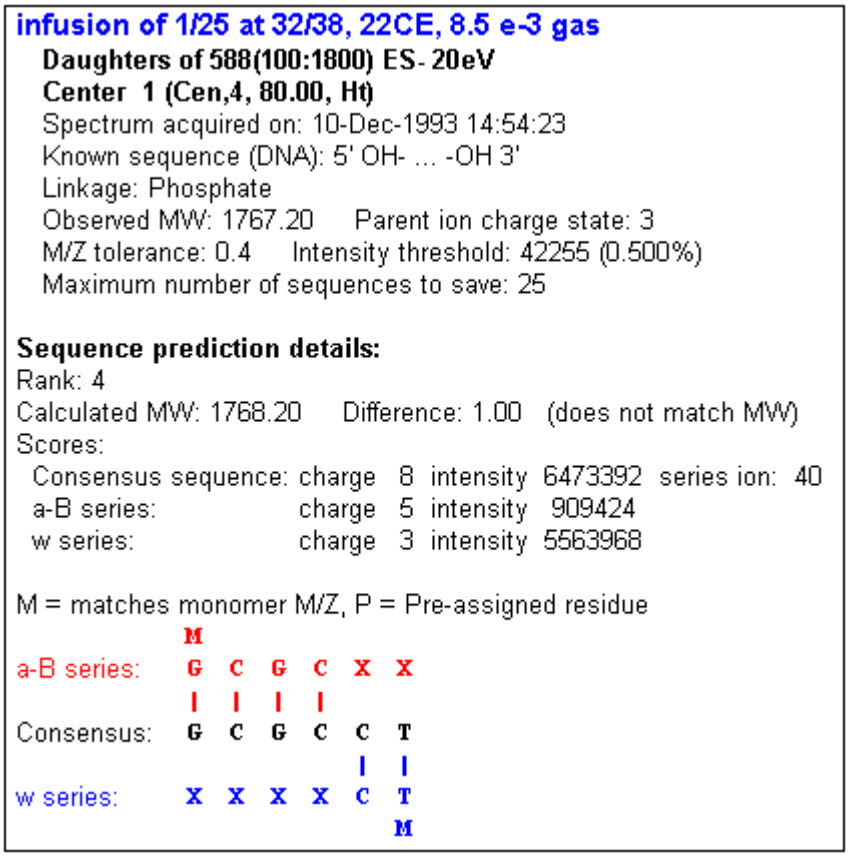


Figure 6.10 The Sequence Report Pane, showing a Sequence Prediction

The bottom pane contains one of three text-based reports describing analysis details for the currently selected sequence.

Sequence Prediction Details	This report summarizes scores and consensus sequence derivation.
Analysis Details	This report describes the series ions expected and matched and their contributions to scoring.
Assignment	This report providing spectral assignments for **w, a-B,** and other series ions from the consensus sequence. The report displayed can be chosen from the Oligonucleotide sequencer **View** menu.

Toolbar

There are four toolbars, each containing buttons for a specific program function.

- The **Standard** toolbar contains buttons for file, edit, printing, and on-line help functions.
- The **Sequence** toolbar contains buttons to invoke the parameter specification dialog, sequence the spectrum and control the labeling of sequencing results on the spectrum display.

- The **Spectrum Manipulation** toolbar has buttons for controlling the format of the spectrum display.

- The **Text Report** toolbar permits quick selection of the display of textual results.

Each of the four toolbars can be independently docked against one of the sides of the Main Window, or it is left floating. To dock a floating toolbar or to change the location of a docked toolbar, simply click on the bar in the background area (between or surrounding the buttons) and drag it. An outline of the bar will appear to aid positioning. When a floating toolbar is brought near the edge of the Main Window, the outline will change to show its orientation when docked. Releasing the mouse will dock the toolbar. To float a docked toolbar, click and drag it off the toolbar. Move it away from the edge of the Main Window and release the mouse. The toolbar will float in a "mini-window" frame, as shown in the figures below.

The Standard Toolbar

Figure 6.11 The Standard Toolbar

The **Standard Toolbar** provides shortcuts for commonly used **File** and **Edit** menu commands.

Press the **New File** button to create a new, empty Oligonucleotide Sequencing document. This document must be configured with a centered spectrum and other parameters before sequencing can proceed. (Menu equivalent: **File, New**).

Press the **Open File** button to invoke the File Open dialog. This retrieves a previously-saved Oligonucleotide Sequencing document and open a new child window containing it. By convention, Oligonucleotide Sequencing documents use the file name extension **.OLI**, so the File Open dialog is pre-loaded with all of the existing **.OLI** files. (Menu equivalent: **File, Open**).

Press the **Save File** button to save the currently selected Oligonucleotide Sequencing document. If the document has not previously been saved the File Save As… dialog will be invoked to prompt for a name. While any name may be chosen, by convention, the extension **.OLI** should be used. (Menu equivalent: **File, Save** or **File, Save As…**, as appropriate).

Press the **Copy** button to make a copy of the currently selected pane in the Windows Clipboard. E.g., to copy the list of consensus sequences to the Clipboard, click in the **Consensus Sequences Pane** to select it, then click the **Copy** button. The contents of the pane are copied into the Clipboard as a Windows metafile, and may be pasted into other applications which accept metafile formats. (Menu equivalent: **Edit, Copy**. Note that the text of the menu item will change to reflect the currently selected pane i.e. **Edit, Copy Spectrum** if the spectrum pane is current).

Press the **Paste Spectrum** button to retrieve a centered mass spectrum from the Clipboard and replace the current spectrum. The Clipboard spectrum must be in the form of a peak list of m/Z, intensity pairs, one pair per line, in increasing m/Z order. Simply load the text file into Notepad or some other text editor, select all the peaks, and copy them to the Clipboard. The **Paste Spectrum** button will then be enabled, and the spectrum can be pasted into the Oligonucleotide Sequencing document. (Menu equivalent: **Edit, Paste Spectrum**).

Press the **Print** button to print a report according to the currently selected reporting options (see **Report Options**, below). (Menu equivalent: **File, Print…**).

The Sequence Toolbar

Figure 6.12 The Sequence Toolbar

The **Sequence Toolbar** provides a set of buttons and controls to define sequencing parameters, run the sequencer and to alter the labeling of the spectrum with sequencing results.

Press the **Parameters** button to invoke the **Oligonucleotide Sequencing Parameters** dialog and specify sequencing parameters and conditions. (Menu equivalent: **Process**, **Parameters**).

Press the **Predict Sequences** button to predict sequences for the spectrum using the current parameter set. (Menu equivalent: **Process, Predict Spectrum**).

Press the **Label w and a-B Series** button to show these series ion labels on matching spectrum peaks. The labels shown are those for predicted ions of nucleotides, of the currently selected consensus sequence, for which matching spectral peaks were found. Pressing the button again will turn the labeling mode off. (Menu equivalent: **Spectrum, Annotate Peaks, w and a-B Series**).

Press the **Label Other Series** button to show labels for all other series ions for the current sequence on matching spectral peaks. Pressing the button again will turn the labeling mode off. (Menu equivalent: **Spectrum, Annotate Peaks, Other Series Ions**).

The Spectrum Manipulation Toolbar

Figure 6.13 The Spectrum Manipulation Toolbar

The **Spectrum Manipulation Toolbar** provides some shortcuts to control the threshold and range of the spectrum display.

Press the **Display All Peaks** button to show every peak in the spectrum, regardless of intensity. (Menu equivalent: **Spectrum, Display Peaks, All Peaks**).

Press the **Display Peaks Above Threshold** button to show only those peaks with intensity exceeding the threshold set in the sequencing parameters. The peaks displayed are those which will be used in sequencing; any peaks of lower intensity will be ignored. (Menu equivalent: **Spectrum, Display Peaks, Peaks Above Threshold**).

Press the **Zoom Out** button to restore the spectrum display to full m/Z range. (Menu equivalent: **Spectrum, Zoom Out**).

The **Zoom In** button decreases the displayed range by half, around the center of the current range. For example, if the current range is 100 < m/Z < 500 (a 400 Da range, centered at 300 Da) the new range will be 200 < m/Z < 400 (a 200 Da range, centered at 300 Da). This is a quick way to zoom in to the currently displayed range. (Menu equivalent: **Spectrum, Zoom In**).

The Text Report Toolbar

Figure 6.14 The Text Report Toolbar

The **Text Report** toolbar is used to change the details displayed in the lower, Sequence Report pane. The Sequence Report pane always shows a report for the currently selected consensus sequence. The report may be one of three types.

Pressing the **Consensus Sequence** button displays the Sequence Prediction Details report summarizing the consensus results for the currently selected sequence. (Menu equivalent: **View, Consensus Sequence**).

Sequence prediction details:
Rank: 8
Calculated MW: 1758.20 Difference: -9.00 (does not match MW)
Scores:
Consensus sequence: charge 13 intensity 5185315 series ion: 138
a-B series: charge 12 intensity 1218339
w series: charge 1 intensity 3966976

M = matches monomer M/Z, P = Pre-assigned residue

```
            M
a-B series: G  C  T  C  T  X
            |  |  |  |  |
Consensus:  G  C  T  C  T  T
                           |
w series:   X  X  X  X  X  T
                           M
```

Figure 6.15 The Sequence Prediction Details Report

The **Sequence Prediction Details** report shows how the consensus sequence was derived from nucleotides in the **w** and **a-B** series.

Scores This section gives the charge and intensity scores for both ion series and the consensus sequence. The **charge** score is the count of matches of expected series ions to spectral peaks; each matching peak (above threshold) adds one to this count.

Intensity Is the sum of intensity for each matching peak, except those denoted as "monomers". Monomer peaks may arise from multiple cleavage fragmentation along the oligonucleotide backbone which results in ions due to a single nucleotide. Since it is not possible to determine whether the ions arise from this type of fragmentation as opposed to a series ion fragmentation, matches to monomer peaks yield ambiguous information and thus cannot be reliably used to indicate sequence. In addition, monomer ions are generally of much higher intensity than other series ions. For these reasons, monomer matches are counted when determining the charge score (credit is given for the *presence* of the residue), but not the intensity score (no credit because of the positional ambiguity).

Consensus Sequence

An additional score: the score derived from matches to the nine series ion (described above) used to determine overall ranking.

The **a-B** series derives sequence from the 5' end; these residues are shown in red, with vertical bars extending to the corresponding nucleotides in the consensus sequence. The **w** series derives sequence from the 3' end; these residues are shown in blue. An "X" marks residue positions that were not considered when aligning series nucleotides to derive the consensus.

Series residues with expected ions matching a monomer ion are labeled with an "M", these ions are used when determining charge score but not intensity score. It is possible to pre-assign both 5' and 3' starting sequence as part of the parameter setup; pre-assigned nucleotides are labeled with a "P".

Pressing the Analysis Details button displays the Analysis Details report (Figure 6.16) for the current consensus sequence. (Menu equivalent: View, Analysis Details). This is shown below:

The **Analysis Details** report describes the construction of the **w** and **a-B** ion series for the current consensus sequence. The numbering convention shown in Figure 6.16 is used for each series. The report is organized as columns by charge state in increasing order (up to the maximum charge state specified), and by row in increasing residue number order. The entry for each residue/charge state combination contains the m/Z expected, the m/Z found and its intensity (if any). The count of charge states and sum of matched intensity for all charge states is listed along the left for each series type. Monomer ions are flagged, but while the intensities for these ions are listed, they are not used in score computations.

The Analysis Details report is useful for determining the correct consensus sequence, when there are several possibilities of the correct molecular weight, or when the analysis has failed to fully sequence the spectrum. For example, coincidental matches to expected ions can be identified as isolated entries in the table (a residue matches only a single charge state) or where intensities do not appear to follow a consistent distribution over charge states or down through a series. In general, for a given charge state, the intensity of matched ions tends to decrease further into the series ($\mathbf{w_1} > \mathbf{w_2} > \mathbf{w_3}$ …).

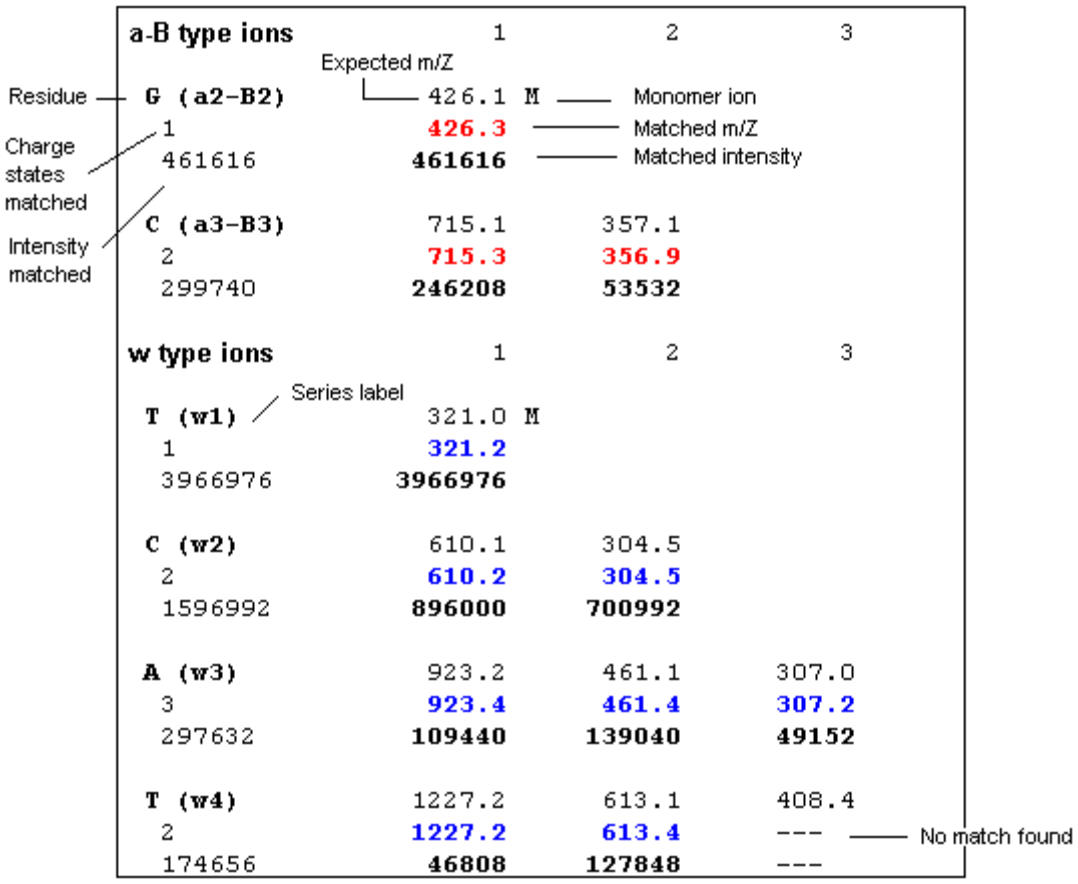


Figure 6.16 The Analysis Details Report

Press the **Series Ion Assignments** button to display the **Series Ion Assignments** report. This report lists the spectral assignments for the **w**, **a-B**, and other ion series, based on the assumption that the selected consensus sequence is correct. The assignment detail for each matched ion includes the m/Z value, series ion label and charge state. For **w** and **a-B** series, the partial sequence is also supplied for the first matching ion. Depending on the sequence, it is possible for assignment redundancies to occur; in this case multiple labels will be listed or the ion may appear in more than series list. (Menu equivalent: **View, Series Ion Assignments**). See Figure 6.17 below.

Partial sequences are listed from either the 5’ or 3’ end of the oligonucleotide, as appropriate. Linkages are denoted as “p”, sugars as “s”, and for **a-B** series the lost base is shown in parentheses.

The Series Ion Assignment report is also useful for evaluating whether a sequence has been correctly predicted. If there are many redundant assignments or gaps in series, it can be an indication that the sequence is not correct.

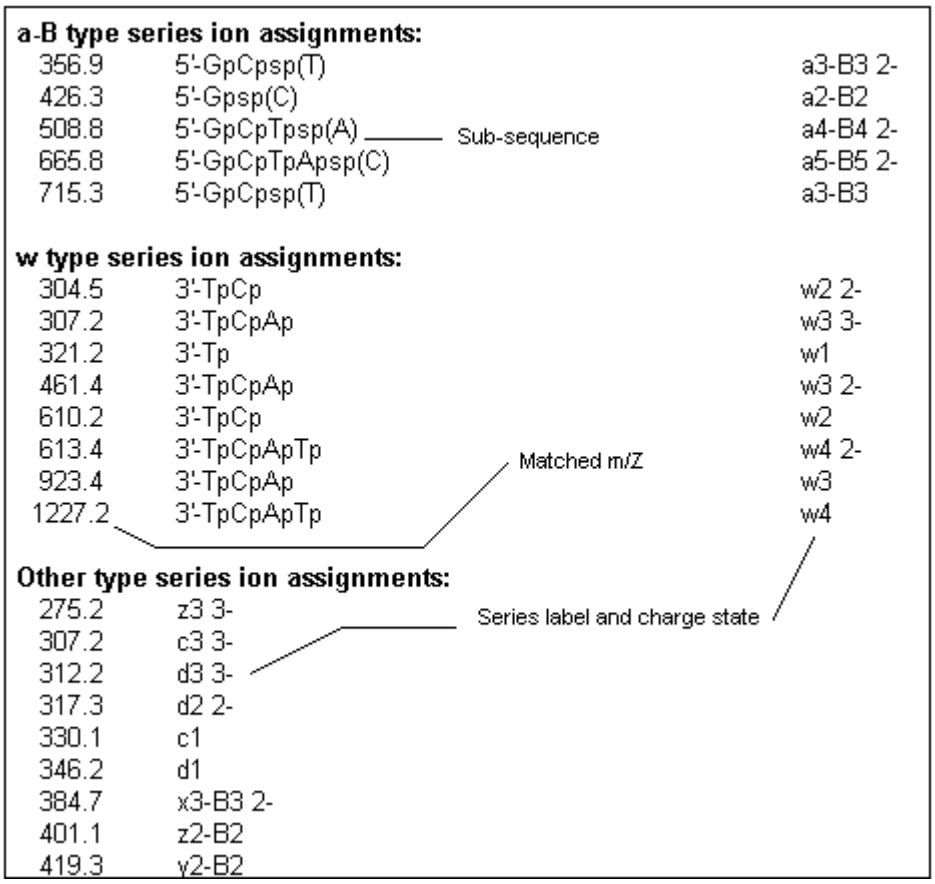


Figure 6.17 The Series Ion Assignment Report

Menu Commands

The toolbars provide shortcuts for most menu items, and these have been described above. There are a few additional menu items, which have no toolbar shortcuts:

File Menu

Import Spectrum Invokes the Spectrum Data Browser to select a MassLynx spectrum. If the Oligonucleotide Sequencing document already contains a spectrum, you will be asked if you wish to replace this data. If you press the **Yes** button, the Spectrum Browser will appear to let you select a new spectrum.

> Only centered spectra may be selected; the Oligonucleotide Sequencer program cannot process continuum data. Once a spectrum has been selected and the dialog has been closed, the spectral data will be copied into the Oligonucleotide Sequencing document. Sequencing results will be deleted, since they were based on the previous spectrum and are no longer consistent.

Close Closes the currently selected Oligonucleotide Sequencing document and removes any Child Windows viewing it. If the document has not previously been saved, you will be prompted to save it and to provide a name. If the document has been previously saved, you will be prompted to save the changes.

Print Setup Invokes the **Print Setup** dialog to permit selection of the printer and its properties.

Print Preview	Displays a preview of the printed report for the current document, according to the printing options. The display is as close as possible to what will actually appear on paper
Report Options	Invokes the **Report Options** dialog. This dialog is used to select the types of printed reports and their appearance. It is a *tab dialog* with two pages. You may specify the report types on the first page in Figure 6.19.
Exit	Exit the Oligonucleotide program. Any open documents will be closed. The layout and appearance of the Main Window, Toolbars and Child Windows will be saved to the registry

Report Options

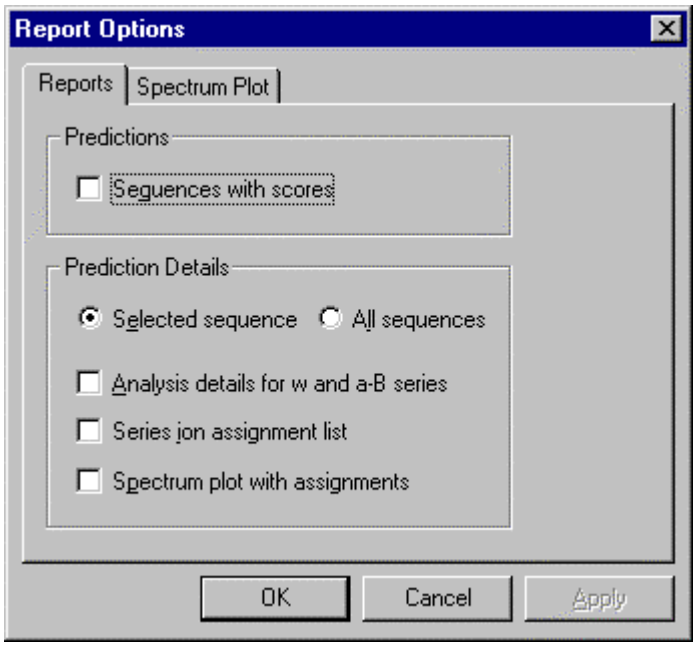


Figure 6.18 Report Options, Reports page

Reports

Sequences with scores	Check this box to print a report listing the consensus sequences and their scores.
Selected sequence	Check this box to print reports for the currently selected consensus sequence, or **All Sequences** to print reports for all consensus sequences. Note the latter will produce many pages.
Analysis details for w and a-B series	Check this box to print the Analysis Details report
Series ion assignment list	Check this box to print the Series Ion Assignments report.
Spectrum plot with assignments	Check this box to print an annotated spectrum plot. You may specify the appearance of the spectrum plot on the second page, shown below. Any annotation options or magnification ranges in effect for the **Spectrum Pane** are also applied to the hard-copy plot. The spectral range may be specified independently for the hard-copy plot without affecting the **Spectrum Pane**.

Spectrum Plot

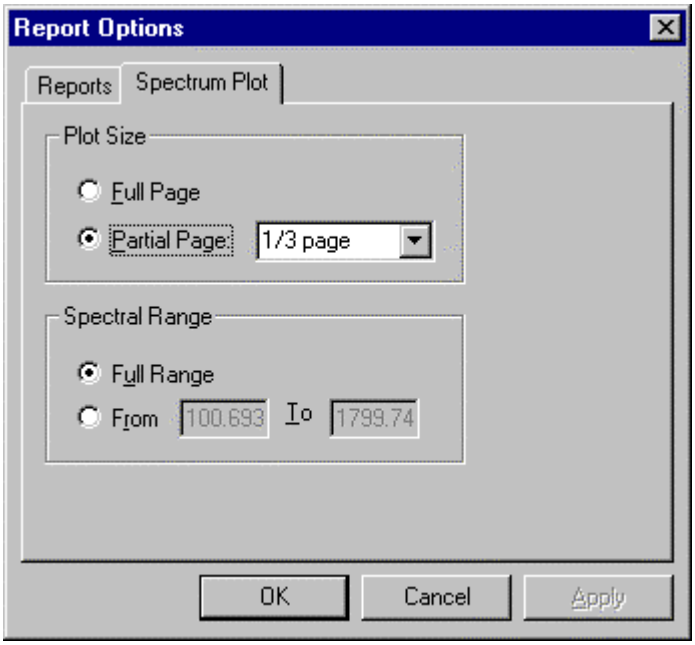


Figure 6.19 Report Options, Spectrum Plot page

Plot Size	Select **Full Page** or **Partial Page**. If **Partial Page** is selected, the user can choose one of Full Page, ½ page, ⅓ page or ¼ page.
Spectral Range	Select **Full Range** or **From** and **To**. If **From** and **To** are selected enter the m/Z range, in Da, of the range required.
Load Last File	When ticked, the last Oligonucleotide Sequencing document viewed will automatically be loaded the next time you use the Oligonucleotide Sequencing program. If you do not tick this item, the program will be started with a new, empty document.
Most Recent Files	This list contains the names of the last four Oligonucleotide Sequencing documents viewed. To open any of them in a new Child Window simply select it from the list.

Edit Menu

Sample Description	Invokes a dialog to let you change the sample title. The title is normally copied from the MassLynx data file. Changes affect only the text in the Oligonucleotide Sequencing document; the original MassLynx file remains unaltered.
Spectrum Description	Invokes a dialog to let you change the spectrum comments. These are normally copied from the MassLynx data file. Changes affect only the text in the Oligonucleotide Sequencing document; the original MassLynx file remains unaltered.

View Menu

Font Sizes	Invokes the **Font Sizes** dialog, to let you change the default font sizes for various components in reports and plots. These sizes are applied to on-screen displays and printed reports.

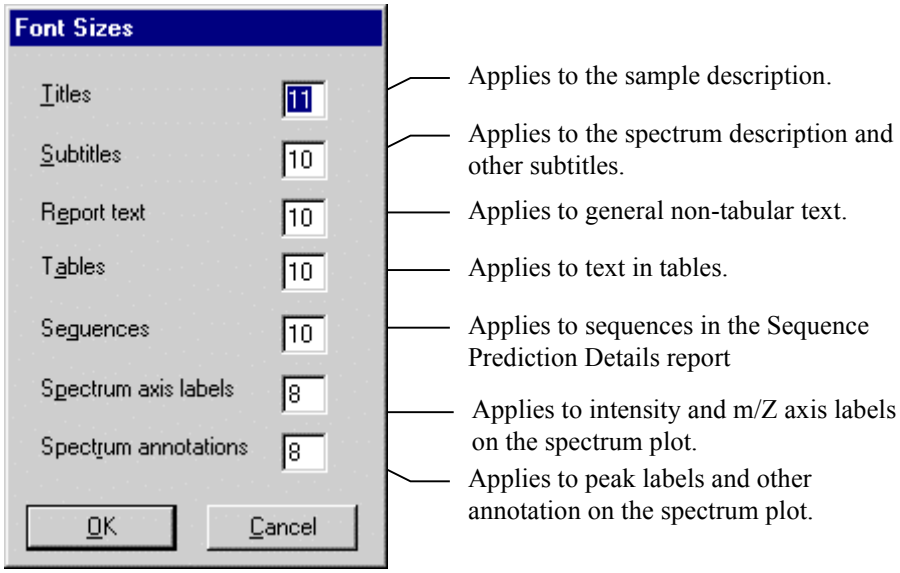


Figure 6.20 The Font Sizes dialog

Toolbars Invokes the **View Toolbars** dialog to change which toolbars are visible. If a box is ticked, the corresponding toolbar will be made visible. If not ticked, the toolbar will not appear on screen. The size and position of a toolbar is independent of visibility.

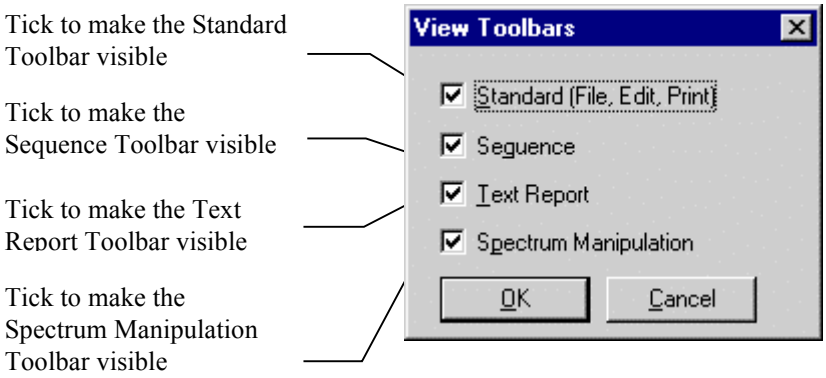


Figure 6.21 The View Toolbars dialog

Status Bar When ticked, the status bar at the bottom of the Main Window is displayed. It is used to display the program state and to provide more information about menu commands and toolbar buttons. To see more information about a menu item, pull down the menu and highlight the desired item; to see toolbar button information, position the cursor over the button. Descriptive information will then appear in the status bar.

Spectrum Menu

Magnify Invokes the **Spectrum Magnify** dialog to let you add or change magnification ranges in the spectrum plot.

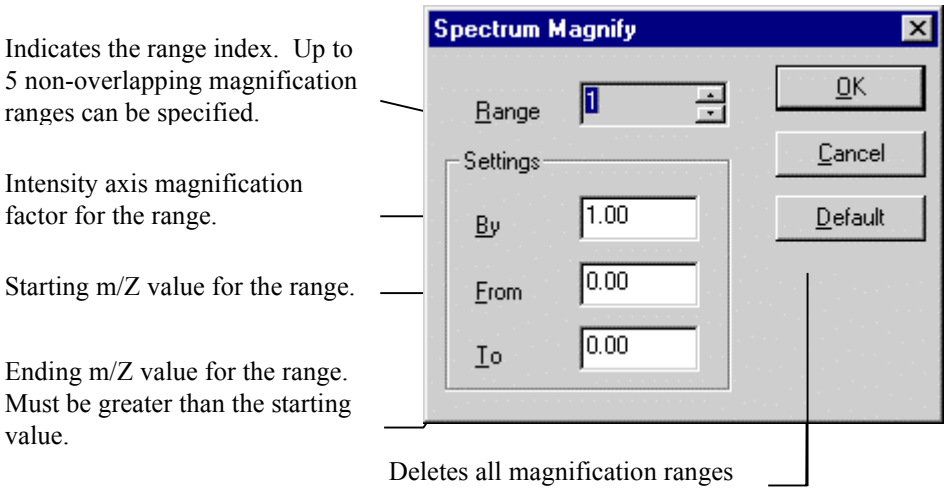


Figure 6.22 The Spectrum Magnify dialog

Window Menu

New Window — Creates a new Child Window containing a second view of the currently selected Oligonucleotide Sequencing document. Results for a different selected consensus sequence may viewed in the second window; however, changes to the *document* contents (such as parameters or re-sequencing) will affect the contents of both windows.

Cascade — If more than one Child Window is open, this command will resize and arrange them in an overlapping cascade within the Main Window.

Tile Horizontally — If more than one Child Window is open, this command will resize and arrange them so that they are stacked one above the next within the Main Window.

Tile Vertically — If more than one Child Window is open, this command will resize and arrange them side-by-side in the Main Window.

Arrange Icons — If any Child Windows are minimized, this command will arrange their icons along the bottom of the Main Window. The icon for a minimized Child Window is shown below:

Figure 6.23 Oligonucleotide Sequencing document icon

Open Documents List — The list of Child Windows containing open Oligonucleotide Sequencing documents appears at the end of the Window menu. The currently active Child Window is ticked. Select a different Child Window to make it active and bring it to the top of the stack of overlapping Child Windows.

References

1. McLuckey, S.A., Van Berkel, G.J., and Glish, G.L., *J. Am. Soc. Mass Spectrom.* **1992**, 3, 60-70

2. McLuckey, S.A. and Habibi-Goudarzi, H., *J. Am. Chem. Soc.* **1993**, 115, 12085-12095

3. Boschenok, J. and Sheil, M.M., *Rapid Comm. Mass Spectrom.* **1996**, 10, 144-149

4. Pomerantz, S.C., Kowalak, J.A., and McCloskey, J.A., *J. Am. Soc. Mass Spectrom.* **1993**, 4, 204-209

Chapter 7 ProteinLynx

Contents

Figures 7-2
Introduction 7-2
Automated Data Processing & Output 7-3
Automated Database Searching using ProteinLynx Global Server 7-3
Data Browser for Rapid Results Review 7-3
Overview 7-3
    Output File Format 7-4
    Data processing 7-5
Installation 7-6
    Installing ProteinLynx on the Instrument PC 7-7

Figures

Figure 7.1 Example Sample List for an Electrospray data processing 7-4
Figure 7.3 Example Sample List - MALDI data processing 7-6

Introduction

ProteinLynx is a suite of software tools for automated high-throughput protein/peptide identification and characterisation using MALDI and/or Electrospray mass spectrometry. The tools comprise:

- Automated Data Processing & Output.
- Automated Database Searching using ProteinLynx Global Server.
- Data Browser for Rapid Results Review.

Automated Data Processing & Output

Currently this can be configured for both MALDI and Q-Tof MS-MS data.

Software is provided to define parameters to enable post-processing of acquired data, prior to database searching and automatic execution as part of a batch analysis.

Automated Database Searching using ProteinLynx Global Server

Software is provided to define parameters used for the automatic searching of a query, based on the processed data, against a protein sequence database.

Data Browser for Rapid Results Review

The ProteinLynx Results Browser is used to review spectra and database search results for a batch of samples. The key features are:

- A microtitre plate view allowing the user to navigate to a specific well i.e. sample.
- A spectrum view, showing the processed spectrum for a selected well for MALDI analysis or annotated MS/MS spectrum of a specific matched peptide from a function switching QTof experiment.
- Database search results (hit list) view for a specific sample.
- A window for either viewing the database entry text, coverage map or matching peptide results.

Overview

The ProteinLynx system requires that a RAW data file be present in the location specified in the File Name column in the sample list. If necessary, create the required acquisition method files. See the relevant sections in the MassLynx User's Guide and the Guide to Data Acquisition for details.

1. Create a new Project in MassLynx, if required.
2. Within the Project, create the relevant acquisition files: MS Tune, MS Method and Inlet Method.

3. Create a Sample List.

4. Create a ProteinLynx parameter file.

5. Run the Sample List to acquire and/or process data to create a ProteinLynx Report file (*.mrp) and if the Full Text option is selected export a text file with hit results and protein ID.

The **Process** column in the Sample List must contain **PeptideAuto** and the **Parameter File** column should contain the name of parameter file created in step 4 above. A list of accessible parameter files may be displayed by double-clicking the right hand corner of any cell in the Parameter File column.

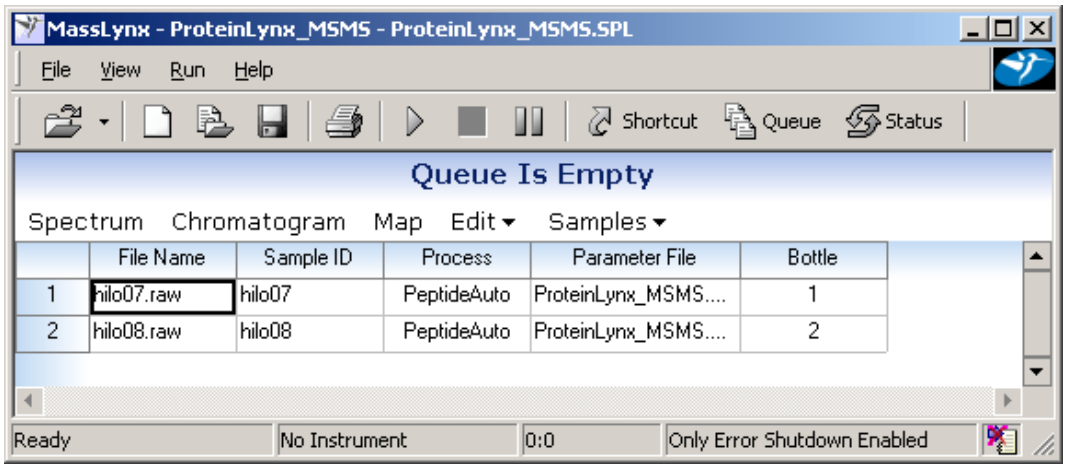


Figure 7.1 Example Sample List for an Electrospray data processing

To acquire and process data select both the **Acquire Sample Data** and **Auto Process Samples** options on the **Start Sample List Run** dialog. To process data check only the **Auto Process Samples** option.

Output File Format

Selecting **Electrospray** on the **Instrument type** page of the **Setup** dialog will allow data to be processed and searched using the ProteinLynx Global Server database search engine. Data can be output in either SEQUEST© compatible or Micromass Peak List (MASCOT© compatible) format. These export MS/MS data in ascii/text or XML format for further processing by third party software.

The SEQUEST-compatible files are named according to the following format:

Filename.function#scan#.dta

The file consists of a header line entirely derived from the survey scan with the following format:

(Precursor m/z)(SPACE)(charge state)

2

115.1027 2.0000 etc…

The Micromass Peak List files are named according to the following format:

Filename.scan#.function#.charge.pkl – where charge is the charge state of the precursor ion obtained from the survey scan i.e. function 1.

If **Append Data to a single file** has been selected all the survey scans are saved in one file.

If **Append Data to a single file** has not been selected in ProteinLynx Setup, the file consists of a header line entirely derived from the survey scan with the following format –

(Precursor m/z)(SPACE)(intensity)(SPACE)(charge state)

The rest of the file contains the peaklist of mass/intensity pairs derived from the MS/MS data. For example:

528.9359 419.0000 3

84.0828 2.0000

101.0962 4.0000

115.1027 2.0000 etc.

MSMS output files are saved in the directory defined on the **Select output** page of the **Setup** dialog; if no directory was specified, files are written to the Windows Temp directory.

Selecting **M@LDI** on the **Instrument type** page of the **Setup** dialog will allow data to be processed and searched using ProteinLynx Global Server database search engine. Data can also be output in either Micromass Peak List (MASCOT© compatible) format (two options) or Protana (PepSea) **format.**

Data processing

A ProteinLynx Report file (*.mrp) as well as text output files are produced automatically when a sample list is processed. The type of output file is defined in the ProteinLynx Setup dialog.

Browser Report Files

To run the browser, select **ProteinLynx Results** from the **ProteinLynx** menu on the top level MassLynx Shortcut Bar. The browser may be run at any time after the auto-processing has begun. Selecting ProteinLynx Results will display the most recent ***.mrp** file, to open a previous file use the **Open Workspace** command on the **File** menu. To update the view as the auto-processing and searching proceed, periodically select the **Refresh** option on the **View** menu, or hit the **F5** key on your keyboard.

The s2 column in Figure 7.2 indicates a **spare** column. There are 5 spare columns available in the **Customize Field Display** option in the sample list. They can be used for adding extra information to the ProteinLynx Report file (gel spot number, pI, molecular weight information etc) for each sample. To appear in the report file this information must be entered in the Sample List.

Text Report Files

Exported files are named according to the following convention:

sample_identifier_outputType.txt

where *sample_identifier* is the name defined in the *Sample ID* column in the Sample List, and *outputType* is the type chosen on the **Select output** page of the **Setup** dialog. E.g. if Sample ID = 980715ROM and output type = mass the file will be named 980715ROM_mass.txt.

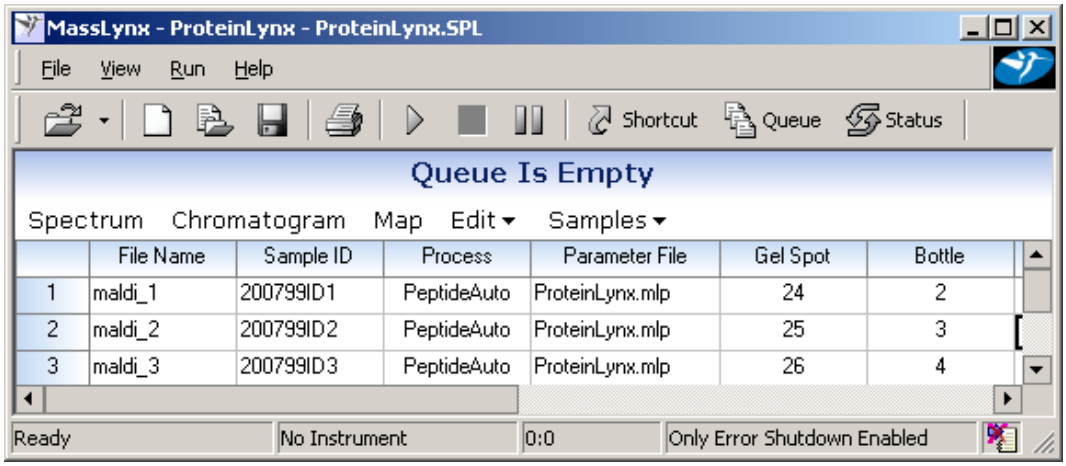


Figure 7.2 Example Sample List - MALDI data processing

XML Report Files

Selecting **M@LDI** or **Electrospray** on the instrument type page allows data to be processed as XML. These files are named as for ***.pkl** except with the ***.xml** extension. However, unlike ***pkl** files you can't have multiple file output and only one single XML file can be output.

Full Text Search Files

The Full Text Search option allows protein ID results to be written to ascii/text files. These files are named according to the following convention:

sample_identifier_outputType.txt

where *sample_identifier* is the name defined in the *Sample ID* column in the Sample List, and *outputType* is the type chosen on the **Select output** page of the **Setup** dialog. E.g. if Sample ID = 980715ROM and output type = mass the file will be named 980715ROM_mass.txt.

For each hit from the database three views are written to the file. View1 is the protein entry name and sequence as it appears in the database. View2 gives detailed information on scores, MW, pI, matching peptides etc. View3 shows the sequence only in FASTA format.

Database Searching

If the **Enable Database Search** option is selected the **Database Setup** button will be enabled. This allows ProteinLynx to automatically generate a query and pass it to ProteinLynx Global Server. The results of the database search are written to a text-format file in the directory containing the relevant RAW data file, and may be viewed by opening the relevant Browser report file.

A summary of the sample list, processing details, search query and results are written to a text-format file in the DATA directory containing the relevant RAW data file. This is the ***.mrp** file which can be opened by selecting ProteinLynx Results from the ProteinLynx menu or selecting Open Workspace from the ProteinProbe File menu. It can also be viewed as a text file in Word, Notepad etc. For each sample a ***.prp** file is written, these results can be viewed in the ProteinProbe.

Installation

ProteinLynx is installed on the instrument PC. In order to use database searching ProteinLynx Global Server is required. Consult the ProteinLynx Global Server User's Guide for details.

Installing ProteinLynx on the Instrument PC

The Instrument PC should be using the **Windows NT/2000** operating system. Install MassLynx on the Instrument PC from the MassLynx CD see the "Introduction" chapter of the MassLynx User's Guide.

During installation, you will be asked which **Options** you wish to install. You should select **MassLynx**, **ProteinLynx**, **BioLynx** and any other options you want to install.

Chapter 8 ProteinLynx - Setup

Contents

Figures 8-2
Overview 8-2
Accessing ProteinLynx Setup 8-3
Getting Started 8-3
Specifying Setup Parameters 8-4
    Instrument Type 8-4
    Combine and Mass Measure 8-5
    Peptide Selection & Centroiding 8-9
    Select Output 8-11
Database Setup 8-14

Figures

Figure 8.1 The Protein Lynx Setup dialog 8-3
Figure 8.2 Select Instrument Type dialog 8-4
Figure 8.3 Combine & Mass Measure dialog 8-5
Figure 8.4 Auto Select Parameters dialog 8-6
Figure 8.5 Select Calibration Method dialog 8-7
Figure 8.6 Monoisotopic Peak Selection dialog 8-8
Figure 8.7 Peptide Selection & Centroiding dialog 8-9
Figure 8.8 Mass Measure dialog 8-10
Figure 8.9 MaxEnt3 Advanced Parameters dialog 8-10
Figure 8.10 Select Output dialog (MSMS Survey) 8-11
Figure 8.11 Select Output dialog (Electrospray) 8-12
Figure 8.12 Select Output dialog (M@LDI) 8-13
Figure 8.13 Query Setup dialog 8-14
Figure 8.14 Exclude from search dialog 8-15

Overview

The ProteinLynx Setup program is used to develop **ProteinLynx parameter files** for use by MassLynx. These files describe how a sample will be processed and which database to search against. The ProteinLynx parameter files have **.mlp** file extensions and should be saved to the main MassLynx installation directory.

Accessing ProteinLynx Setup

Select **ProteinLynx Setup** from the **ProteinLynx** menu on the MassLynx screen. The ProteinLynx Setup dialog displays the most recently opened ProteinLynx parameter file.

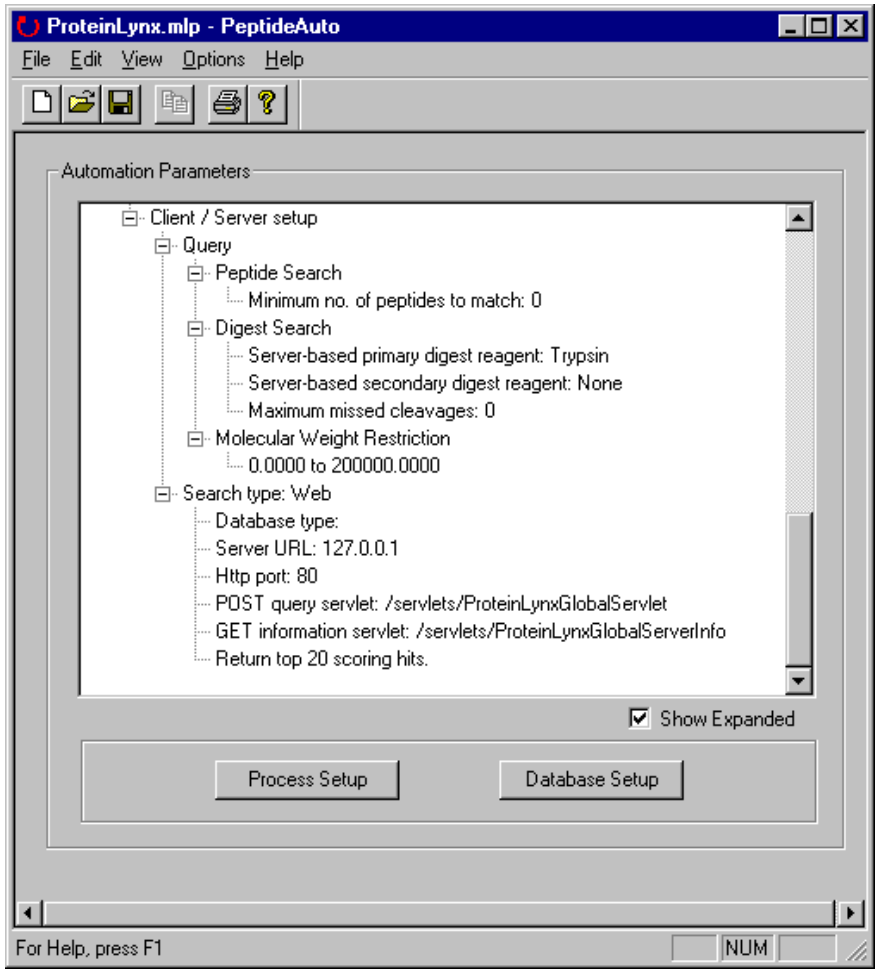


Figure 8.1 The Protein Lynx Setup dialog

Getting Started

Select **ProteinLynx, ProteinLynx Setup** from the Shortcut Bar and the MassLynx Bar respectively

To Create a ProteinLynx Parameter File

1. Press the Toolbar button, or select **New** from the ProteinLynx Setup **File** menu.

2. On each page of the ProteinLynx Setup, enter the required data, as described in this chapter.

3. Press the toolbar button, or select **Save** or **Save As** from the ProteinLynx Setup **File** menu. Enter a name for the new ProteinLynx parameter file and press **OK**.

To Open an Existing ProteinLynx Parameter File

1. Press the Toolbar button, or select **Open** from the ProteinLynx Setup **File** menu.

2. Select the required ProteinLynx file (*.mlp) and press **Open**.

Specifying Setup Parameters

Press the **Process Setup** button. This displays the **Select instrument type** dialog.

Instrument Type

Figure 8.2 Select Instrument Type dialog

Select the type of data you want to acquire by clicking M@LDI, Electrospray or MSMS Survey.

MSMS Survey is a special option to process MSMS data. Instead of building a query from MSMS data (the normal electrospray option) it processes the survey scans and outputs a list of precursor ions in a text file.

Press the **Next** button. If **M@LDI** was selected the **Combine & Mass Measure** dialog is displayed (see page 8-5). If **Electrospray** or **MSMS Survey** was selected the **Peptide Selection & Centroiding** dialog is displayed (see page 8-9).

Combine and Mass Measure

If you selected an M@LDI instrument, the Combine and Mass Measure dialog is displayed.

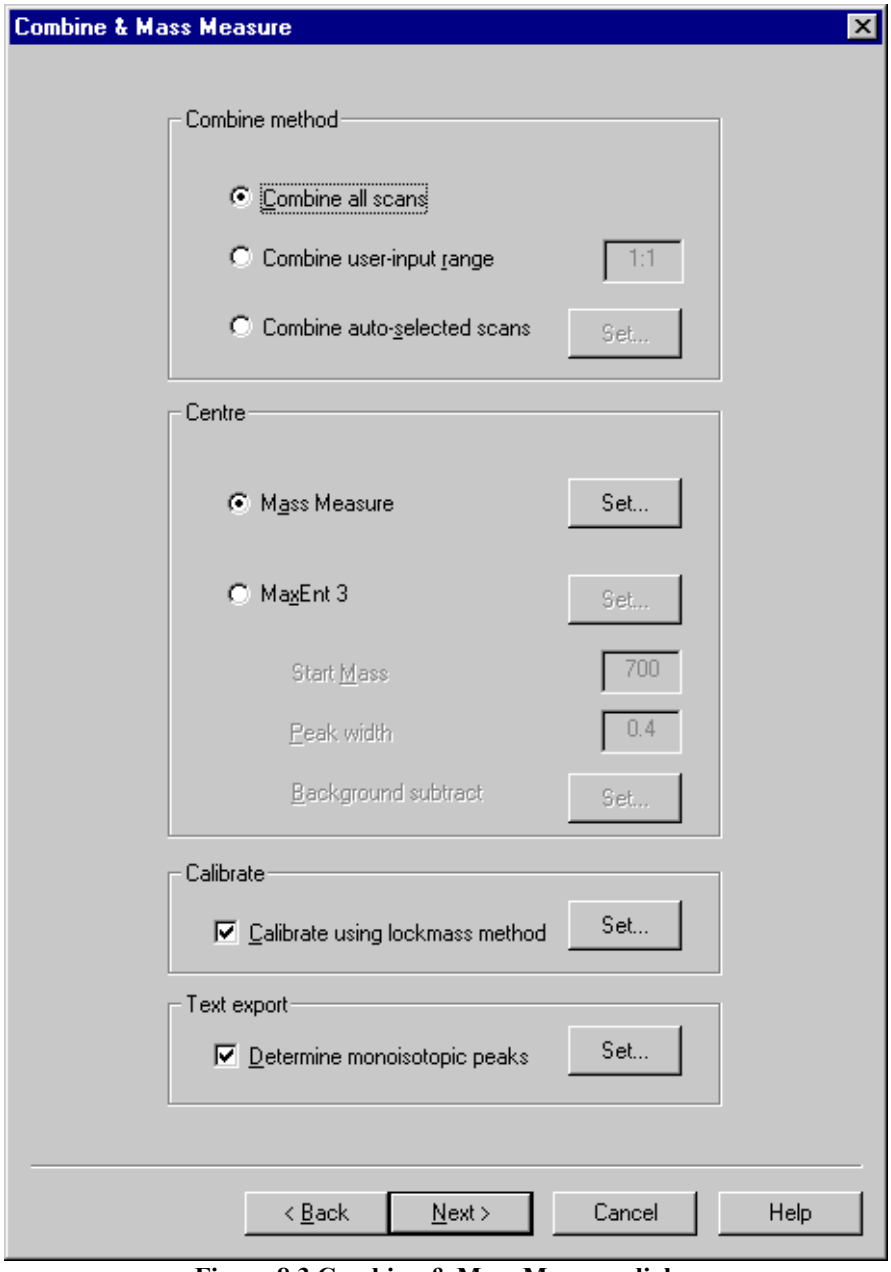


Figure 8.3 Combine & Mass Measure dialog

Combine Method

Combine all scans	Select this option to combine all scans acquired for the sample.
Combine user-input range	Select this option to combine the scans defined in the range box to the right of the field. Enter the range of scans to combine, separated by a colon (**:**).
Combine auto-selected scans	Select this option to combine the scans which pass the auto selection criteria (see Auto Select Parameters on page 8-6 for details). If this option is selected, the **Set** button is enabled. Press this button to display the **Auto Select Parameters** dialog.

Auto Select Parameters

This dialog allows the auto select parameters to be defined. Only scans that fall within these parameters will be combined.

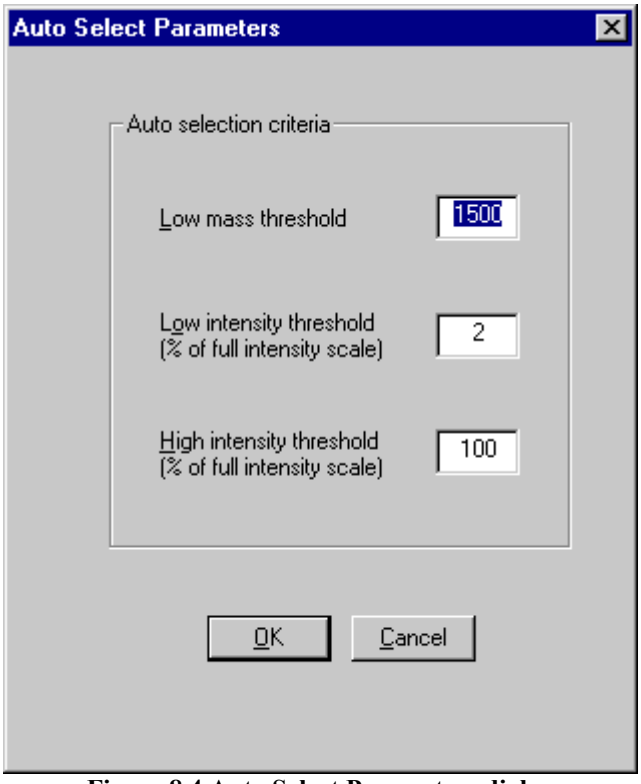


Figure 8.4 Auto Select Parameters dialog

Low mass threshold	Scans will be included for masses above the mass specified.
Low intensity threshold	Scans will be included if at least one peak is above the intensity specified.
High intensity threshold	Scans will **not** be included if any peak is above the intensity specified.

Centre

Mass Measure Select this option if you require the acquired data to be Mass Measured. Press the **Set** button to define the required mass measure parameters. For details on these settings see the Spectrum chapter of the MassLynx NT User's Guide.

MaxEnt 3 Select if MaxEnt 3 processing is required. Press the **Set** button to define the required parameters.

If a lockmass calibration is required, the calibration factor must be determined from the Mass Measured data before MaxEnt is applied, so ensure that the Mass Measure parameters are set in this case. For details on the MaxEnt 3 settings, see the Spectrum chapter of the MassLynx NT User Guide.

Calibrate

Calibrate using lockmass method Check this box to allow a calibration factor to be applied to the located peaks.

Press the **Set** button to display the calibration parameters dialog.

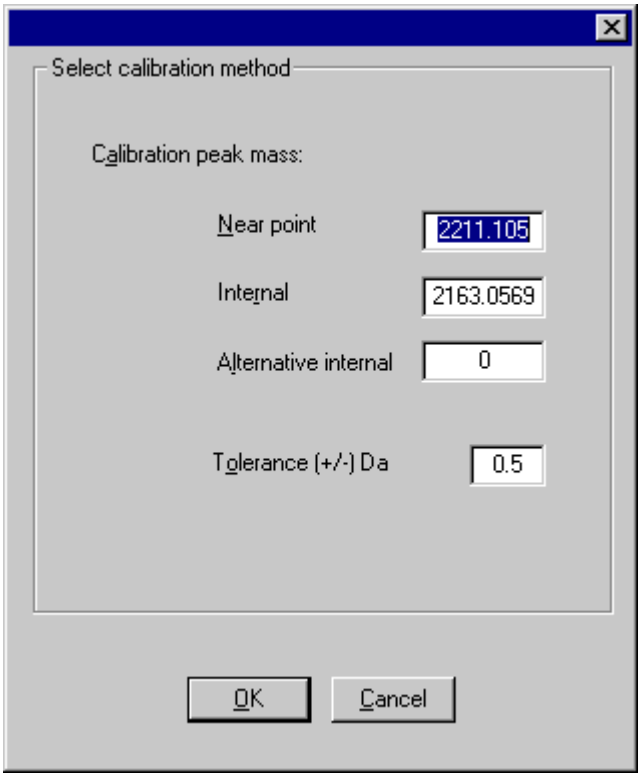


Figure 8.5 Select Calibration Method dialog

Enter the masses of known fragments in the **Near Point**, **Internal** and **Alternative Internal** boxes. The software will then search for the mass, if it is found within the **Tolerance** then any drift from the expected position will be applied to all peaks in the spectrum.

Monoisotopic Peak Selection

Determine monoispotopic peaks

Check this box to allow further definition of the peaks that will be selected for **text output** for 3rd party database searching.

Press the **Set** button to view the Monoisotopic peak selection parameters dialog.

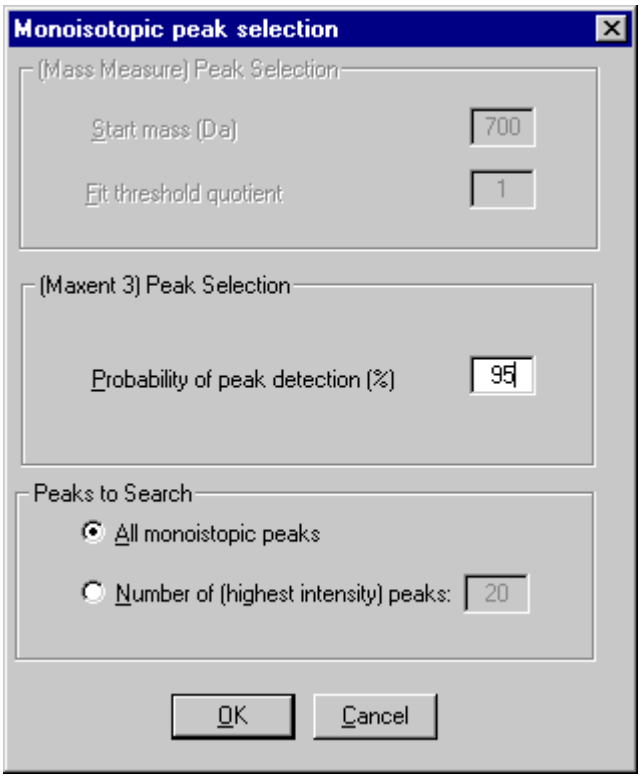


Figure 8.6 Monoisotopic Peak Selection dialog

Only one of the Peak Selection parameters will be enabled (either Mass Measure or MaxEnt 3) depending on the Center method selected on the Combine and Mass Measure dialog.

Start Mass

Specifies a mass at which to begin the search for monoisotopic peaks (default is 700 m/z) this avoids low mass peaks which may be due to matrix or contaminants.

Fit threshold quotient

This defines a divisor for the fit threshold value. The larger this value the more peaks selected. Default is 10.

For Maxent 3 processed Maldi spectrum the **Maxent 3 Peak Selection** section will be enabled and the **Peak Selection** section will be disabled. The criteria for the peak selection is based on the **Probability of peak detection** expressed in percentage. Default is 95%.

All monoisotopic peaks

Select this option to output all monoisotopic peaks.

Number of (highest intensity) peaks

Select this option to output the most intense peaks and then enter the number of peaks to include.

Peptide Selection & Centroiding

The Peptide Selection & Centroiding dialog is displayed if you selected the **Electrospray** or **MSMS Survey** option in the Instrument Type dialog (8-4).

Figure 8.7 Peptide Selection & Centroiding dialog

All MS/MS Select this option to allow all MS/MS data to be output and/or used in the database search.

Apply peptide filter Select this option to allow all MS/MS data that meets the required **QA Threshold** to be output and/or used in the database search. An algorithm is applied which attempts to determine whether the MSMS data represents a peptide. It looks at mass differences between peaks to see if any represent the mass of an amino acid residue. Values between 30 and 100 are valid.

All MS This option is automatically selected if you selected MSMS Survey in the Instrument Type dialog. All scans in the survey function will be processed, with no prior data assessment or filtering.

Mass Measure All Check this box to mass measure the complete datafile i.e. centroid the whole file not only MS/MS data.

Mass Measure Check this box if you require the acquired data to be Mass Measured. Press the **Set** button to define the required mass measure parameters.

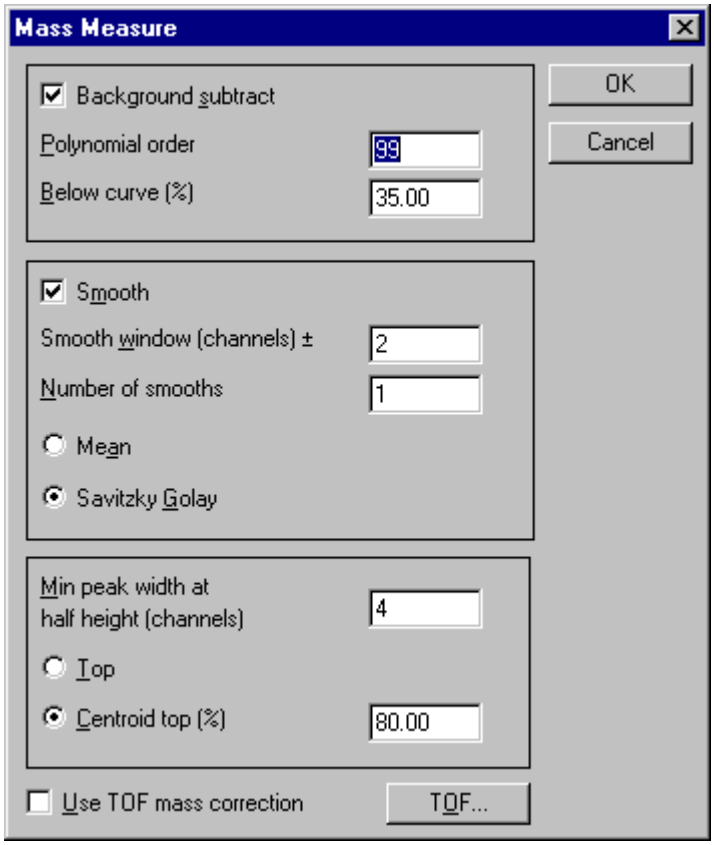


Figure 8.8 Mass Measure dialog

For details on these settings see the Spectrum chapter of the MassLynx NT User's Guide.

MaxEnt 3 Select this option if you require the acquired data to be processed using MaxEnt 3. Press the **Set** button to define the required parameters.

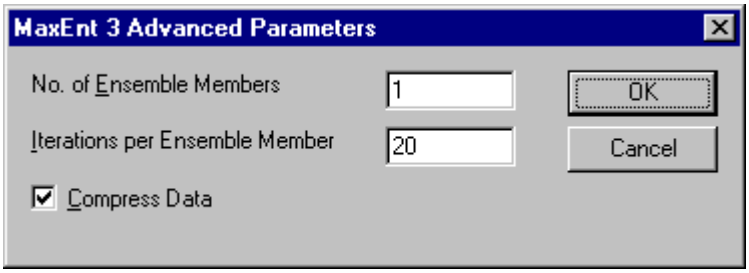


Figure 8.9 MaxEnt3 Advanced Parameters dialog

Press the Mass Measure (survey scan) Set button to define the required parameters. For further details on these settings see the Spectrum chapter of the MassLynx NT User Guide.

Press the **Next** button. The Select Output dialog is displayed.

Select Output

Two possible dialog boxes are displayed depending on the instrument selected (see page 8-4). For MSMS Survey the dialog shown in Figure 8.10 is displayed, if Electrospray is selected then Figure 8.11 is displayed and if M@LDI is selected Figure 8.12 is displayed.

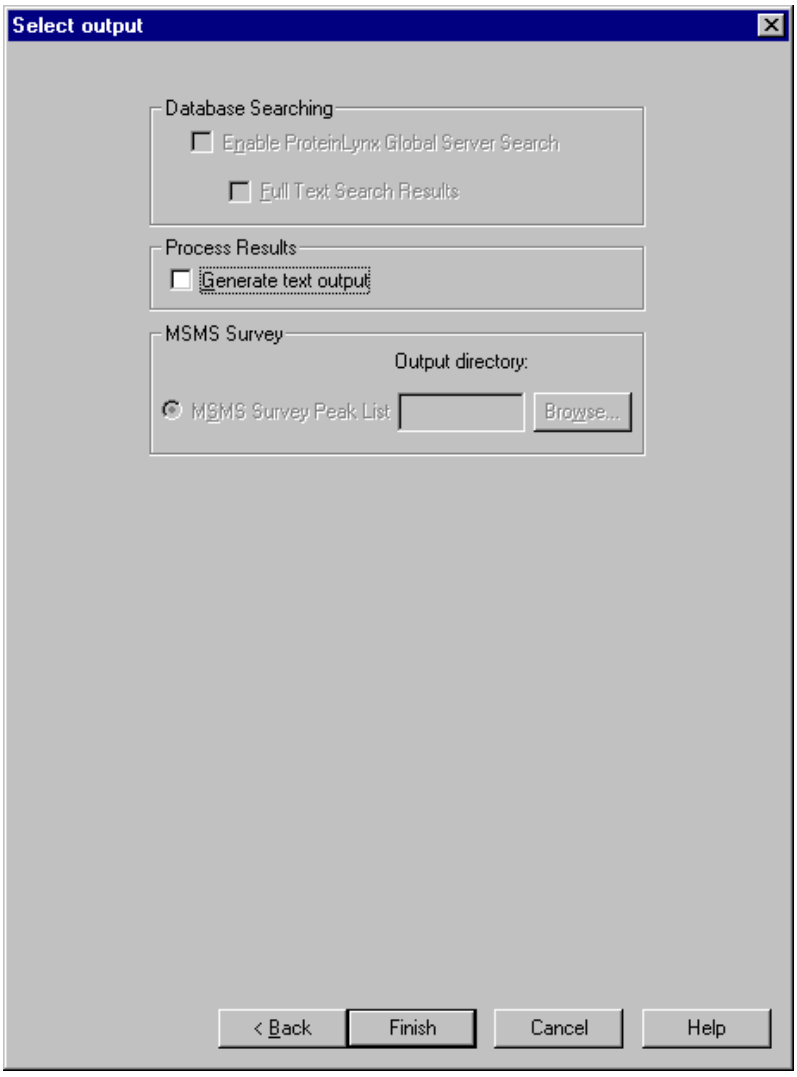


Figure 8.10 Select Output dialog (MSMS Survey)

All three dialog boxes allow output files to be written to a specified directory. If no output directory is defined then the files will be written to the Windows Temp directory.

Whichever machine has been chosen select the type of output file (**Mass:Intensity**, **Mass**, **PepSea compatible**, **Micromass Peak List** or **SEQUEST compatible**) and either type in the directory name, including the full path, or press the **Browse** button and select a directory from the dialog displayed. The default directory is the Temp directory.

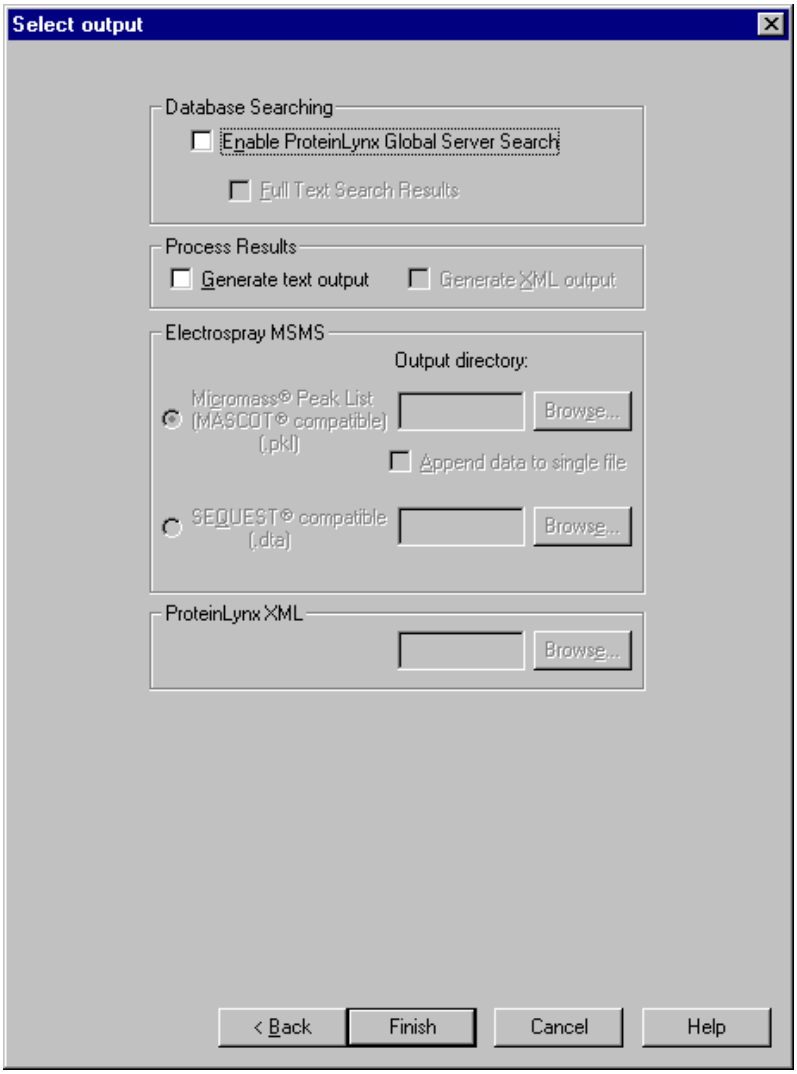


Figure 8.11 Select Output dialog (Electrospray)

Database Searching

Enable ProteinLynx Global Server Search — Check this box to carry out a database search using the peaklists or MS/MS data on a user defined FASTA formatted peptide sequence or EST database. Selecting this option will enable the **Database Setup** button on the ProteinLynx Setup dialog.

Full Text Search Results — Check this box to allow output of the database search results to an ascii/text file for incorporation in the BioRad gel-imaging software suite.

Process Results

Generate Text output — Check this box to create a text file containing the search output data.

Generate XML output — Check this box to generate an XML file containing the search output data (M@LDI and Electrospray only). The destination of this file is chosen in the **ProteinLynx XML** edit box.

MSMS Survey / Electrospray MSMS / M@LDI

The options available in these group boxes become available **Generate text output** box is checked. Enter a name and location for the search output data, or press the **Browse** button and select a directory from the dialog displayed.

ProteinLynx XML

Only available on the M@LDI and Electrospray Select Output dialogs, this option becomes available when the **Generate XML output** check box is selected.

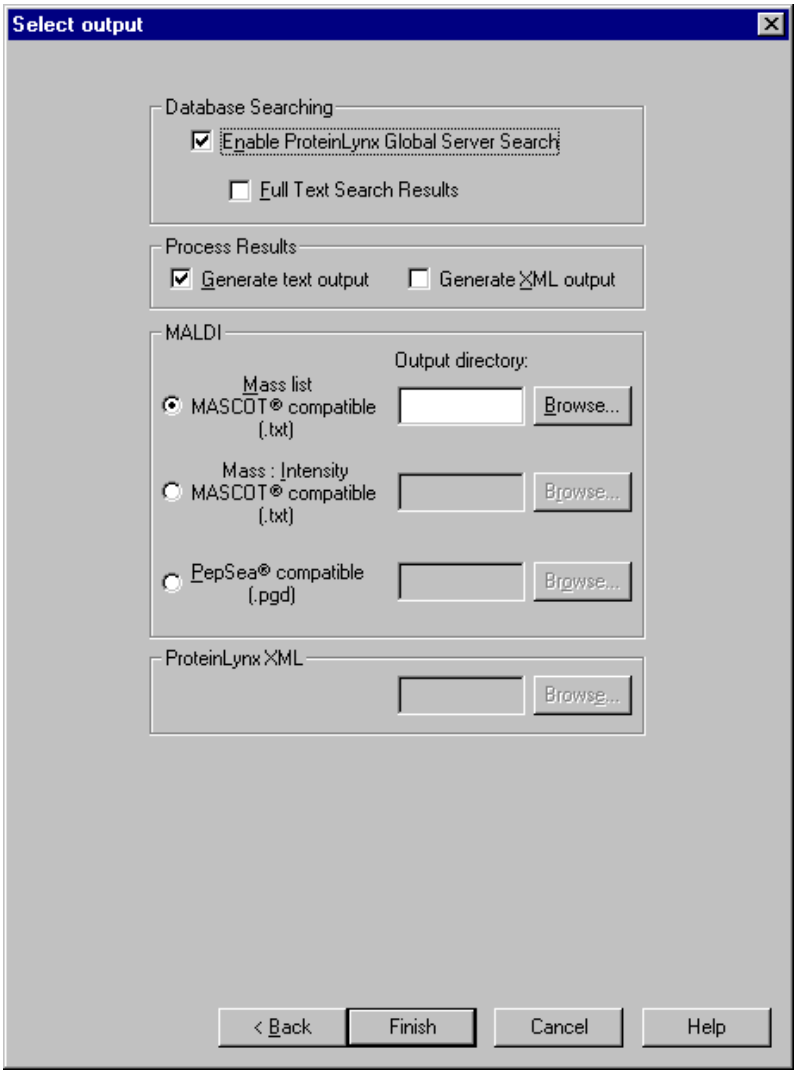


Figure 8.12 Select Output dialog (M@LDI)

Press the **Finish** button.

Press the **Back** button on any dialog to display the previous dialog.

Database Setup

Database Connection

If ProteinLynx Global Server and MassLynx are installed on the same PC, the default local host Web Search Server URL: 127.0.0.1 is applicable.

If ProteinLynx Global Server and MassLynx are not installed on the same PC, you may need to amend the URL of the Web Search Server. From the ProteinLynx Setup dialog, select **Web Server** from the **Options** menu and enter the correct **Web Search Server URL**.

> To view the Database Setup page, you must have an Internet connection and Internet Explorer version 5, or higher, must be installed.

Database Setup Page

If **Enable ProteinLynx Global Server Search** is selected on the **Process Setup**, **Select output** page, then the **Database Setup** button becomes enabled. Press this button to define the database to search against.

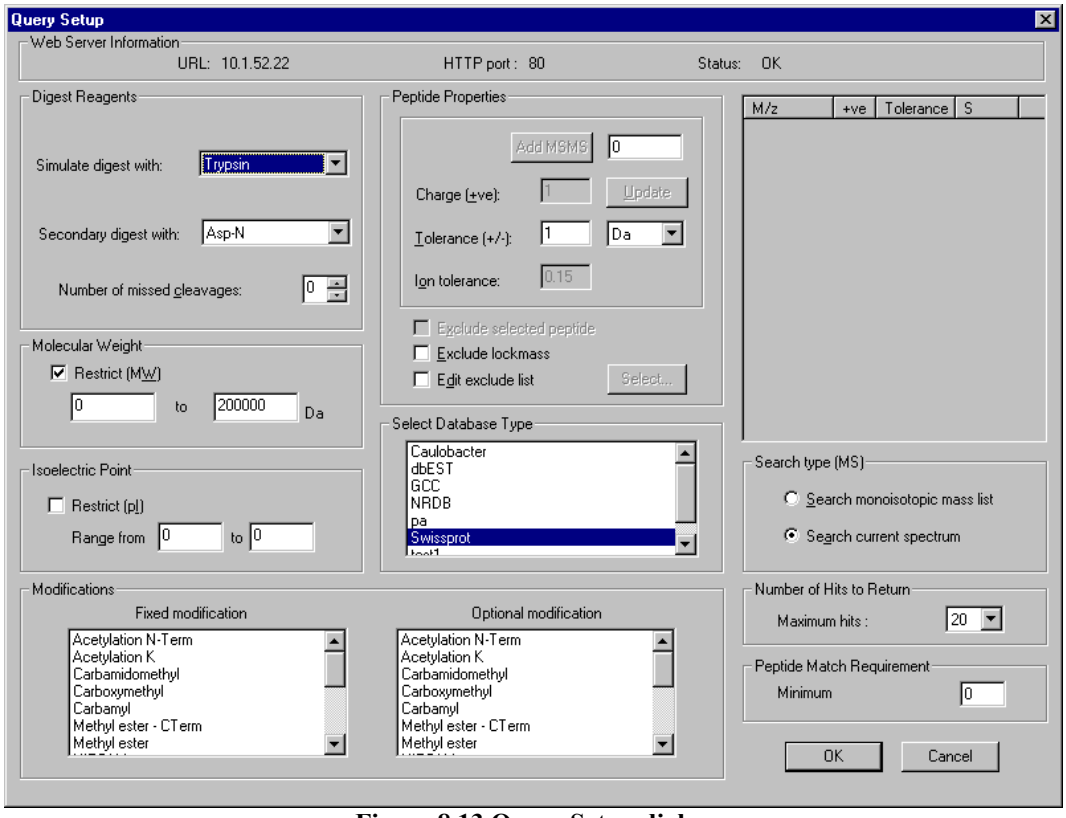


Figure 8.13 Query Setup dialog

Digest Reagents

Simulate digest with — If the peptide masses entered are the result of an enzymatic or chemical cleavage, then select an enzyme from the drop down list box.

Secondary digest with — Select a secondary digest reagent from the drop down list box, if applicable.

Number of missed cleavages

Press the arrows to take into consideration missed cleavage sites. A maximum of 2 missed cleavage sites is currently allowed. If you do not want to consider missed cleavage sites then set this to 0. Computing all possible nearest neighbor partial fragments adds significantly to the number of peptides entered in the database (by a factor of two). The major effect of this is to increase the background score by increasing the number of random molecular weight matches, which can significantly reduce discrimination.

Peptide Properties

Add MSMS

This is only available when setting up queries manually (i.e. from ProteinProbe). See the Database Searching chapter for details.

Charge (+ve)

This option is enabled for MSMS queries only. Enter a charge state and press the **Update** button to apply the charge state to a peptide mass peak.

Tolerance (+/-)

Enter an allowed mass error and select **%**, **ppm** or **Da** from the drop down list box.

Ion tol (Da)

This option is enabled for MSMS queries only. Enter the tolerance of the ions in a fragmented peptide daughter spectrum.

Exclude selected peptide

Check this box to exclude from the search the peptides that are defined in the Exclude from search dialog (see below).

Exclude lockmass

Check this box to exclude from the search any lockmass peaks found in the data.

Edit exclude list

Check this box and press the **Select** button to display the **Exclude from search** dialog.

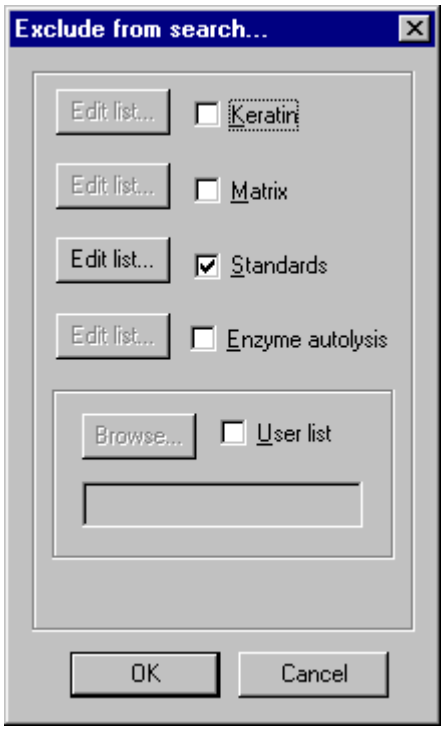


Figure 8.14 Exclude from search dialog

Check the box next to the relevant peptide group and press the associated **Edit list** button. Check the box(es) on the dialogs displayed to exclude specific peptide masses from the search.

Molecular Weight

Restrict (MW) Check this box if there are restrictions on the molecular weight of the peptide and enter a Range from and to in the relevant boxes.

Isoelectric Point

Restrict If there are restrictions on the Isoelectric Point, check this box and enter values in the **Range from** and **to** boxes.

Select Database Type

Select Database Type A list of the databases available on the server is displayed. Click on the database(s) that you want to search.

Modifications

Modifications can be set up for individual or groups of residues. For example, you can specify that the serine, threonine and tyrosine residues are phosphorylated. This modification will then be taken into account during the search.

Fixed modification If there are modifications to the residues, click on the modification(s) that you would like to be considered during the search. Selected modifications are highlighted.

Optional modification If there are suspected modifications to the residues, click on the modifications that you would like to be considered during the search, in addition to searching for the normal mass. Selected modifications are highlighted.

Search type (MS)

Search monoisotopic mass list Check to perform a search on the monisotopic mass list.

Search current Spectrum Check to perform a search on the current spectrum.

Number of Hits to Return

Maximum hits From the drop down list box, select the maximum number of hits to display.

Peptide Match Requirement

Minimum Enter the minimum number of peptides that a protein must contain in order to be returned as a 'hit'.

Chapter 9 ProteinLynx - Results

Contents

Creating Report Files 9-3
Accessing ProteinLynx Results 9-3
The ProteinLynx Workspace Screen 9-3
    The Plate Pane 9-4
The Workspace Toolbar 9-5
Getting Started 9-6
Menu Commands 9-7
    Workspace Menu 9-7
Workspace Properties 9-7
    Properties - General Page 9-7
    Properties - Default Plate Page 9-8
    Properties - Colors PagePage 9-9
    Properties - Report Options Page 9-10
    Properties - Sample and Results Pages 9-11

Figures

Figure 9.2 The Proteinlynx Workspace 9-3
Figure 9.3 The Plate Pane 9-4
Figure 9.4 The Floating Plate pane 9-5
Figure 9.5 The Workspace Print Control dialog 9-6
Figure 9.6 The Workspace Properties General page 9-8
Figure 9.7 The Workspace Properties Default Plate page 9-8
Figure 9.8 The Workspace Properties Colors page 9-10
Figure 9.9 The Workspace Properties Report Options page 9-10
Figure 9.10 The Workspace Properties Sample tab 9-11

Overview

ProteinLynx Results are viewed using the Workspace within the ProteinProbe it is therefore recommended that you read the ProteinProbe Interface section of the Database Searching chapter. The Workspace is used to view ProteinLynx Report files, print results or export results to other applications (such as LIMS systems or Excel) via the clipboard or TAB delimited text files.

The ProteinLynx Results can only be viewed on computers that have MassLynx with the BioLynx or ProteinLynx option installed.

Creating Report Files

A ProteinLynx Report file (*.mrp) is created when a job submitted from MassLynx is processed. Report files are created by specifying PeptideAuto in the **Process** column of a Sample List and the **ProteinLynx Setup file (*.mlp)** in the **Process Parameter** column.

One ProteinProbe file (*.prp) is created for each sample in the sample list and one ProteinLynx Report file (*.mrp) is created for each sample list. If a sample list is reprocessed then the *.prp and *.mrp files are overwritten.

Sample List file names in different sample lists (within the same project) should be unique as the *.prp file will be overwritten.

Accessing ProteinLynx Results

To access the ProteinLynx Results select **ProteinLynx Results** from the **ProteinLynx** Shortcut Bar of the main MassLynx screen or select **File, Open Workspace** from the ProteinProbe menu and select the required file from the dialog displayed. When **ProteinLynx Results** is selected the last processed *.mrp file within the project is opened automatically.

The ProteinLynx Workspace Screen

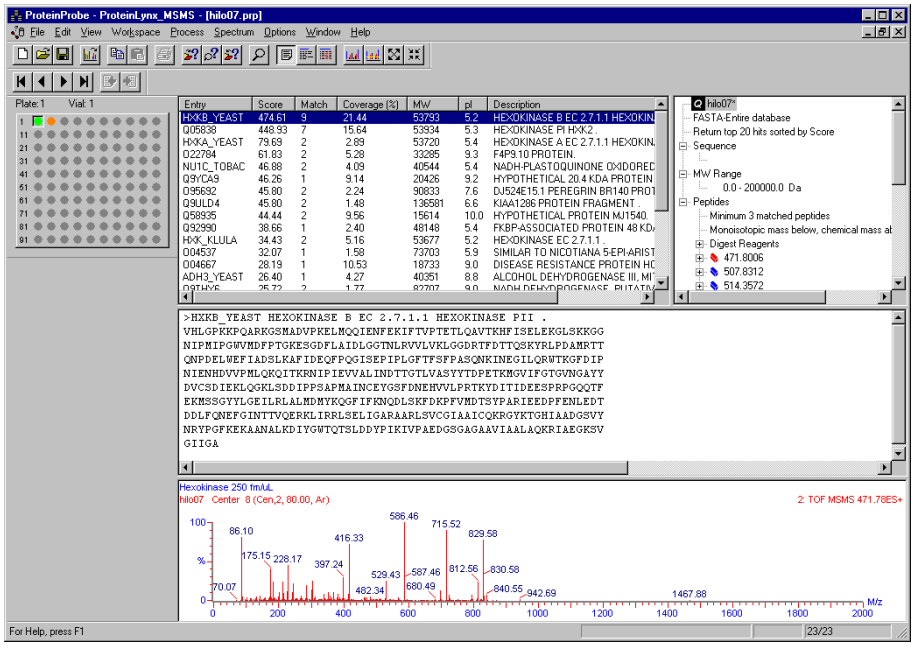


Figure 9.1 The Proteinlynx Workspace

The screen is split into 5 panes (Figure 9.1).

- The Plate pane showing the plate layout and proteins found.
- The Results List pane showing a summarized list of database entries of proteins which match the sample.
- The Query pane showing the current query.
- The Detailed Results pane showing the result highlighted in the Results List pane in detail.
- The Spectrum pane showing a centered spectrum of the raw acquired data for the well highlighted.

Apart from the Plate pane the functionality of the other four panes is the same as that described in the ProteinProbe Interface section of the Database Searching chapter.

The Plate Pane

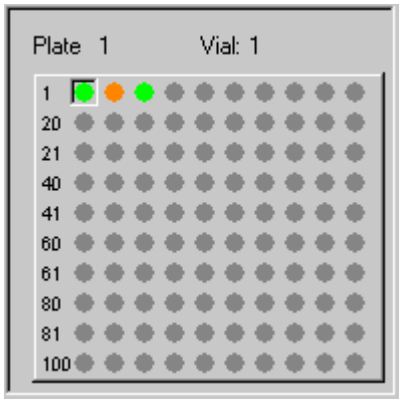


Figure 9.2 The Plate Pane

The Plate Pane (Figure 9.2) displays information about all the samples from the currently selected plate. The vials are color coded to indicate:

- The sample contains a protein that is a good match with one or more entries in the database (default is light green).
- The sample contains a protein that is a possible match with one or more entries in the database (default is orange).
- The sample does not contain a protein that matches any entry in the database (default is red).
- No sample in this position (default is gray).

The criteria that determine whether the sample is a good or a possible match is based on the score. A threshold score can be defined and if the score value for a sample is above this threshold then that sample is coloured green, i.e. a good match. If the value for a sample is below the threshold then the sample is coloured orange, i.e. a possible match.

The score threshold value and the default colors can be changed see Properties - Colors on page 9-9.

The currently selected vial is shown recessed. Change the current vial by clicking on a new vial or using the vial selection toolbar buttons (, , and).

Selecting a new vial will display detailed information about that vial in the other panes of the screen.

The appearance of a plate can be modified enabling the number of rows and columns, reference labels and vial properties to be set. See Properties - Default Plate on page 9-8 for more details.

To Reposition the Plate Pane

The Plate pane can be docked against one of the sides of the Main Window, or can be left floating (Figure 9.3)

To dock a floating plate pane or to change the location of a docked plate pane, click on the pane in the background area (i.e. not on the plate) and drag it until it touches the edge of the main window. An outline of the pane will appear to aid in positioning it. Releasing the mouse will dock the pane.

To float a docked pane, click and drag it off the main window using the mouse. Move it away from the edge of the Main Window and release the mouse. The pane will float in a “mini-window” frame, as shown below.

Alternatively double clicking on the background area will dock an undocked pane and undock a docked pane.

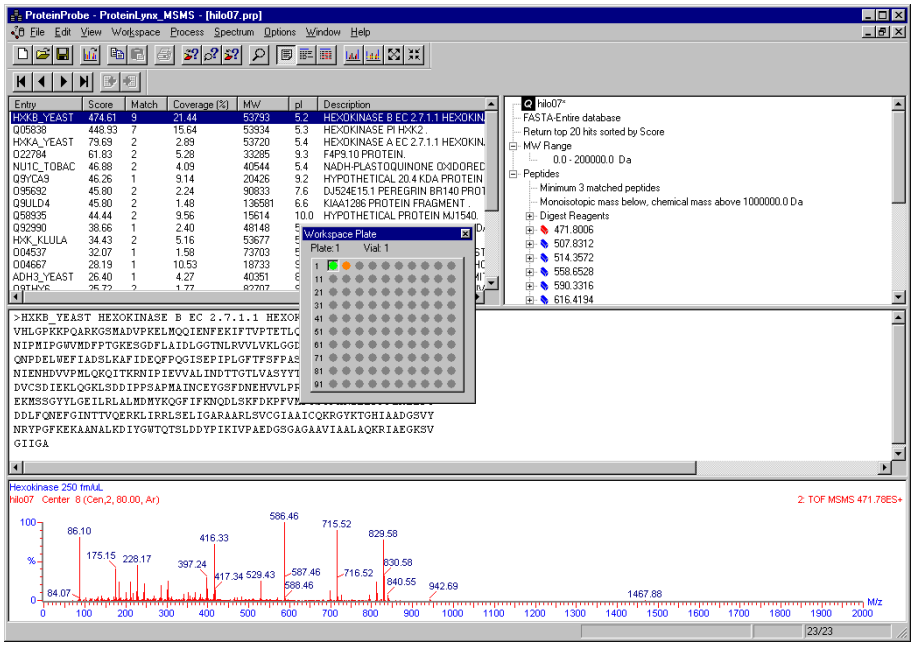


Figure 9.3 The Floating Plate pane

The Workspace Toolbar

The following toolbar buttons are only available when the Workspace is opened. The remaining toolbar buttons are described in The ProteinProbe Interface section of the Database Searching chapter.

Toolbar button	Purpose
	Go to first vial

Go to previous vial

Go to next vial

Go to last vial

Go to next plate

Go to previous plate

Getting Started

To Open an Existing ProteinLynx Report File

1. Select **Open Workspace** from the **File** menu.

2. Select the required ProteinLynx Results file (*.mrp) and press **Open**.

To Print a ProteinLynx Results File

Press the Toolbar button, or select **Print** from the **File** menu. This will display the Print control dialog as described in the ProteinProbe Interface section of the Database Searching chapter. This allows printing of individual samples in the style selected in the Report Options style required.

-or-

Select **Print** from the **Workspace** menu, this will allow printing of samples as defined on the **Workspace**, **Properties**, Report Options page. The **Workspace Print Control** dialog will be displayed.

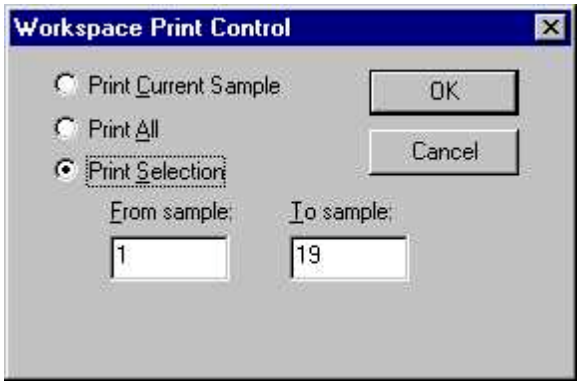


Figure 9.4 The Workspace Print Control dialog

Print Current Sample Select this option to print information for the currently selected well.

Print All Select this option to print information for all wells.

Print Selection Select this option and enter a range of samples to include in the printed Results File.

When you press **OK** the standard Windows print control dialog is displayed. This allows the selection of printers and printer properties, press **OK** again to print the report.

To Change the Size of the Panes

Position the mouse pointer on the line between the two panes until a symbol appears. Hold down the mouse and drag until the pane is the required size.

Menu Commands

The toolbar provides shortcuts for most menu items, and these have been described above. There are a few additional menu items that have no toolbar shortcuts. The following menu items are available only in the Workspace. The purpose of all the remaining menu items is described in the ProteinProbe Interface section of the Database Searching chapter.

Workspace Menu

Plate Select this option to display the plate pane. When selected a tick mark appears next to it. To hide the plate pane select the menu item a second time.

Toolbar Select this option to display the toolbar. When selected a tick mark appears next to it. To hide the toolbar select the menu item a second time.

Refresh The ProteinLynx Results file is updated after each sample is processed and can be viewed as they are being created. Select Refresh to re-read the *.mrp file and update the Workspace with the new details.

Print Invokes the Workspace print control dialog to permit selection of samples to be printed.

Print Preview Displays a preview of the printed report for the current document, according to the printing options. The display is as close as possible to what will actually appear on paper.

Workspace Properties

To display the Workspace Properties dialog (Figure 9.5), select **Properties** from the Workspace Menu.

Properties - General Page

Replace current sample on selection: When you select a new sample (using the plate pane or toolbar buttons) the current sample document is replaced by the selected sample document.

If this box is not checked then a new ProteinProbe document is opened when a new sample is selected. These documents can be viewed simultaneously by selecting Cascade or one of the Tile options from the Window menu.

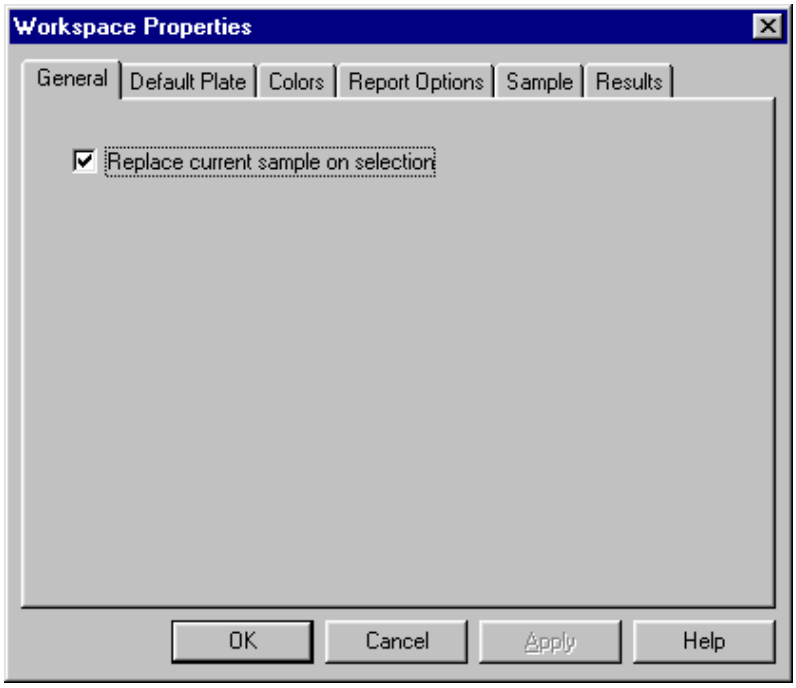


Figure 9.5 The Workspace Properties General page

Properties - Default Plate Page

For autosamplers controlled by the MassLynx software a plate layout will have been defined (see the MassLynx Guide to Inlet Control). This cannot be changed.

For autosamplers not controlled by the MassLynx software, this page is used to organize the data acquired so that it can be displayed in plate format.

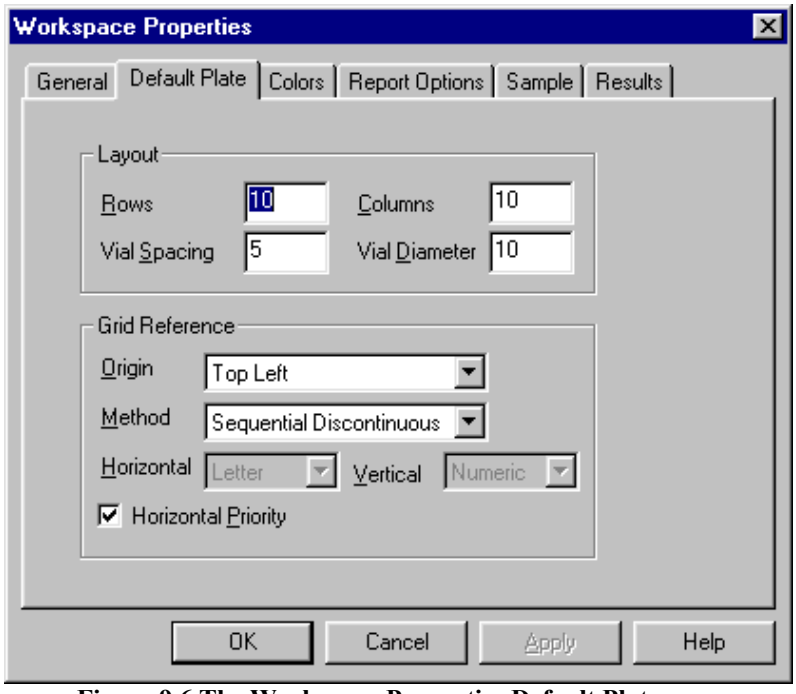


Figure 9.6 The Workspace Properties Default Plate page

Rows Enter the number of rows on the plate to be used.

Columns Enter the number of columns on the plate to be used.

Vial Spacing or Vial Diameter These values are used to define the distance between and the size of the vials on the plate pane display..

Origin This is the corner of the rack from which the vial grid referencing starts. Select one of Top Right, Top Left, Bottom Right or Bottom Left from the drop down list box.

Method This is the method of numbering the wells on a plate. There are three options:

XY which references the vials A1, B1 etc.

Sequential Discontinuous which numbers the vials 1, 2, 3 across a row, left to right if origin is top left, and then starts the next row from the left again.

Sequential Continuous which numbers the vials 1, 2, 3 across a row, left to right if origin is top left, then continues number the next row, right to left etc.

Horizontal If the method chosen is X,Y then this box becomes enabled. It allows horizontal referencing of the plate to be a number or a letter.

Vertical If the **Method** chosen is X,Y then this box becomes enabled. It allows vertical referencing of the plate to be a number or a letter.

Horizontal Priority Check this box if samples are to be acquired horizontally across the plate.

If Referencing = X,Y, Horizontal = Letter, Vertical = Number and Horizontal Priority is checked, this will result in samples being acquired in the order A1, A2, A3. If the Horizontal Priority box is not checked samples will be acquired in the order 1A, 1B, 1C etc.

If Referencing = sequential continuous or discontinuous and Horizontal Priority is checked, this will result in samples being acquired from row 1 then row 2. If the Horizontal Priority box is not checked samples will be acquired from column 1, then column 2 etc.

Properties - Colors Page

The colors page is shown in Figure 9.7.

Score Threshold This refers to the **value** shown in the Score column on the Hit List Results Pane. Enter a value above which the unknown protein is considered a good match to the database entry.

Select a color from the drop down list box for each of the categories. These will be used to show whether a good, possible or no protein match has been found at this position on the plate pane.

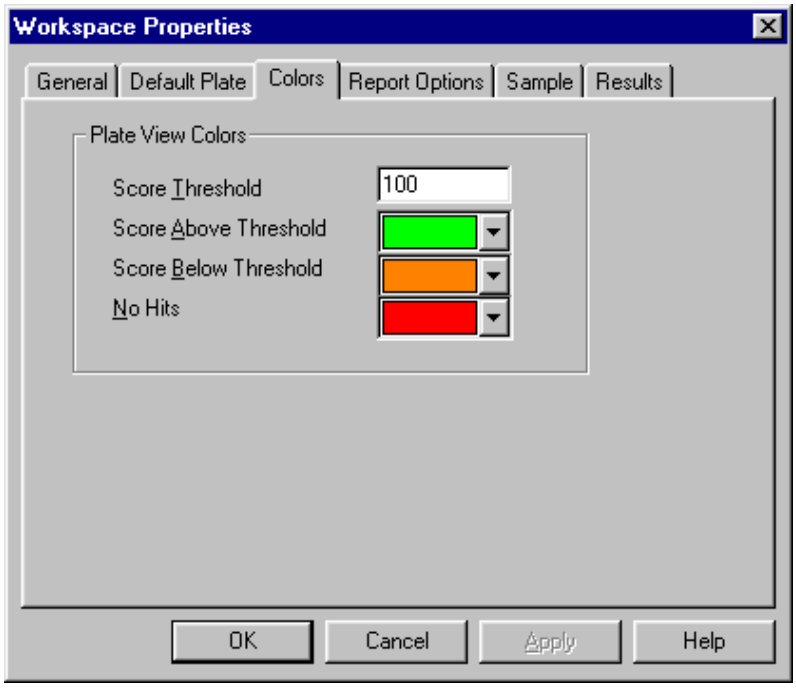


Figure 9.7 The Workspace Properties Colors page

Properties - Report Options Page

This dialog is used to select the types of printed reports and their appearance. Report options are stored in the document as well as in the Windows Registry.

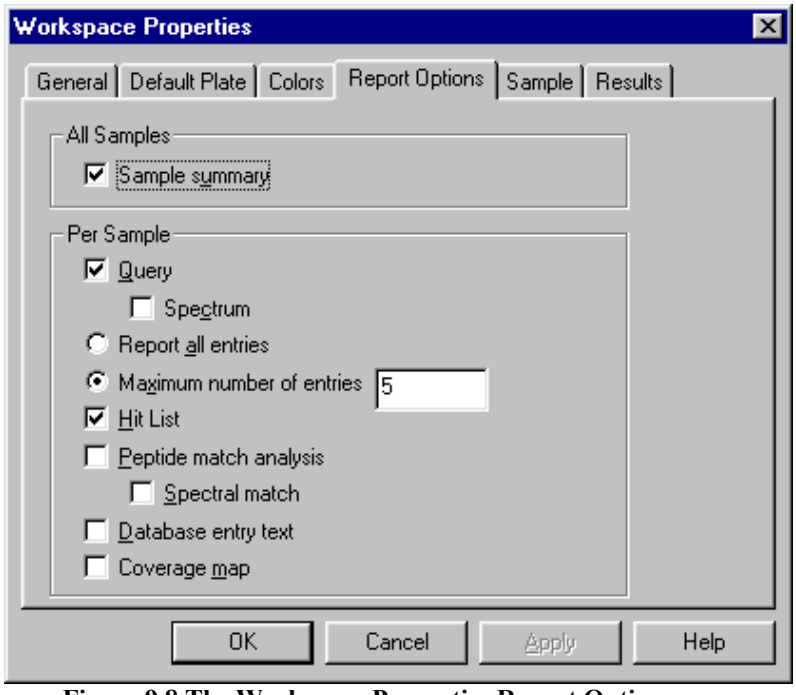


Figure 9.8 The Workspace Properties Report Options page

Sample Summary Check this box to print the information defined on the **Sample** and **Results** tabs, described below.

Query	Check this box to print a summarised query report similar to that shown in the Query pane. Check the **Spectrum** box to print the query spectra. **Note**: If you have altered the display range of the spectrum in the Spectrum Pane then this new range will be used for all spectra printed.
Report all entries	Select this option to print all the returned hit results for each sample as displayed in the Result List pane. Alternatively select the **Maximum number of entries** option and enter the maximum number to print for each sample.
Hit List, Peptide match analysis, Database entry text and Coverage map	Check these boxes to print the details as they appear in the corresponding ProteinProbe panes. If the **Spectral match** box is checked the corresponding matching MSMS spectrum will be printed. **Note:** If you have altered the display range of the spectrum in the Spectrum Pane then this new range will be used for all spectra printed.

Properties - Sample and Results Pages

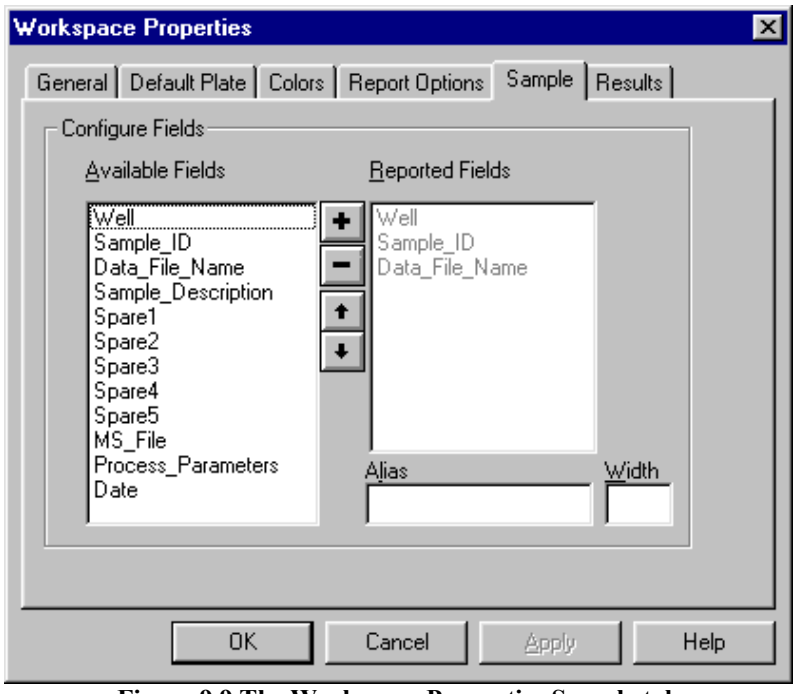


Figure 9.9 The Workspace Properties Sample tab

The basic format of the Sample page and Results page are the same, each has a list of fields which can appear as column headings on the summary report. The text that appears in the column heading and the width allocated for that column can be changed in the **Alias** and **Width** edit box. The fields in the **Sample** tab are printed first followed by the fields in the **Result** tab.

To Add a Field

Click on a field in the **Available Fields** box and press the button. The new field is added to the bottom of the list.

To Remove a Field

Click on the field to remove, in the **Available Fields** box and press the button. The field will be removed from the list.

To Change the Order of Fields

Click on the field to move, in the **Available Fields** box and press the or button until the field is in the required position.

Index

Access
  Chain editor, 1-9
Add
  Disulfide bridge, 1-7
  Field, 9-11
  Monosaccharide, 4-8, 4-9
  New Modifier, 4-9
  Peak Mass, 3-10
  Terminal Group, 4-9
  text, 1-11
Amino Acid Composition, 1-10
Amino Acid Residue
  Mass Differences, 1-41
Auto Select Parameters, 8-6
Base combination parameters, 5-12
Base Composition, 5-12
beta lactoglobulin, 1-26
Calculating Compositions, 4-7
Calculations, 5-12
Calibrate, 8-7
  using lock mass, 8-7
calibration parameters, 8-7
CarboTools, 4-1
CarboTools
  Calculating Compostions, 4-7
  Configuration, 4-5
  Deleting Results, 4-10
  Load Spectrum, 4-6
  Manually Identifying Residues, 4-9
  Menu Commands, 4-4
  New Document, 4-5
  Results Display, 4-10
  Saving and Loading Files, 4-11
  Series Assignments, 4-9
  Spectrum Display Settings, 4-6
carboxy terminal, 1-13
Centre, 8-7
Centroiding, 8-9
chain editor
  PTM button, 1-15
Chain Editor, 1-9
Child Windows, 2-15
Chromatogram Annotation, 1-26
CID electrospray, 1-33
Clear
  Disilfide bridge, 1-8
  modified residue, 1-17
Clear a modified base/s, 5-10
Cleavage Parameters, 1-24
Clipboard, 5-11
Colors Page
  ProteinLynx Results, 9-9
Combine, 8-5
Combine & Mass Measure
  Centre, 8-7
Combine and Mass Measure, 8-5
Cone Voltage Fragmentation, 1-28
Consensus Sequence, 6-16, 6-17
Copy
  modified residue, 1-16
  Tag, 2-18
Copy a modified base, 5-10
Copy modified residue
  to database, 1-17
Create
  modified residue, 1-16
  terminal group, 1-14
Creating Report Files, 9-3
C-term, 1-14
CVF, 1-28
Data processing, 7-5
Database Connection, 3-4
Database Searching, 3-1
  Loading a Spectrum, 3-4
  Searching, 3-7
Database Searching, 7-6
DataBase Searching
  New Documents, 3-3
Database Setup, 3-4, 8-14
Database Setup Page, 3-4
Delete
  modified residue, 1-16
  Peak Mass, 3-10
  terminal group, 1-15
  text, 1-11
Delete a modified base, 5-10
Delete a Sequence, 2-15
Delete masses, 1-21
dialog
  Charge state parameters, 1-39
  cleavage parameters, 1-24
  elemental composition, 1-10
  isotopic labelling parameters, 2-13
  Mass Spectral Fragmentation, 1-28
  Match Mass Spectral fragments, 1-30
  Post translational modification editor, 1-15
  Report Options, 2-21
  residue search, 1-19
  Spectrum Interpretation, 1-34
  terminal group editor, 1-14
  View Options (PepSeq), 2-23
Dialog
  calibration parameters, 8-7
  Exclude from Search, 8-15
  Select Output, 8-11
Digest Reagents Page, 2-11
Digest Simulation, 1-23
Digest Targeting, 1-25, 1-26
Display Options, 1-6
Disulfide Bridge Editor, 1-7
Disulfide bridges
  Add a bridge, 1-7

View a bridge, 1-7
Edit
modifed residue, 1-16
terminal group, 1-15
Edit a modified base, 5-10
Editing masses, 1-20
Editing Text, 1-11
Nuclic Acids, 5-6
Electrospray, 8-5
elemental composition, 1-10
Entry Report Pane, 3-9
Exclude from search dialog, 3-6, 8-15
Extend Tag, 2-17
FASTA, 3-3
Field
Change Order, 9-12
Remove, 9-11
Find Residue Mass, 1-22
Find Tag, 2-16
Fragmentation overview, 1-33
Full Text Search Files, 7-6
Global Server, 7-3
Hydropathic Profiles, 1-36
Insert a Sequence, 2-15
Installing ProteinLynx, 7-6
intra-/intermolecular disulfide bridges, 1-7
Ion Display Setup, 1-29
ion series
customizing, 1-32
Isoelectric Point, 8-16
Isotopic Labelling, 2-13
M@LDI, 8-5
MASCOT, 7-4
Mass Measure, 8-5, 8-7, 8-10
Mass Search, 1-19, 1-21
Mass Search parameters, 1-20
mass spectral fragmentation, 1-31
Mass Spectral Fragments, 1-28
Masses
Deleting, 1-21
Editing, 1-20
Entering, 1-20
MaxEnt 3, 8-7
Modifications, 8-16
Modifications Page, 2-9
modified base
clear, 5-10
copy, 5-10
create, 5-9
delete, 5-10
edit, 5-10
Modified Bases, 5-9
modified residue
clear, 1-17
copy, 1-16
create, 1-16
delete, 1-16
edit, 1-16
Modify
terminal group, 1-14
Modify a Sequence, 2-15
MS/MS, 8-9
MSMS Survey, 8-5
New
Modifier, 4-9
Monosaccharide, 4-8, 4-9
Terminal Group, 4-9
N-term, 1-14
Nucleic Acids, 5-1
Olgonucleotide Sequencing
Spectrum, 6-9
Oligonucleotide Fragmentation, 6-3
Oligonucleotide Sequencer
Accessing, 6-5
New Document, 6-5
Specifying the Spectrum, 6-5
Oligonucleotide Sequencing, 6-1
Documents, 6-11
Interface, 6-11
Parameters, 6-6
Report Options, 6-20
Sequence Report Pane, 6-13
Sequence Toolbar, 6-15
Series Ions, 6-8
Spectrum Manipulation Toolbar, 6-15
Spectrum Paranmeters, 6-8
Text Report Toolbar, 6-16
Toolbar, 6-13
Troubleshooting, 6-9
Output File, 7-4
Panes, 2-15
peptide filter, 8-9
peptide fragment elution, 1-27
Peptide Fragmentation, 2-5
Peptide Parameters Page, 2-7
Peptide Properties, 8-15
Peptide Selection, 8-9
Electrospray, 8-9
MSMS Survey, 8-9
Peptide Selection & Centroiding, 8-9
Peptide Sequence
Delete, 2-15
Insert, 2-15
Modify, 2-15
Select, 2-15
Peptide Sequencing, 2-1
Accessing, 2-6
Child Windows, 2-15
Digest Reagents, 2-11
Extend Tag, 2-17
Find Tag, 2-16
Modifications, 2-9
New Document, 2-6
New Troubleshooting, 2-12
Parameters, 2-7
Sequence Page, 2-15
Sequence Report Pane, 2-19
Spectrum Pane, 2-16

Tree Views pane, 2-18
Peptide Sequencing Documents, 2-14
Peptide Specifying the Spectrum, 2-6
Plate Pane, 9-4
Post Translational Modifications, 1-15
Print Report, 1-37
ProteinLynx, 7-1
ProteinLynx
Data Processing, 7-5
ProteinLynx Client/Server Installation, 7-8
ProteinLynx Parameter File, 8-4
Create, 8-3
Open, 8-4
ProteinLynx Report Files
Opening, 9-6
Printing, 9-6
ProteinLynx Results, 9-1
ProteinLynx Results
Accessing, 9-3
Colors Page, 9-9
Default Plate Page, 9-8
Plate Pane, 9-4
Properties, 9-7
Report Options Page, 9-10
Sample and Results Pages, 9-11
Workspace Screen, 9-3
ProteinLynx Setup, 8-1
Instrument Type, 8-4
ProteinProbe
Child Windows, 3-7
Documents, 3-7
Entry Report Pane, 3-9
Interface, 3-7
Main WIndow, 3-7
Query Pane, 3-9
Spectrum Pane, 3-10
ProteinProbe™, 3-7
Proteins, 1-1
PTM, 1-15
Query Pane, 3-9
Repeat Find Tag, 2-17
Replace
text, 1-11
Report Options, 2-21, 6-20
Nucleic Acids, 5-13
Peptides, 2-21
Proteins, 1-37
Residue Mass Combination, 1-22
Residue Search, 1-18
Residue search dialog, 1-19
Sample and Results Pages, 9-11
Scoring Algorithm, 2-5
Search type (MS), 8-16
Select
text, 1-11
Select a Sequence, 2-15
Sequence Editor, 5-4
Sequence Name, 1-18
Sequence Pane, 2-15
Sequence Parameters Page, 2-9
Sequence Prediction Details, 6-16
Sequence Report Pane, 2-19, 6-13
Sequence Source, 1-18
Sequence Toolbar, 6-15
Sequence Window
Nucleic Acids, 5-3
Proteins, 1-5
Sequencing Algorithm
Olgonucleotides, 6-3
SEQUEST, 7-4
Series Ion Assignments, 6-18
Setup Parameters, 8-4
simulated digest, 1-26
Specifying the Spectrum, 2-6
Spectrum
Analysis, 2-11
Spectrum Annotation, 1-25
Spectrum Interpretation, 1-33
Spectrum Manipulation Toolbar, 6-15
Spectrum Pane, 2-16, 3-10
Oligonucleotide, 6-12
Spectrum Parameters Page, 6-8
terminal group
modify, 1-14
Terminal Group Modifications, 1-13
terminal/linkage group
Create, 5-8
delete, 5-9
edit, 5-8
Select, 5-8
Terminal/Linkage Group Modifications, 5-7
Text Report Files, 7-5
Text Report Toolbar, 6-16
The Oligonucleotide Sequencing Interface, 6-11
Toolbar
CarboTools, 4-3
CarboTools Main Toolbar, 4-3
CarboTools Spectrum Toolbar, 4-4
Database Searching, 3-10
Nucleic Acids, 5-3
Oligonucleotide Sequencing, 6-13
Peptides, 2-19
Proteins, 1-8
Workspace, 9-5
Tools (PepSeq), 2-24
Tree Views Pane, 2-18
Troubleshooting, 2-12
Oligonucleotide Sequencing, 6-9
tryptic digest, 1-26
Undo Tag, 2-17
User Defined
Terminal Groups, 4-8
User Defined Modifiers, 4-8
User-Definable Elements, 1-17
View
Disulfide bridge, 1-7
View Options dialog, 2-23

Workspace
ProteinLynx, 9-3
Workspace Menu, 9-7
Workspace Properties, 9-7
Workspace Toolbar, 9-5
XML, 7-6
XML Report Files
Report Files, 7-6