Bioinformatics Solutions Inc.

BSi

PEAKS Studio 2.4
User Manual

For use with the PEAKS Studio
De novo sequencing
and protein identification
software

www.bioinformaticssolutions.com

Peaks Studio v2.4 User's Manual

Written by: Iain Rogers
Please contact the author for questions
or suggestions for improvement.

Introduction

Introduction to Peaks Studio v2.4

PEAKS makes the interpretation of MS/MS data much easier and much faster.

P EAKS is an innovative software system designed to derive amino acid sequences and identify proteins from tandem mass spectrometry data. After running MS/MS on a protein sample, PEAKS performs *de novo* sequencing and database search identification of the protein(s) and peptides using raw experimental data.

Peaks Studio v2.4 provides peptide sequence and protein identification results via an intuitive interface, allowing for rapid visual interpretation. PEAKS provides both auto and manual *de novo* sequencing tools for detailed examination of MS/MS spectra, providing the flexibility to manually modify auto *de novo* results when searching for additional sequence possibilities.

With Peaks Studio v2.4, we will enjoy the following features that distinguish PEAKS from other software packages:

- **Supreme Accuracy.** PEAKS correctly derives more peptide sequences from any tandem MS/MS data than any other software currently available.

- **Post-Translational Modification Detection.** The user can tell PEAKS to expect one or more post-translational modifications, or even define new ones, and PEAKS will use this information in its analysis.

- **Full Automation.** Simply point PEAKS at the instrument's data file and derive the peptide without the need for further human interaction. Depending on the system hardware, PEAKS can easily handle processing a few thousand standard spectra at a time. For help with processing more spectra, please contact technical support.

- **Convenient Output.** PEAKS Exports to both FASTA and XML formats.

- **Superior Efficiency and Speed.** PEAKS typically derives one sequence in under five seconds.

- **No Database Required.** PEAKS efficiently evaluates all possible combinations of amino acids using a global optimization algorithm. Thus, the peptide is sequenced *de novo*. No protein/DNA database is required.

- **Database comparison.** Though no database search is required, PEAKS can use its *de novo* results to efficiently search a peptide database. This yields faster results than a conventional database search and allows us to compare the two results.

- **Great Error-Tolerance.** PEAKS automatically estimates the noise level and the accuracy of the instrument that produced the data, adjusts for calibration errors, and ignores spurious peaks.

- **Instrument Optimization.** PEAKS is optimized for Q-TOF, TOF-TOF, FTMS and ion-trap instruments. PEAKS accepts a peak list in Micromass (.PKL), Sequest (.DTA), text (.txt) or Mascot Generic Format (.MGF), Peaks is also proud to support the proposed data standard: mzXML.

- **Platform Independency.** PEAKS runs in any Java 1.3.1 (or better) environment with reasonable hardware and memory requirements:

- **Comprehensive manual sequencing tools**, including algorithms for y/b ion search, C/N terminal search and generation of sequence tags.

- **Intuitive graphical interface.** Peaks Studio v2.4 shows a simple, informative interface that provides access to spectra, sequence, confidence and y/b ion information.

- **Spectra view manipulation.** Peaks Studio v2.4 provides a closer look using zooming, labeling, annotation, peak distance tools and an alignment view so we can keep track of where we are on the spectrum.

- **HTML Report Editing.** PEAKS can produce reports in HTML format, and provides easy access to spectra, fragment and sequence information for fast document creation.

How to use this user's manual

This user's manual is intended to help us get started using Peaks Studio v2.4, acquaint us with its functionality, show us how to customize PEAKS to our application, allow us to work efficiently with the interface, provide a task based reference, and help us with troubleshooting. As such, this manual is organized into chapters based on these

categories. Use the table of contents at the front of this manual to access the relevant section. If searching for the definition of a particular term (or abbreviation), please consult the glossary – found in this section. The glossary will tell us what a particular term means, but it will not tell us how it applies to PEAKS usage.

Scope

PEAKS users are assumed to be familiar with computer usage, and the operating system environment. As such, it is beyond the scope of this manual to instruct the user on the use of windows, dialogue boxes, menus, file storage etc. Please refer to the operating system's manual, or computer help books for such information. Similarly, PEAKS users are expected to be familiar with mass spectrometry, standard operating practices and data.

Terminology and Abbreviations Glossary

m/z: mass to charge ratio.

Deconvolution: rearrangement of the spectrum to show each monoisotopic peak as if it were singly charged. Thus, to reposition them on the scale, PEAKS multiplies the m/z of ion's that were doubly charged by two. Note that the deconvolved scale PEAKS shows is 'at +1.'

a-ions: an N-terminal fragment holding at least one charge; similar to b-ions and c-ions. This is a prefix fragment of the peptide. The a-ion's mass will be the sum of the masses of the N terminal group, plus the intervening neutral amino acid residues, subtract the mass of Carbon Monoxide.

b-ions: an N-terminal fragment holding at least one charge; similar to a-ions and c-ions. This is a prefix fragment of the peptide. The b-ion's mass will be the sum of the masses of the N terminal group, plus the intervening neutral amino acid residues.

c-ions: an N-terminal fragment holding at least one charge; similar to a-ions and b-ions. This is a prefix fragment of the peptide. The c-ion's mass will be the sum of the masses of the N terminal group, plus the intervening neutral amino acid residues, plus the mass of ammonia.

x-ions: a C terminal fragment holding at least one charge; similar to y-ions and z-ions. This is a suffix fragment of the peptide. The x-ion's mass will be the sum of the masses of the C terminal group, plus the intervening neutral amino acid residues, plus the mass of Carbon Monoxide.

y-ions: a C terminal fragment holding at least one charge; similar to x-ions and z-ions. This is a suffix fragment of the peptide The x-ion's mass will be the sum of the masses of the C terminal group, plus the intervening neutral amino acid residues, plus the mass of H_2.

z-ions: a C terminal fragment holding at least one charge; similar to x-ions and y-ions. This is a suffix fragment of the peptide. The z-ion's mass will be the sum of the masses of the N terminal group, plus the intervening neutral amino acid residues, subtract the mass of ammonia.

Residue: as used in this manual, a residue refers to what remains of an amino acid once it has become part of a peptide, or peptide fragment. In this manual, residues are referred to by their original amino acid names.

Resolution: refers to the precision of an instrument. On a spectrum, this is reflected by how close together can two PEAKS be and still be told apart.

Mass accuracy: this refers to the accuracy of the spectrometer and its resulting data. On a spectrum, this is reflected by how close the PEAKS are to the actual masses of the ions they represent.

ESI (Electrospray Ionization): A method for ionizing a sample into the mass spectrometer.

MALDI (Matrix-Assisted Laser Desorption/Ionization): A method for ionizing a sample into the mass spectrometer. This has a characteristic effect of only producing singly charged ions.

PTM (Post Translational Modification): A protein, just translated and hence newly formed, may differ from its final form as a result of interaction with the cellular environment (or the experimental environment). As they interact chemically with the environment, residues may gain or lose molecules. This change is referred to as a post-translational modification. Since PTM changes the mass of residues, it must be accounted for when sequencing peptides by mass spectrometry.

Built-in PTM: PEAKS comes equipped with a library of possible post-translational modifications. These can be incorporated into a *de novo* analysis at the click of a button.

Customized PTM: If the post-translational modification we are looking for is not in the PEAKS PTM set, we may create our own entry, or modify an existing one. This will appear as a customized PTM in the set.

Enzyme: The residues PEAKS can find in different positions in the sequence. This is based on information about the enzyme used to digest our protein sample.

PTM set: A listing of all possible (built-in and custom entered) post-translational modifications that PEAKS can use as a part of its analysis.

Enzyme/PTM set: Combined, the enzyme information and post-translational modification information provide PEAKS with the relevant parameters of the experiment/sample. This will be applied to the corresponding data set when PEAKS performs its *de novo* analysis. It is a required parameter.

Fixed modification: selecting a post-translational modification as a fixed modification tells PEAKS that this modification is applied to all occurrences of the residue(s) that the PTM can act on.

Variable modification: selecting a post-translational modifications as a variable modification tells PEAKS that this modification may or may not be applied to any given occurrence of the residue(s) that the PTM can act on.

FASTA (Fast-All): A standard sequence database file format used for protein identification. PEAKS can identify proteins from any FASTA-format database of proteins.

PKL: The file format associated with Micromass instruments

DTA: The file format associated with SEQUEST software

MGF: The file format associated with Mascot. software

BSI (Bioinformatics Solutions Inc.): The makers of PEAKS and other fine bioinformatics software.

ANN (Annotation): PEAKS' XML-based ANNotated spectrum file format. ANN files preserve all the information from the PEAKS session.

ANN data file: contains the MS/MS information and peptide information of one spectrum.

ANN index file: Used to organize multiple datasets, the ANN index file links to a directory containing multiple ANN data files.

Getting started with Peaks Studio v2.4

Everything we need to know from the beginning and step by step.

T his section of the manual will guide us through the process of installation and configuration of Peaks Studio v2.4. If we run into any problems we can refer to the frequently asked questions section of this manual. If problems persist, contact technical support.

What we will need

Package contents

The Peaks Studio v2.4 package should contain:

- This manual

- Peaks Studio v2.4

- Release Notes

System requirements

Peaks Studio v2.4 will run on most platforms with the following requirements:

- Equivalent or superior processing power to a Pentium at 500 MHz.

- At least 512 MB of memory (RAM). 1024MB is recommended.

- JAVA Virtual Machine 1.4 or better.

Instrumentation

Peaks Studio v2.4 will work with data from any type of tandem mass spectrometer.

It was originally built for use with Q-TOF instruments, but has been tuned to work with many other instrument types (Ion-Trap, TOF-TOF, FTMS, etc.). The more resolution and mass accuracy an instrument gives, the better the results from PEAKS Studio's *de novo* analysis.

A note on ion-trap instruments

Ion trap data is usually of lower resolution than Q-TOF data, but Peaks Studio v2.4 can still use it to sequence peptides. One simply has to configure PEAKS for use with ion trap. This setting can be set by selecting "ION Trap" as the instrument in the auto *de novo* dialogue box – discussed later.

PEAKS will accept data directly from Waters/Micromass (.pkl) instruments, Thermo-Finnegan (.dta) instruments, ABI 4700 (directly from the database). PEAKS will accept data from any other instrument provided it can be exported to .dta, .pkl, mzXML or .mgf.

Installation

If we already have PEAKS installed on our system, we should uninstall it before proceeding.

1. Close all programs that are currently running and end all non-system tasks.

2. Insert the Peaks Studio v2.4 disc into the CD-ROM drive. This is the BSI disk which lists its contents as

   - PEAKS Software

   - PEAKS Movies

   - PEAKS Tutorial

3. Auto-run should automatically load the installation software. If it does not, find the CD-ROM drive and open it to access the disc (the disc is labeled "PEAKS STUDIO V2.4"). Click on the exe file.

4. A menu screen will appear with the title "PEAKS Studio 2.4". Select "Install Peaks Studio".

5. The installation utility will begin the install. Wait while it does so. When the "PEAKS Studio v2.4" installation dialogue appears, click the "Next" button.

6. Read the license agreement. If we agree to it, we change the radio button at the bottom to select "I accept the terms of the License Agreement" and click "Next".

7. Next we choose the amount of memory we would like to run PEAKS with by clicking the button next to an amount of memory. We must select an amount that is less than (or equal to) the amount of physical memory our system possesses. Click "Next".

8. Next we choose the folder/directory in which we'd like to install Peaks Studio v2.4. Press the "Choose…" button to browse our system and make a selection, or type a folder name in the textbox. Click "Next".

9. Choose where we'd like to place icons for Peaks Studio v2.4. The default will put these icons in the programs section of our start menu. Click "Next".

10. Review the choices we have made. We can click "Previous" if we'd like to make any changes or click "Next" if those choices are correct.

11. Peaks Studio v2.4 will now install on our system. We may cancel at any time by pressing the "Cancel" button in the lower left corner.

12. When installation is complete, click "Done". The "PEAKS Studio 2.4" menu screen should still be open. You may view movies and materials from here. To access this menu again, we simply insert the disc in our CD-ROM drive.

Registering PEAKS

The first time we run PEAKS, we will be told that the product is not registered. Press the "Ok" button and a dialogue will appear. Enter the registration key that came with the product – whether it be a key for the full version or time limited trial version. We must also enter our name, the name of our organization, and the MAC address of the machine we are going to use PEAKS on. If we are connected to the internet, registration will be completed automatically. If all is well, a dialogue will show "Registration Successful" and PEAKS will load.

Re-registering PEAKS may be necessary if our license has expired or if we wish to update the license. We will need to obtain a new registration key from BSI. Once we have obtained this new key, select "Register Peaks" from the Help menu. The "License Upgrade" dialogue box will appear, cautioning us that we are about to update the license. Press the "Ok" button to continue. Now we must

Database Configuration

In addition to *de novo* sequencing of peptides, Peaks Studio v2.4 also has the ability to search through a database search to identify proteins. But in order to use this function, PEAKS must have access to a protein or EST database in FASTA format or an EST database of DNA sequences. We can point PEAKS to an existing database on our system, or download one. This is database configuration.

> We can use PEAKS without the database search; PEAKS will perform *de novo* sequencing only.

WARNING: Downloading a database can take a long time (8+ hours), depending on connection speed.
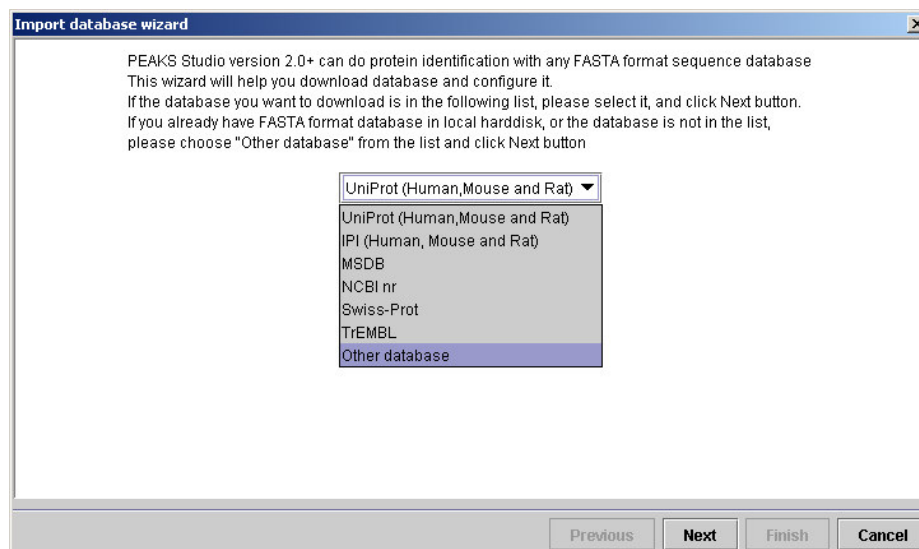
To configure a database:

1. Load Peaks Studio v2.4. If we have not yet configured a database, the wizard will appear automatically. Otherwise...

2. In the 'edit' menu, select "Configuration", then "Import Database Wizard"

3. The Import Database Wizard will load and ask us to select a database to download – from the dropdown list. If we already have a database we wish
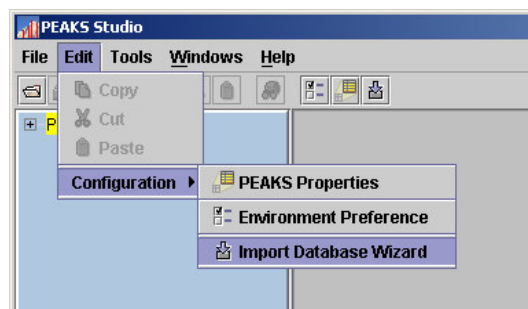
to use, we can select "Other database" from the dropdown list and skip to step6. Click "Next"

Import database wizard

PEAKS Studio version 2.0+ can do protein identification with any FASTA format sequence database
This wizard will help you download database and configure it.
If the database you want to download is in the following list, please select it, and click Next button.
If you already have FASTA format database in local harddisk, or the database is not in the list, please choose "Other database" from the list and click Next button

UniProt (Human,Mouse and Rat)
IPI (Human, Mouse and Rat)
MSDB
NCBI nr
Swiss-Prot
TrEMBL
Other database

Previous | Next | Finish | Cancel

4. Having selected a database, the Import Database Wizard will provide us with some information about that database. If this is, in fact, the database we wish to use, click the provided link to begin downloading. A dialogue box will appear with instructions on downloading using file transfer protocol (FTP). It does not matter where we put the download file, but we must remember where it is.

A note on downloading

The links in the Wizard may be outdated because the owners of those download locations may change their URL periodically. If this is the case, remove all but the domain name and browse from there.

```
ftp://ftp.ebi.ac.uk/pub/databases/MassSpecDB/msdb.fasta.z
```
becomes:
```
ftp://ftp.ebi.ac.uk/
```

5. The database we downloaded may be in a compressed file. We must find the file and use a decompression utility, such as WinZip or PK-Zip, to extract its contents. It makes sense to place the extracted file in a location that Peaks Studio v2.4 can easily access.

6. Return to the Import Database Wizard and click the "Next" button. This screen will allow us to configure the database. Click on the hyperlink next to each field for more information.

- PEAKS will ask us to enter the database nickname. This is a nickname that we chose to represent the database we are configuring. We must enter at least one character.

- The Path textbox shows where the database is located. It will be blank, so we must tell PEAKS where the database is. Type the location of the file into the textbox, or we can browse to find the file on our system We must sure to select the FASTA database, **not the compressed file of the same name**.

Database header format is important for protein ID result reports. If parsed correctly, accession numbers and protein names will be shown in full.

- If we chose one of the public standard databases (in step: 3) its format style will be displayed in the advanced options box. The selected database format is shown in the dropdown list. Accession number information and the way PEAKS parses the database headers – i.e. the parsing rules –are shown in the textboxes below.

- If we chose an "other database" (in step: 3) we must enter parsing parameters ourselves by typing in the textboxes. Alternatively, if our database format is very similar to one of the public databases, we can choose to apply that database's format when PEAKS reads our database. Select the database that is similar to ours from the dropdown list and **press the apply button** to fill the textboxes with the appropriate parsing rules.

The delimiter is the character used to separate multiple headers.

7. Press the "Finish" button to complete the database configuration.

We can repeat this process to configure a number of another databases. Once configured, a database need not be configured again unless we update the database itself.

Chapter

3

Features Walkthrough

Let's familiarize ourselves with PEAKS

T his section of the manual will walk us through most of the basic functionality of Peaks Studio v2.4. After completing this section we will have seen how easy it is to load and view a spectrum data file, perform *de novo* sequencing, and database search protein identification.

Begin the walkthrough

Download and configure the Swiss-Prot database. The procedures for doing so are outlined in the previous section.

Run Peaks Studio v2.4 and import a datafile by clicking the 🖆 icon on the toolbar in the upper left corner of the PEAKS window, or selecting "Open" from the "File" menu.

Sample data is located in the "C:\Program Files\PEAKS Studio\data\multiple spectra" folder. Load the file "Cytochrome-C.pkl" by clicking on it, then clicking "Open."

The data file will appear in the left-hand frame. Make sure "Cytochrome-C.pkl" (i.e. the data file) is selected.
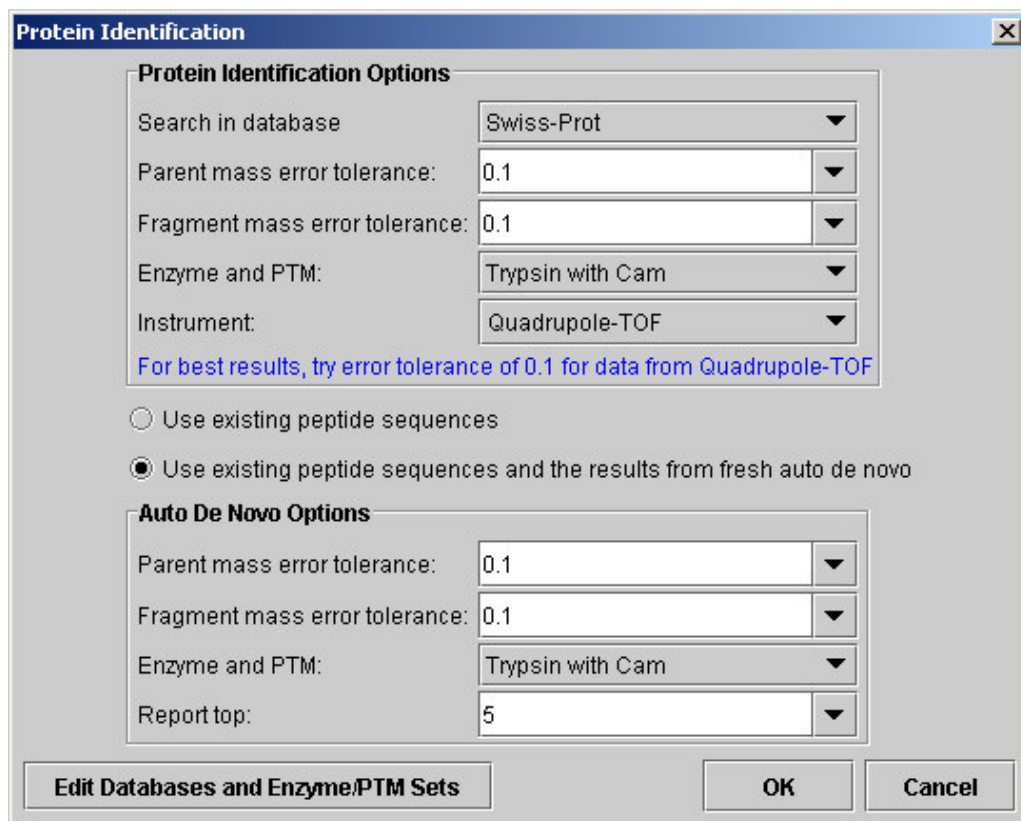
In the "Tools" menu, select "Protein Identification"
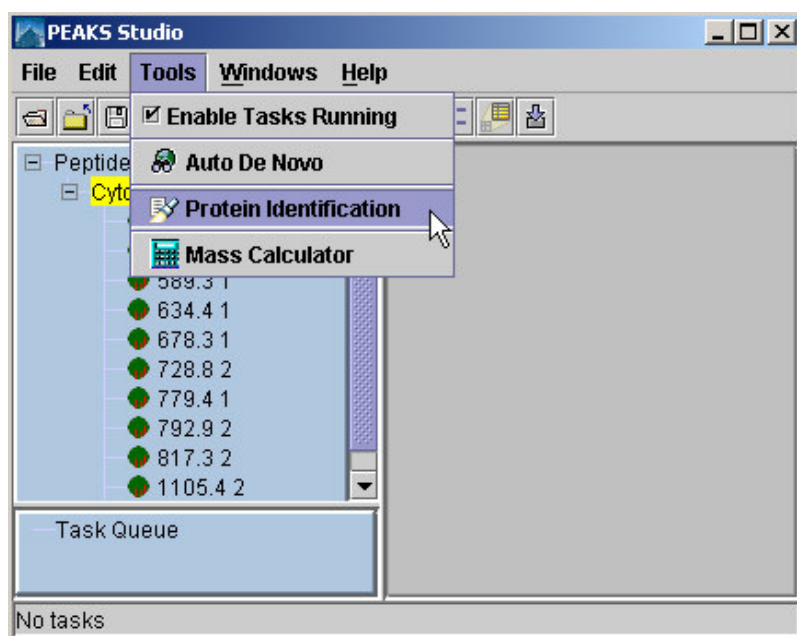
The following dialogue should appear:

Enter the settings as shown. Settings can be changed by clicking on the drop down list and selecting one of the options.

Click "OK", after entering the settings shown, to commence PEAKS auto *de novo*.

The process will take about one minute, after which PEAKS database search will proceed. In total, the process takes less than two minutes for this sample (depending on the system's processing speed and memory). Peptide sequence candidates are derived using the PEAKS auto *de novo* algorithm. With the database search functionality, these *de novo* results can then be used to identify protein candidates for consideration. The PEAKS auto *de novo* algorithm derives sequence candidates for each of the eleven spectra in our example data file. These sequence candidate results for all eleven spectra in the example are then used for the database search component of Peaks Studio V2.4.

Click "Protein ID Result" under the file name in the Peptide Data frame. Peaks Studio v2.4 presents a list of proteins that it believes to be the best match for the sample. This index lists them by accession number, ranked in descending order from highest score on downward.

The numbers under the "Mass" heading represent the mass of the protein displayed. The "Coverage" numbers represent the percentage of the proteins sequence covered by the matching peptides.

Result from database SwissProt(0.1 0.1 Trypsin with Cam)

Accession		Mass	Score	Coverage	Description
CYC_BOVIN	more	11547.01	99.22%	50%	(P00006) Cytochrome c.
SRD1_YEAST	more	25853.277	32.89%	8.44%	(P09007) Pre-rRNA processing protein SRD1.

Since CYC_BOVIN is at the top of the list it is most likely that the sample contains Cytochrome c. Since one cannot distinguish between Bovine Cytochrome c and Equine Cytochrome c (or that of some other organisms), it is not necessarily Bovine Cytochrome c; CYC_BOVIN is displayed simply because it is the first entry in the database. Clicking the **more** hyperlink next to CYC_BOVIN will show a listing of other possible Cytochrome c.

The listing, as shown above, is simply an index. To see why PEAKS chose Cytochrome c as its best possible protein candidate, click the **CYC_BOVIN** hyperlink. The report will scroll to the following:

CYC_BOVIN more... (P00006) Cytochrome c.
Mass:11547.01 **Score:**99.22% **Coverage:**50% [Show missed de novo results]

Mz	Charge	Mr(calc)	Start	End	Score	Peptide
482.7	2	945.5171	92	99	99%	EDLIAYLK
584.8	2	1149.6042	28	38	99%	TGPNLHGLFGR
634.4	1	615.3744	9	13	93.58%	IFVQK
678.3	1	659.36426	74	79	84.58%	YIPGTK
728.8	2	1437.6525	40	53	99%	TGQAPGFSYTDANK
779.4	1	760.43054	80	86	86.4%	MIFAGIK
792.9	2	1565.7474	39	53	97.83%	KTGQAPGFSYTDANK

[Return to top]

Above, 8 of the original 11 spectrum were found to have a matching sequence with Cytochrome c. For example, the search found the peptide EDLIAYLK in the 482.7 spectrum. Each peptide match shows a high confidence: strong evidence for having found the correct protein. We can see a list of good spectra that did not match up to this protein by clicking the [Show missed *de novo* results] link.

We can also see exactly where the peptide fits into the protein sequence. To see where the match was made click the **CYC BOVIN** hyperlink (the one right above the peptide results). This brings up a "Protein View" window with the matching sequences highlighted in red. Close this window now.

As mentioned above, the peptide sequence results are based on a database search guided by an initial *de novo* analysis. Let's see how the *de novo* was able to help. Click on the **EDLIAYLK** hyperlink. This will bring up the spectrum 482.7 2 results window. Look in the top right frame to see the *de novo* and database results.

Color coding shows positional confidence scores. By the letters coded in red we can see that the PEAKS auto *de novo* analysis returned, with >90% confidence, the peptide sequence LAYLK, but was not as sure of the first three residues. The PEAKS DB Search was able to confirm this result, returning EDLIAYLK.

Clicking on the Peptide Data Frame, and selecting another spectrum, e.g. 589.3 1, will allow us to view the results from that spectrum without having to return to the protein identification result. Alternatively, click on "Protein ID Result" to return to the report.

Thus concludes our walkthrough of Peaks Studio v2.4's basic features.
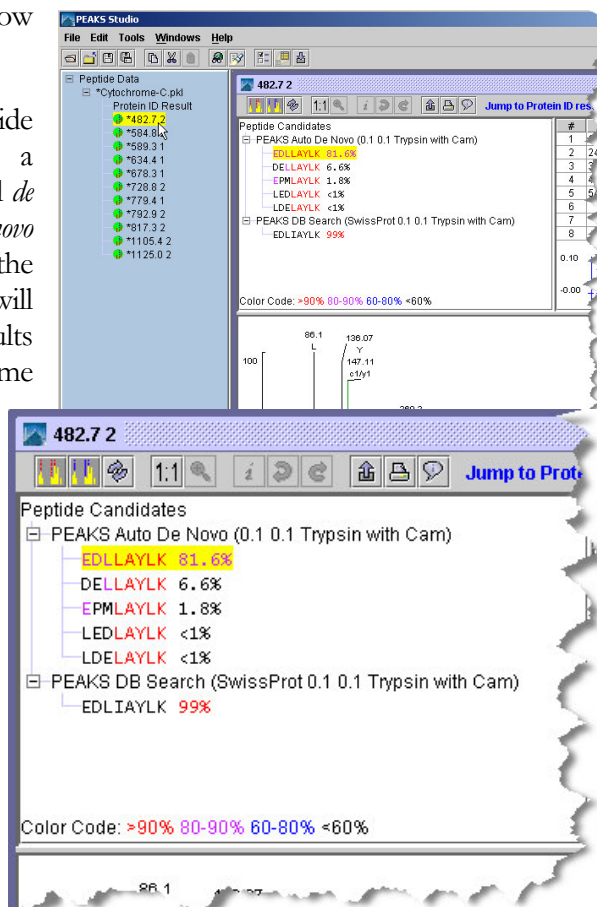
Using Peaks Studio v2.4

A task based guide to the most powerful peptide sequencing software

T his section of the manual will provide in depth explanation of Peaks Studio v2.4 usage. If we are stuck on a certain task, this section will walk us through that task with an in-depth series of instructions.

The first part of this chapter deals with **interface elements**. This tells us what certain dialogue boxes, windows and frames do and how to read them.

The second part of this chapter deals with **configuration tasks**. These should be performed before running auto *de novo*, manual *de novo*, or protein identification functions. However, Peaks Studio v2.4 allows us the flexibility to do this on the fly.

The final part of this section will provide instructions on how to efficiently use all of Peaks Studio v2.4's **auto *de novo*, manual *de novo*, protein identification, data manipulation and mass calculation functions**. Refer to the appropriate section below for help. For help integrating all these operations, perhaps it would be best to follow the walkthrough provided in chapter 3.

It is a good idea for us to become familiar with the interface, as described below, before proceeding.

Peaks Studio v2.4 interface

Windows, Dialogues, Frames and Reports
Peaks Studio v2.4 main window
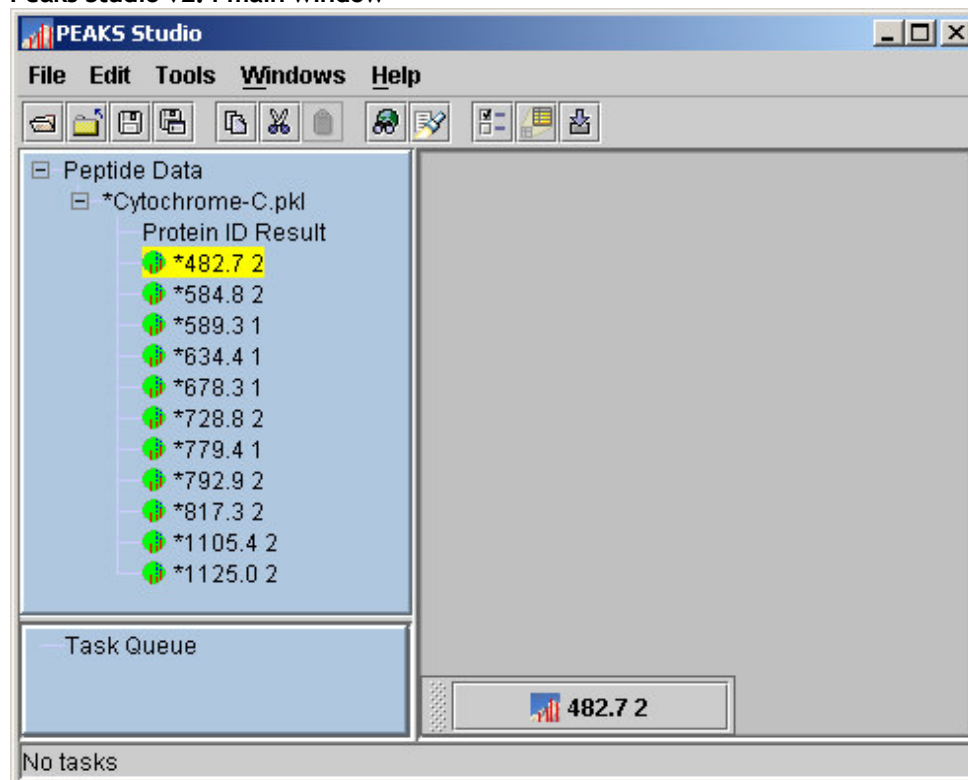
Comprises:

- **Peptide data frame (left):** Each spectrum is represented by its precursor ion information (m/z value followed by the charge of the precursor ion that generated the spectrum). Select all spectrum data files by clicking on "Peptide Data", an individual spectrum data file by clicking on its name, or a spectrum within a spectrum data file by clicking on it. The colored dot by each spectrum shows dark green for unprocessed, or light green for sequenced (or partially sequenced).

- **Task Queue frame (bottom left):** Shows running tasks, sorted by priority.

- **Working area (right):** This is where the Protein Identification Result Window and the Main Processing windows appear. Shown above: processing window for spectrum 482.7 2, minimized. These are discussed below.

- **Menu bar:** access file, edit, tools, windows, and help commands.

- **Main window toolbar:** quick access to many commands. See "Toolbars" section below

Auto de Novo Parameters Dialogue

This dialogue will appear each time we run auto _de novo_. If we always use the same settings, set them as default and _de novo_ will proceed automatically.
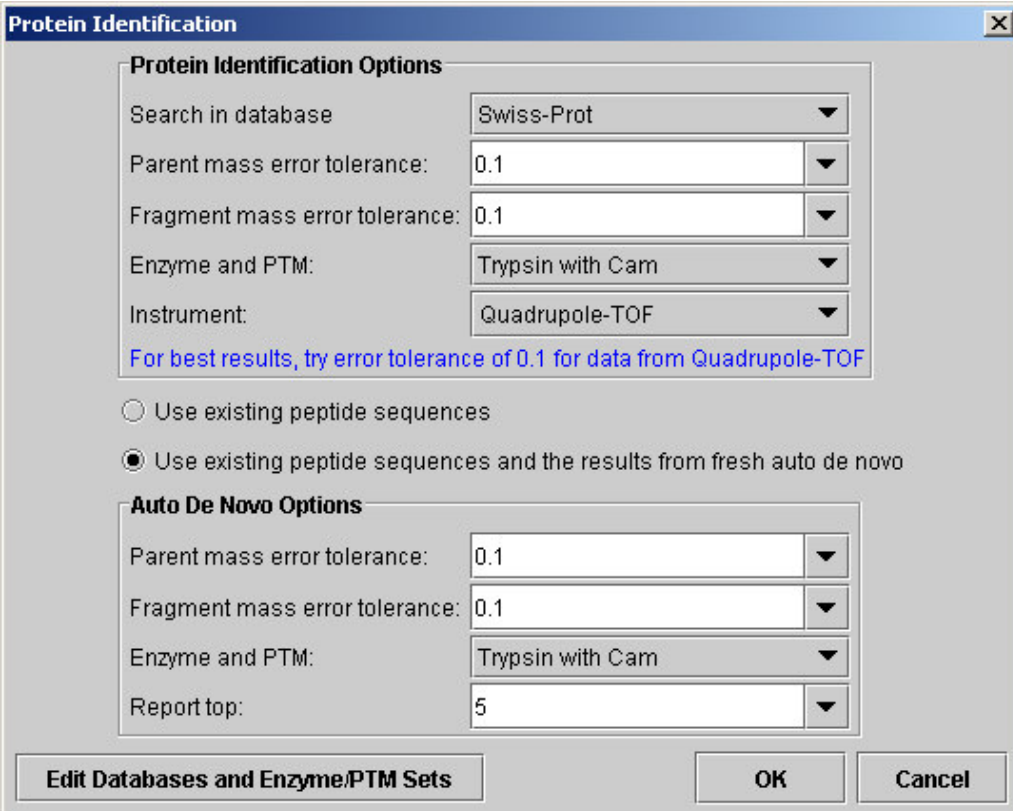

- **Parent mass error tolerance:** determines how much random and systematic experimental error on the parent/precursor ion PEAKS will account for in its analysis. Select a tolerance from the dropdown list.

- **Fragment mass error tolerance:** determines how much random and systematic experimental error on the fragment/daughter ion PEAKS will account for in its analysis. Select a tolerance from the dropdown list.

- **Enzyme and PTM:** choose from a dropdown list of enzymes that we used to digest our protein sample, and PTM that have acted on it.

- **Report top:** set how many peptide sequences PEAKS will report. Choose from a dropdown list.

- **Instrument:** choose the type of spectrometer that produced our data. Choose from a dropdown list.

- **Use this configuration as default:** check this box to keep these settings. The PEAKS Properties dialogue will no longer appear before auto _de novo_. It can be accessed any time by pressing the "Edit

Environment Preference" ▣ toolbar button, to load the Environment Preference Dialoged and then clicking the "auto *de novo*" tab.

- Click the "Edit PEAKS Properties" button to see the PEAKS Properties dialogue (discussed below).

Protein Identification Parameters Dialogue

This dialogue shows the same options for auto *de novo*, plus some options specific to protein identification:

- **Parent mass error tolerance:** determines how much error PEAKS will allow on the parent/precursor mass when searching through the database. Select a tolerance from the dropdown list.

- **Fragment mass error tolerance:** determines how much error PEAKS will allow on the fragment/daughter mass when searching through the database. Select a tolerance from the dropdown list.

- **Enzyme and PTM:** choose from a dropdown list of enzymes that we used to digest our protein sample, and PTM that have acted on it.

- **Instrument:** choose the type of spectrometer that produced our data. Choose from a dropdown list.

- **Use existing peptide sequences:** Select one of these radio buttons to either use sequence data that we've already created, or to perform auto *de novo* sequencing before the database search portion of our protein identification. The database search phase of protein identification uses whatever sequence information we already have, to filter through the database. If we have not already *de novo* sequenced our data (or if we wish to re-sequence *de novo*) we should perform auto *de novo* sequencing prior to the database search process.

- **Auto *de novo* options:** these are the same set of options as can be accessed from the auto *de novo* parameters dialogue.

- Click the "Edit Databases and ENZYME/PTM Sets" button to see the PEAKS Properties dialogue (discussed below).

PEAKS Properties Dialogue
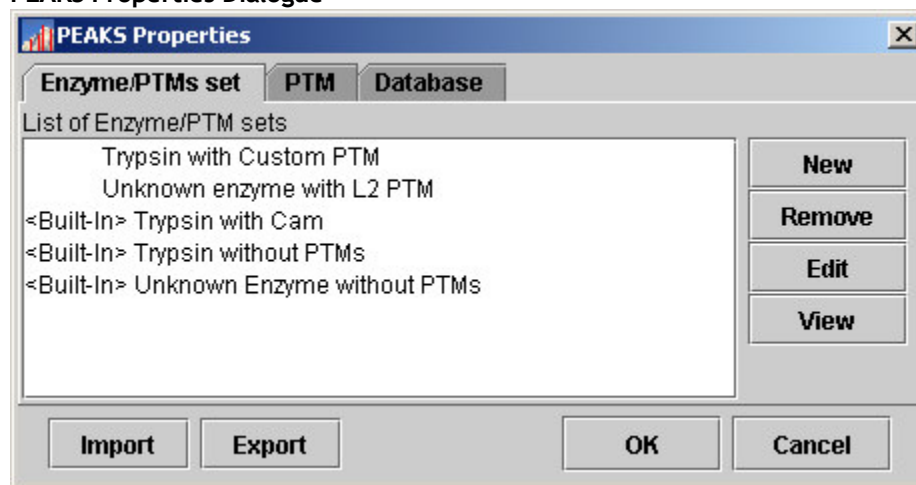
- **Enzyme/PTMs set tab:** Displays a list of built-in and user-defined Residues/PTMs sets. You may edit and create Enzyme/PTMs sets from here.

- **PTM tab:** Displays a list of built-in and user-defined PTMs. You may edit and create PTM from here.

- **Database tab:** Displays a list of databases available to PEAKS. You may make new databases available to PEAKS from here.

Enzyme/PTM Editing Dialogue
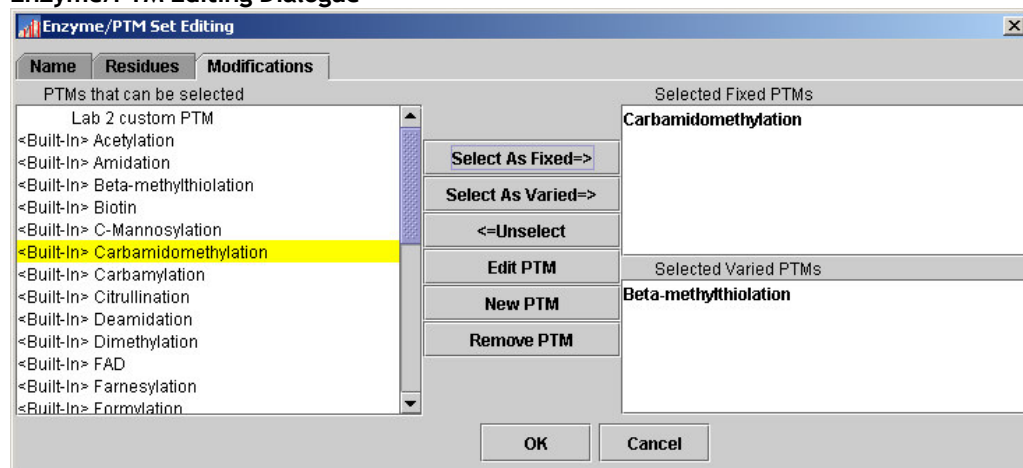
This dialogue is used in the process of creating or editing an ENZYME/PTM set. The three tabs in this dialogue allow you to establish different attributes of the set. (Shown above: the Post Translational Modifications tab). The process of editing and creating an ENZYME/PTM set is described in another section of this manual.

PTM Editing Dialogue

This dialogue allows us to create or edit a PTM

- **Name:** This will appear in the PTM list

- **Abbreviation:** This will appear in the auto *de novo* results, if it is found.

- **Mass (monoisotopic):** The mass that the residue gains or loses as a result of the PTM. Enter this numerically here, or enter the chemical formula below.

- **Neutral Loss Mass:** The mass that the modified residue loses as a result of fragmentation. E.g. 28 would signify a loss of 28 Daltons.

- **Formula:** The chemical formula of the PTM. This will automatically enter the mass.

- **Residues that can be modified:** Enter residues that can be modified anywhere, residues that can only be modified at the N-terminus and residues that can only be modified at the C-terminus.

- **Rule:** user entered, a comment for our reference.

Protein Identification Result Window

The Protein Identification Result Window is used to examine the results of a *de novo* guided database search. It presents the proteins that are, most likely, found in the sample.

- **Database Frame:** If we chose to search through more than one database (or with different sets of parameters) to identify the protein, our search results will be slightly different in each database. Selecting a database (and parameters) from the list will display the protein

candidates returned from that database (and parameters). Those search results will be displayed in the Search Result Report frame below. In the example above, the database "sprot" is selected with 0.1 error tolerance, and "Trypsin without PTMs" as the Enzyme/PTMs set.

▪ **Search Result Report Frame:** Displays the results of the *de novo* guided database search. It is divided into two parts. The index and the peptide match report.

  ▪ **Index (top section):** PEAKS presents a list of proteins that it believes to be the best match for the sample. This index lists them by accession number, ranked in descending order by score. *Very similar proteins (i.e. ones that contain most of or all of the sequences identified by PEAKS) are grouped together – only the first entry in this group is shown here. Access the whole group by clicking the **More…** hyperlink.* In the example above, CYC_BOV is the top ranked protein candidate.

  ▪ **Peptide Match Report (bottom section):** PEAKS presents each protein candidate with a peptide match list beneath it. Each peptide that matched the protein sequence is shown, in order, by spectrum. The confidence that the correct peptide sequence was found is displayed next to each peptide sequence.

Main Processing Window

The main processing window is used to perform manual *de novo* sequencing, and to examine the results of auto *de novo* sequencing
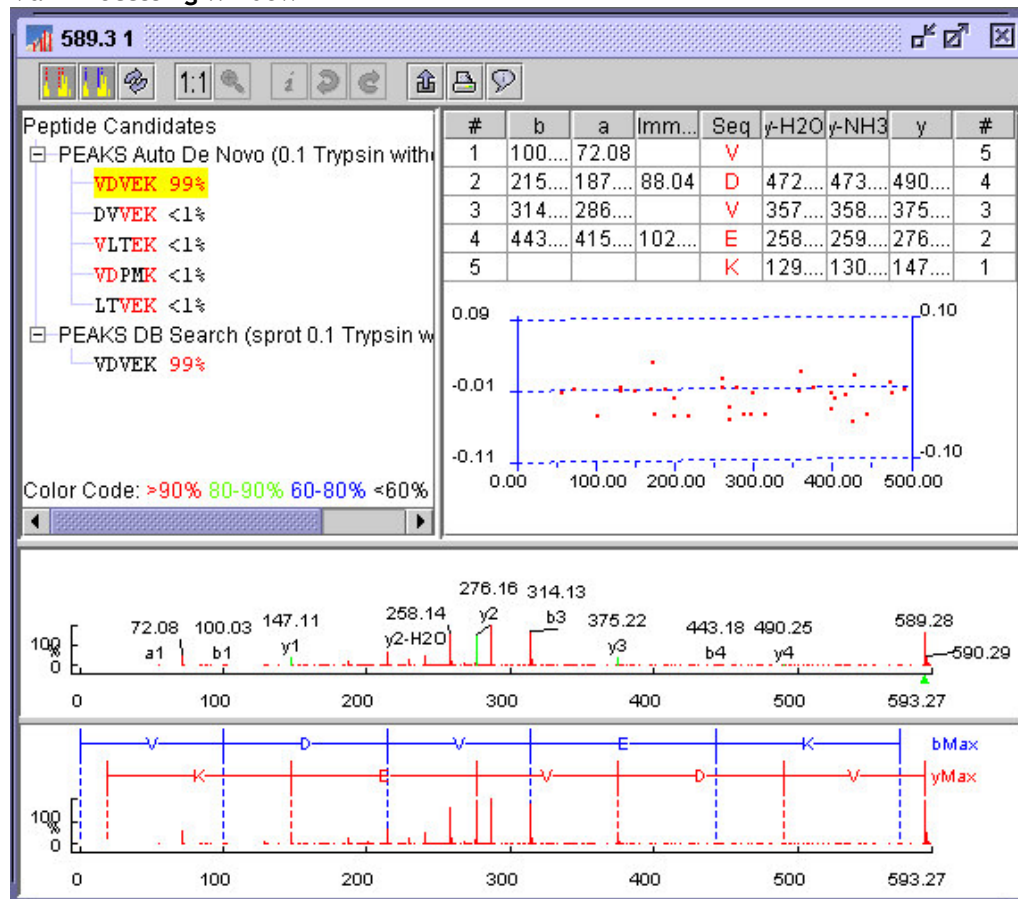
- **Main Processing Window Toolbar:** quick access to many processing functions. See "Toolbars" section below.

- **Peptide Candidates Frame (top left):** PEAKS shows peptide sequence candidates, ranked by score for the selected spectrum. Peptide sequences are grouped by the headings: "Auto *de novo*", "Manual *de novo*", "user defined result type" and "database search results" depending on how they were derived. For *de novo* results, positional confidence is color coded on each residue. More specific positional confidence appears when the mouse is held over a sequence – this shows the confidence in each of its parts.

- **Ion Table Frame (top right):** The Ion Table shows the proposed b, a, immonium, yH20, yNH3 and y ions, with their corresponding masses. (i.e. the mass of the y1 ion is shown in the bottom right corner.)

The Ion Table Frame also contains an error plot *(it may be necessary to scroll down to see the error plot)*. The error plot shows the confidence each ion is assigned. The most confident results lie on the centerline. Clicking a cell or column in the Ion Table highlights the corresponding points on the error plot and corresponding PEAKS on the spectrum.

- **Spectrum View Frame (middle):** Shows a graphical representation of the spectrum. PEAK masses are labeled, as are the PEAKS associated with identified ions. You can zoom in on the spectrum by clicking and dragging over an area.

- **Spectrum Alignment Frame (bottom):** Shows a graphical representation of the spectrum. This view always shows the whole spectrum and is used as a tool to help us navigate the spectrum view frame. A blue bar along the horizontal axis of the alignment view indicates the range of the spectrum view in the Spectrum View Frame. The Spectrum Alignment Frame can also show b-ion and y-ion PEAKS and the derived peptide sequence between them.

Ion Editor

The Ion Editor is used when performing manual *de novo* sequencing.

- **Selected peak information:** displays information about the currently selected peak.

- Under "**Please choose ion type**", the radio buttons set whether the ions in the "ion choice list" are C terminal ions or N terminal ions.

- **Ion choice list: (left):** lists the ions we can apply to the selected peak

- **Selected ion list (right):** lists the ions we have selected add or remove them using the "Add" and "Remove" buttons.

- **Apply button:** applies the ions in the "selected ion list" to the selected peak.

Export Image Dialogue
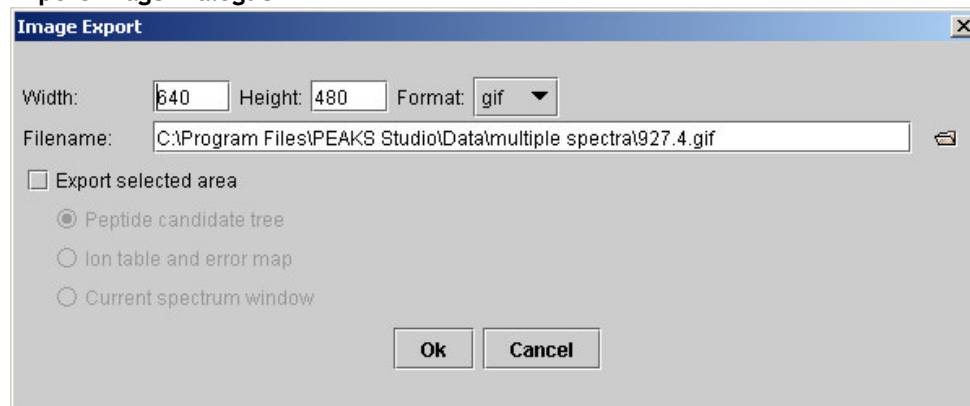
- **Width and Height:** together these determine the size of the output image (measured in pixels).

- **Format:** select an image file format from the drop down list. Bitmap, JPEG and Graphics Interchange Format are supported

- **Filename:** type in the textbox, or browse to a file ⌐, to enter the file name of the image that will be created.

- **Export selected area:** the default image output is the full spectrum as shown in the spectrum alignment frame, checking this box will allow us to print one of the other items. For example, if we have zoomed in on a portion of the spectrum and wish to print that zoomed in view, we click the export button and select the "Export selected area" checkbox, then the "current spectrum window" radio button.

Print Image Dialogue
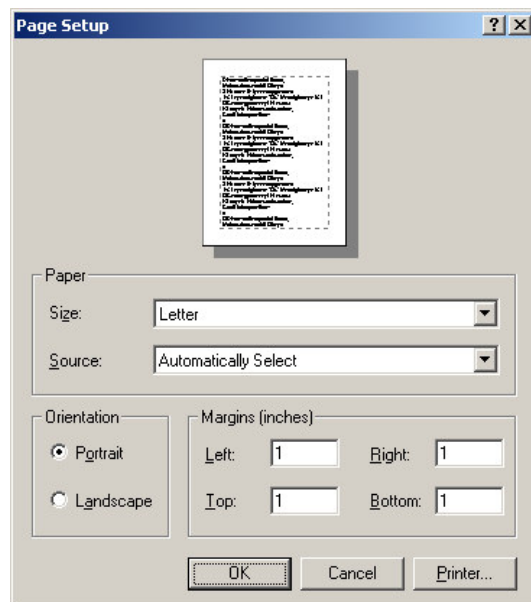
- **Orientation:** paper orientation is shown in the picture at the top. Change this by clicking the "Portrait" or "Landscape" radio buttons.

- **Paper:** Set the paper size and source by selecting from the appropriate dropdown list.

- **"Printer" button:** pressing this will bring up another dialogue where we can select from a list of printers installed on our machine.

- **"Ok" button:** this will commence printing.

The default print output is the full spectrum as shown in the spectrum alignment frame. If we wish to print something else, we must use the export image functions and then print the image from another application.

Toolbars
Main window toolbar

"Open data file" button: This allows us to open a raw data file built by our mass spectrometer, or a PEAKS data file (in ANN format) that also contains peptide analysis data. The file should be in PKL, DTA, MGF or ANN format.

"Close data file" button: Close the selected data file. Press this after selecting a data file in the Peptide Data Frame.

"Save data file" button: Save any changes made to the file (a '*' will appear next to any file that has been changed). The file will be saved in the ANN format. Press this after selecting a data file in the Peptide Data Frame.

"Save all files" button: Save all files. Any changes to files will be saved in the ANN format.

"Copy" button: Copy selected spectrum data.

"Cut" button: Cut selected spectrum data.

"Paste" button: Paste spectrum data into the selected spectrum data file.

"Automatic *De novo*" button: perform auto *de novo* for a selected spectrum data file, spectrum or list of spectrum data files. Press this after selecting one or more spectrum data files (or spectra) in the Peptide Data Frame. An auto *de novo* options dialogue will allow us to set parameters before we begin.

"Protein Identification" button: perform protein identification a selected spectrum data file. Press this after selecting one or more spectrum data files (or spectra) in the Peptide Data Frame. A protein identification options dialogue will allow us to set parameters before we begin.

"Environment Preference Configuration" button: configure the environment, spectrum color-coding, auto *de novo* parameters and manual *de novo* parameters.

"PEAKS Properties Configuration" button: define PTM/ENZYME sets, PTM, and add FASTA protein or EST databases.

"**Import Database Wizard" button:** help user download and configure database.

Main Processing Window Toolbar

"y ion Alignment" button: toggle (show/hide) the location of PEAKS corresponding to y-ions and the corresponding proposed peptides between them.

"b ion Alignment" button: button: toggle (show/hide) the location of PEAKS corresponding to b-ions and the corresponding proposed peptides between them.

"Deconvolve" button: toggle (on/off) deconvolution of the mass spectrum scale.

"1:1 zoom" button: return spectrum to original 1:1 zoom.

"Undo Zoom" button: return to previous zoom ratio.

"Edit Ion" button: set or edit the type of ion associated with a peak in manual *de novo*. Press this button after having selected a peak in the spectrum view frame.

"Next Peptide" button: redo changes to the peptide in manual *de novo*.

"Previous Peptide" button: undo changes to the peptide in manual *de novo*.

"Export Results" button: export the spectrum view, ion table, or to a picture (bmp, gif, or jpg format) with ions, masses, PEAKS and peptides marked.

"Print Results" button: print the spectrum view with ions, masses, PEAKS and peptides marked.

"View Results" button: show, in HTML format: the spectrum view with ions, masses, PEAKS and peptides marked; peptides and confidence scores; the ion table; and the error plot.

"**Jump to Protein ID result**" link: opens the protein identification result window.

File type handling

PEAKS handles spectrum data files in the following formats

- PKL

- DTA

- MGF

- ANN – the XML based file format associated with PEAKS

- XML format files from MicroMass ProteinLynx

As we know, the DTA file and some other MS data files only support one spectrum in each data file. To support protein identification for spectra from multiple data files, Peaks Studio v2.4 provides a tool to load all MS data files from one directory.

When saving our work, Peaks Studio v2.4 creates ANN files. The ANN data files contain the MS/MS information and peptide information of one spectrum and are stored all in one directory. The ANN index file is organizes multiple datasets. It links to a directory containing multiple ANN data files.

Configuring Peaks Studio v2.4

We can use Peaks Studio v2.4 without the need to configure; default settings will be used. However, to increase efficiency we should set environmental preferences, and PEAKS properties. This will enable us to customize the tool to our requirements. Configure PEAKS before processing data files.

PEAKS Properties Configuration

One of Peaks Studio v2.4 preferences, PEAKS Properties configuration sets the parameters that the algorithm will use in processing our data file. PEAKS properties include: PTM, Enzyme/PTMs set and FASTA database. Peaks Studio v2.4 provides tools to edit PEAKS properties for convenient use in *de novo* sequencing and protein identification. To edit PEAKS Properties:

- Click the ▣ icon in the main window toolbar;

- *Or,* from the "Edit" menu, select "Configuration" then "PEAKS Properties"

- *Or,* Click the "Edit PEAKS Properties" button in the Protein Identification (or Auto *De novo* Options) dialogue that appears before each Protein Identification (or auto *de novo*) operation.

The PEAKS Properties dialogue will then appear. This dialogue box has three tabs: "**Enzyme/PTMs Set**", "**PTM**" and "**Database**". Clicking a tab will allow us to edit the PEAKS Properties corresponding to that tab. We can also **import** or **export** our preferences to/from a file.

Enzyme/PTMs Set editor

Enzyme and PTMs are enzyme information and post-translational modification information, respectively. The Enzyme/PTMs set is the combination of ENZYME and PTM that we used in our experiment/sample. If we select an Enzyme/PTM set when we run auto *de novo* or protein ID, Peaks will look for those PTMs in sequence and return only peptides corresponding to the digest. For example, if we select 'Trypsin with Cam' as our set, Peaks will look for Carbamidomethylation on Cysteine.

To see a list of Enzyme/PTMs sets, select the corresponding tab in PEAKS Properties editor. From here we can create a new Enzyme/PTMs set, remove non-built-in Enzyme/PTMs sets, edit non-built-in Enzyme/PTMs sets and view the configuration of selected Enzyme/PTMs set.

PEAKS Properties editor will show all existing Enzyme/PTMs sets in the list. There are two types of Enzyme/PTMs sets: <built-in> sets and user defined sets. PEAKS software contains three built-in Enzyme/PTMs sets: 'Trypsin with Cam', 'Trypsin without PTMs' and 'Unknown Enzyme without PTMs'. These built in sets cannot be removed or edited by user. They cannot be overwritten, even if we try to create a new
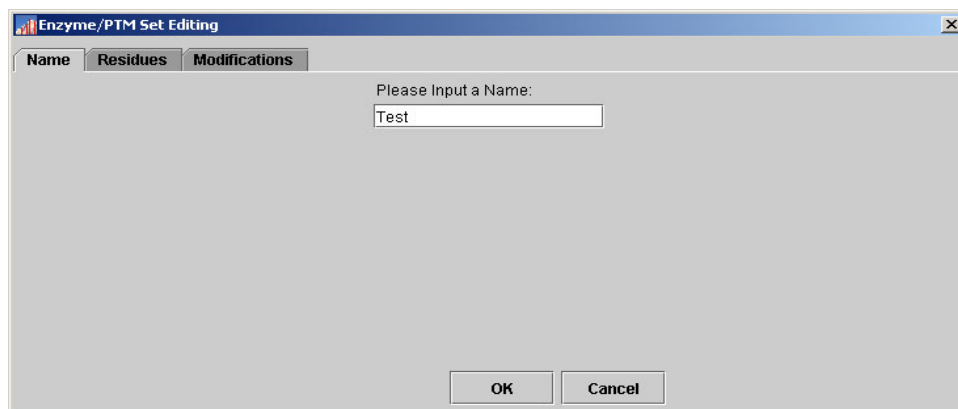
Enzyme/PTMs set with the name of built-in set. User defined sets can be removed, edited or overwritten at any time.

To create a new Enzyme/PTMs set, we open the PEAKS Properties dialogue, ensure that the 'Enzyme/PTM' set tab is selected and click the "New" button. The Enzyme/PTM Editing dialogue will appear. Click the "Cancel" button at any time to exit, discarding changes.

From the Name tab, we must enter a name for our new Enzyme/PTMs set. This is for our reference only, but it makes sense to enter a name that will remind us of the enzyme information and PTM that we are to enter (in the above example we've simply entered 'Test'). *Clicking the finish button after entering only a name, our Enzyme/PTMs set will be created using 'Trypsin' as the enzyme information and 'no PTM' as the PTM information.*

Clicking on the Residues tab brings up the following screen:
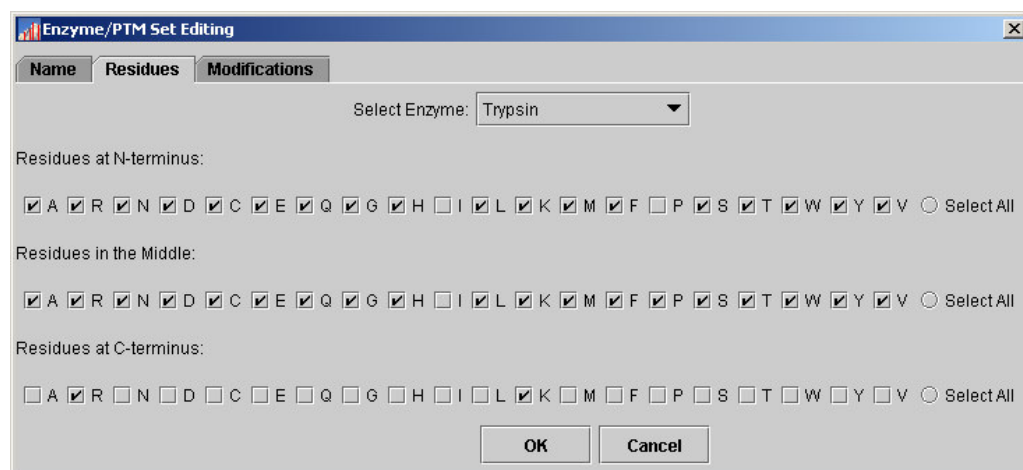
Here we can enter enzyme information into our set. We can choose pre-defined enzyme from the drop-down list. Based on the digestion properties of the enzyme, the

residues that are likely to be found at the N-terminus, C-terminus and in the middle are marked with checkboxes (the check appearing to the left of each letter symbol.) If the residue positions are not quite right for our enzyme we can customize residues for different positions by selecting appropriate checkboxes (the original definition of the enzyme will not be overwritten). Alternatively, we can choose 'Unknown Enzyme' from the dropdown list to define a new enzyme from scratch. If we click the "Finish" button without specifying PTM, PEAKS will assume that we do not wish to specify any PTM:

To specify PTM, click the Modification tab to bring up the following screen:

Selecting a PTM as fixed tells PEAKS that the PTM will is applied in all instances. Selecting a PTM as Varied tells PEAKS that it may or may not be applied.
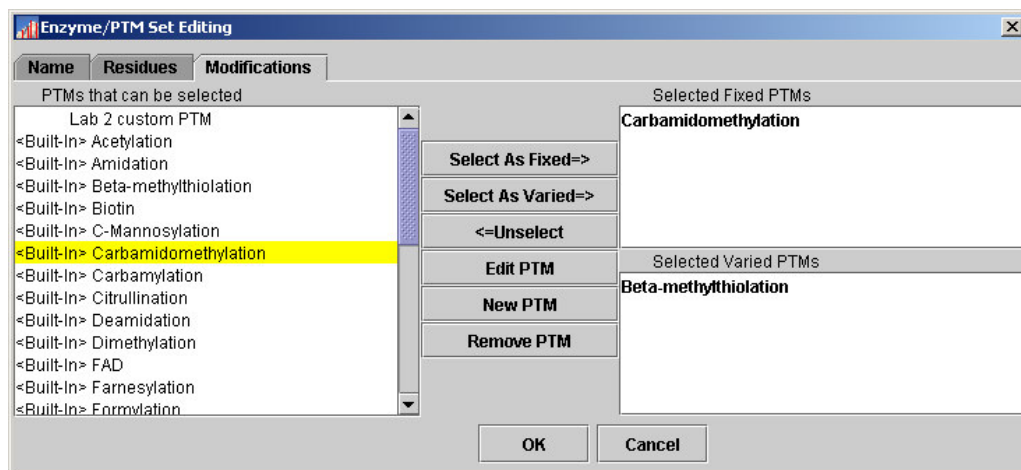
Here we can from a list of available post-translational modifications. We can choose any PTM as Fixed PTM or Varied PTM (to tell PEAKS that it may or may not be applied). To make this selection, click on a PTM in the list at left, and then click the "Select As Fixed=>" or the "Select as Varied=>" button. If a PTM is already selected as a fixed PTM, it cannot be selected as varied PTM, and vise versa.

If we change our mind about a PTM after having selected it, it is still possible to unselect it. Click the erroneous PTM from the list of 'Selected Fixed PTM' or 'Selected Varied PTM', and then click the "<=Unselect" button to remove it from either list of Selected PTM.

> PEAKS software ships with some pre-defined PTMs. If we want to create a new PTM other than the built-in PTMs, we can click 'New PTM' to create a new one. The **Editing PTM** section below describes how this is done.

If we click the "Finish" button without specifying residues, PEAKS will assume that this is a tryptic digest, and apply the appropriate residues.

After having built our Enzyme/PTMs set, we can click the "Finish" button to save the new Enzyme/PTMs set, or the "Cancel" button to exit, discarding changes. After clicking the "Finish" button we return to the PEAKS Properties dialogue to find that our new Enzyme/PTMs set is listed in the 'List of Enzyme/PTM sets'.
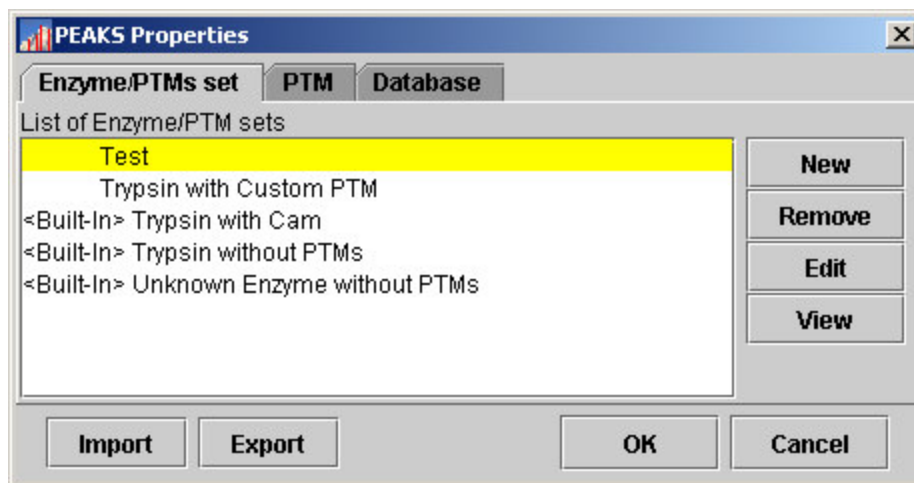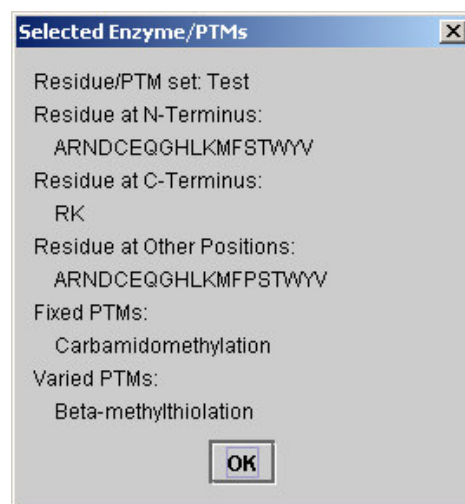

EDITING AN ENZYME/PTMS SET

To edit an existing Enzyme/PTMs set, we open the PEAKS Properties dialogue and ensure that the 'Enzyme/PTM' set tab is selected. Choose a user defined Enzyme/PTMs set from the list and click the "Edit" button. *We cannot edit the built-in Enzyme/PTMs sets.* The Enzyme/PTM Editing dialogue will appear. We now follow the same procedure as we would if we were creating a new Enzyme/PTMs set (see above section).

VIEWING AN ENZYME/PTMS SET

To view the configuration of an existing Enzyme/PTMs set, we open the PEAKS Properties dialogue and ensure that the 'Enzyme/PTM' set tab is selected. Choose an Enzyme/PTMs set from the list and click the "View" button. This presents us with a summary of the selected Enzyme/PTMs set, including enzyme and PTM information. Click the "Ok" button when finished. We cannot edit the selected Enzyme/PTMs set from this screen.

To remove a user defined Enzyme/PTMs set, we open the PEAKS Properties dialogue and ensure that the 'Enzyme/PTM' set tab is selected. Choose a user defined Enzyme/PTMs set from the list and click the "Remove" button. *Built in Enzyme/PTMs sets cannot be removed.*

PTM Editor

If we know that our sample protein may have been modified since translation, we need to apply this information to our analysis. This is done by creating an Enzyme/PTMs set which incorporates enzyme information and PTM information (see above section). To edit the list of PTM available to these sets we use the PTM Editor. To select PTM editor, click the tab 'PTM' in PEAKS Properties editor.

PTM Post Translational Modifications affect the mass of modified proteins and residues
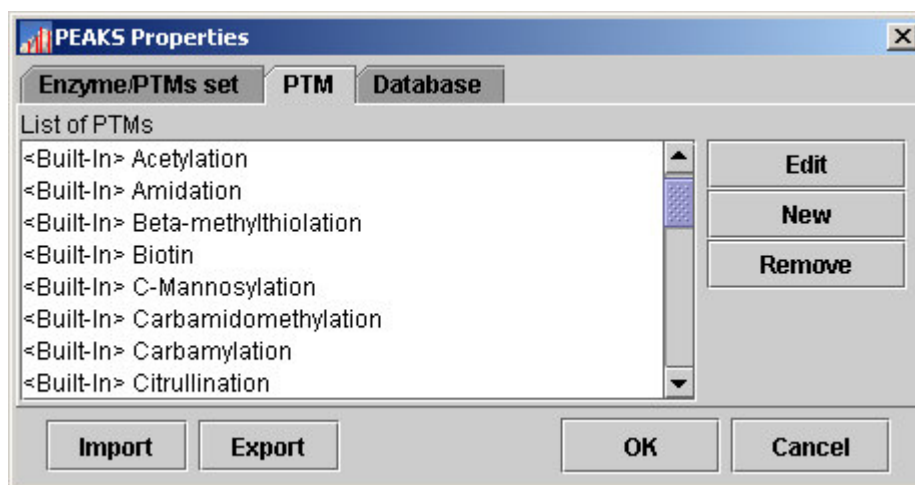
All PTM are listed here. There are two types of PTM available to PEAKS: <built-in> PTM and user defined PTM. From here, we can create a new PTM, edit an existing PTM or remove a PTM from the list. Built-in PTM cannot be removed from the list, but can be edited.
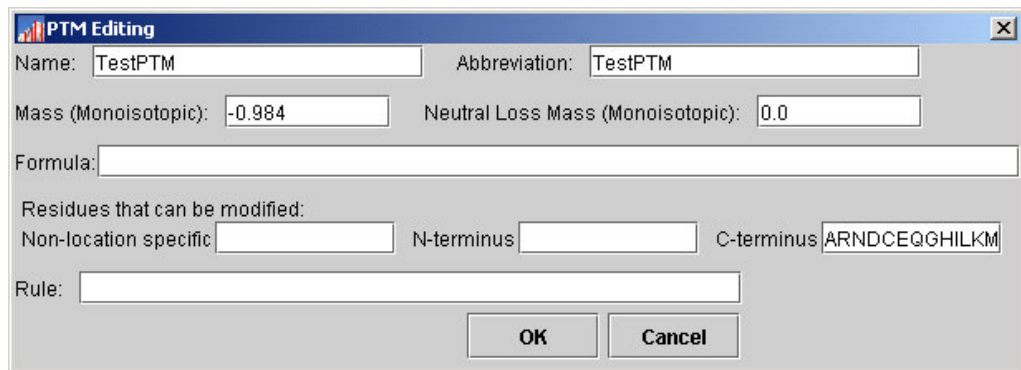
Editing a Built in PTM

It is possible to modify a built in PTM. PEAKS will save the modification and treat this PTM as a customized PTM. It will temporarily overwrite the built-in PTM - we will not be able to see the original, built-in, PTM until we remove the customized one. We can remove this customized PTM at any time and the built-in PTM will reappear.

To create a new PTM, we open the PEAKS Properties dialogue, ensure that the 'PTM' tab is selected and click the 'new' button. *To create a new PTM 'on the fly' while making an Enzyme/PTMs set, click the "new PTM" button while selecting PTMs.* The PTM Editing dialogue will appear:

Figure 1 Create new PTM

Now we type information pertaining to our PTM in the appropriate boxes (see above section on **Interface** for a more in depth explanation of these fields). At a minimum, we must enter a name, a mass and one residue that may be modified.

Enter the mass of the modification either by typing in its monoisotopic mass difference directly, or by entering its empirical formula. It is unnecessary to do both; each will override the other.

Click the "Ok" button to save changes and create our new PTM, or click the "Cancel" button to exit, discarding changes.

After clicking the "Ok" button we return to the PEAKS Properties dialogue to find that our new PTM is listed at the top of the PTMs list.
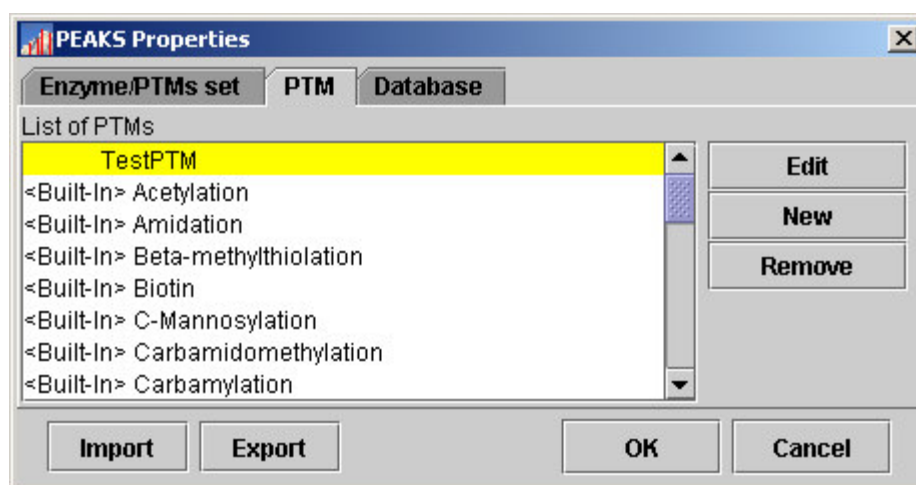
To edit a PTM, we open the PEAKS Properties dialogue, ensure that the 'PTM' tab is selected, and select a PTM from the list by clicking on it and click the 'Edit' button. *To edit a PTM 'on the fly' while making an Enzyme/PTMs set, click the "edit PTM" button while selecting PTMs.* The PTM Editing dialogue will appear. Now we follow the same procedure (see above) as we would if creating a new PTM.
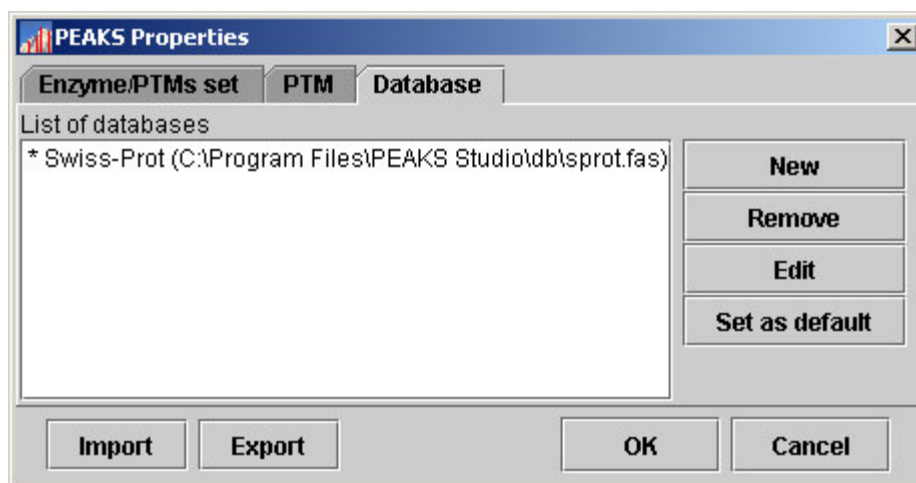
To remove a PTM, we open the PEAKS Properties dialogue, ensure that the 'PTM' tab is selected, select a PTM from the list by clicking on it, and click the 'Remove' button. *To remove a PTM 'on the fly' while making an Enzyme/PTMs set, click the "edit PTM" button while selecting PTMs.* Built in PTM cannot be removed.

Database Manager

Peaks Studio v2.4 needs a protein or EST database (in FASTA format) to identify protein candidates. Since databases are being constantly updated, PEAKS does not ship with a protein or EST database. Thus, we need to download it from the Internet and tell PEAKS where the database is located. PEAKS provides the Database manager as a tool to help us do this. To see a list of databases available to Peaks Studio v2.4, load the PEAKS Properties dialogue and click the 'Database' tab. From here we can edit a database's properties, load a new database, or remove a database.

LOAD/CONFIGURE A NEW DATABASE

For an in depth look at configuring a database, see the Database Configuration section in Chapter 2.

To configure a new database, we open the PEAKS Properties dialogue, ensure that the 'Database' tab is selected and press the "New" button. Now we open up our web browser to find a database to download. Find one, download it, and unpack it. Return to Peaks Studio v2.4 and find the file on our system where we unpacked it. Name the database and select the header format to use (or we can define our own). Click "Ok". The new database will now appear, listed by our chosen name, in the list of databases.

REMOVE A DATABASE

To remove a database, we open the PEAKS Properties dialogue, ensure that the 'Database' tab is selected, select a database from the list of databases, and press the "Remove" button. This will not permanently remove it from our system, it may be reloaded (follow procedure for configuring a new database) at any time.

EDIT A DATABASE

This tool allows us to change the name that PEAKS associates with a database, and the header parsing rules for that database. It does not edit the database itself. To edit a database, we open the PEAKS Properties dialogue, ensure that the 'Database' tab is selected, select a database from the list of databases, and press the "Edit" button.

SET THE DEFAULT DATABASE

Setting the default database changes the database that the protein identification search will use by default. You will still be able to select another database before doing the search if we so choose. To edit a database, we open the PEAKS Properties dialogue, ensure that the 'Database' tab is selected, select a database from the list of databases, and press the "Set as default" button. The default database is marked with a * before its name in the list of databases.

If we choose to move a database to another directory, or delete it entirely, we must let PEAKS know. We must remove the database from the list and re-load it. Until we do so, the database name will appear in red in the list of databases, and any protein identification using that database will fail.
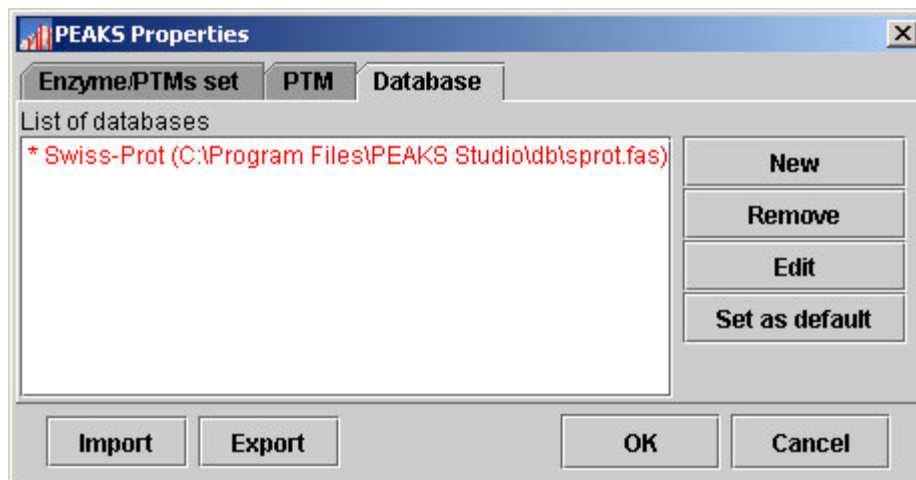
If we choose to update the database by downloading the latest database file and overwriting the old database file, PEAKS will show the database information in light gray. *A light grey colour could also mean that the database does not have header information.*
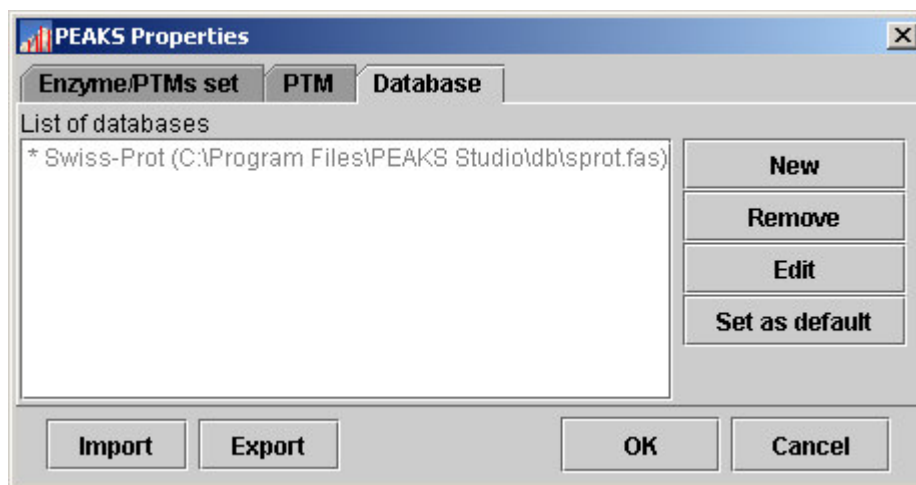
Importing and Exporting PEAKS Properties

We may wish to use Peaks Studio v2.4 on another system. However, if we have a large number of user defined PTM and Enzyme/PTMs sets it could take a great deal of our time to re-input those. This is where importing and exporting of PEAKS properties is useful.

The export function will save PEAKS Properties information in a XML file. The import function can read a PEAKS properties XML file and overwrite local PEAKS Properties with the information from XML file. If we are using our PEAKS properties on a colleague's system, we must remember to export their properties to a file so that they can import them back later.

A note on sharing sequences with PTM

Sequence data and protein identification results for a given spectrum are stored in an .ANN file. Any modifications that were found in the sequence are also included. As such, user-defined modifications will still show up if the file is viewed on another machine. It is not necessary to import all PEAKS properties to view these modifications. Also, user-defined modifications can be extracted from an .ANN file and added to the local PEAKS properties.

To export PEAKS properties to a file, open the PEAKS Properties dialogue and press the "Export" button. Type in a file name and press the "Save" button.

To import PEAKS properties from a file, open the PEAKS Properties dialogue and press the "Import" button. Select a file or type in a file name and press the "Open" button. This must be a PEAKS configuration file in XML format.

To save user defined PTM to an .ANN file for sharing, we simply save our work to an .ANN file as normal. Any user defined PTM will be preserved in the .ANN file.

To import a user defined PTM from another user's .ANN file, we open the .ANN file and find a sequence containing the user defined modification. Right click on that sequence to bring up the popup menu. Click the "View Modifications" menu item. This brings up a dialogue box named "Modifications". Select the PTM of interest from the dropdown list (in this example: 'Lab 2 custom PTM') and click the import modification button.
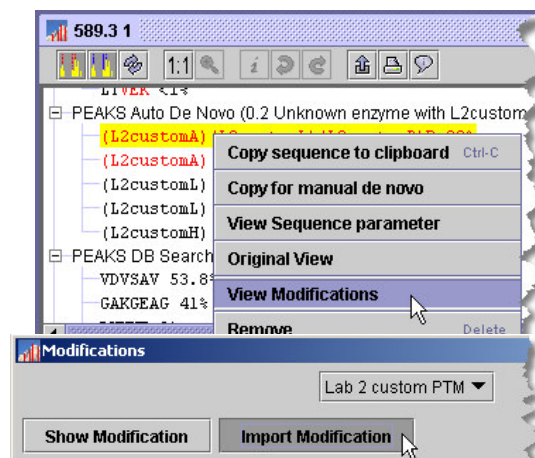
PEAKS Environment Preference Configuration

One of Peaks Studio v2.4 preferences, PEAKS Environmental Preference allows us to customize Peaks Studio v2.4 to our needs. PEAKS Environmental Preferences include: Environment, Color, Auto *de novo*, and Manual *de novo*. **To edit Peaks Studio v2.4 Environment Preferences:**

- Click the ▦ icon in the toolbar;

- *Or,* from the "Edit" menu, select "Configuration" then "Environment Preference"

The Environment Preference dialogue will then appear. This dialogue box has four tabs: '**Environment'**, '**Color'**, '**Auto *de novo'*** and '**Manual *de novo***'. Clicking a tab will allow us to edit the Environmental Preferences corresponding to that tab.

Environment

We can change the environment settings so that, when we are browsing our system to find or save data files, PEAKS always starts looking in the folder we specify.

To change the working environment, we open the Environment Preference dialogue and ensure that the "Environment" tab is selected. The current working folders for data input and data output are shown. We can choose to have Peaks Studio v2.4 use the last folder we loaded from (saved to) as the current working folder, or toggle the appropriate radio button to "User directory" to set it ourselves so that it will be the same each time.

The directory where PEAKS stores its preferences information cannot be changed in this version of PEAKS.

Additionally, we can choose whether or not we'd like PEAKS to close a spectrum view window once we have finished with it and move on to another spectrum. Click the checkbox at the bottom of the window.
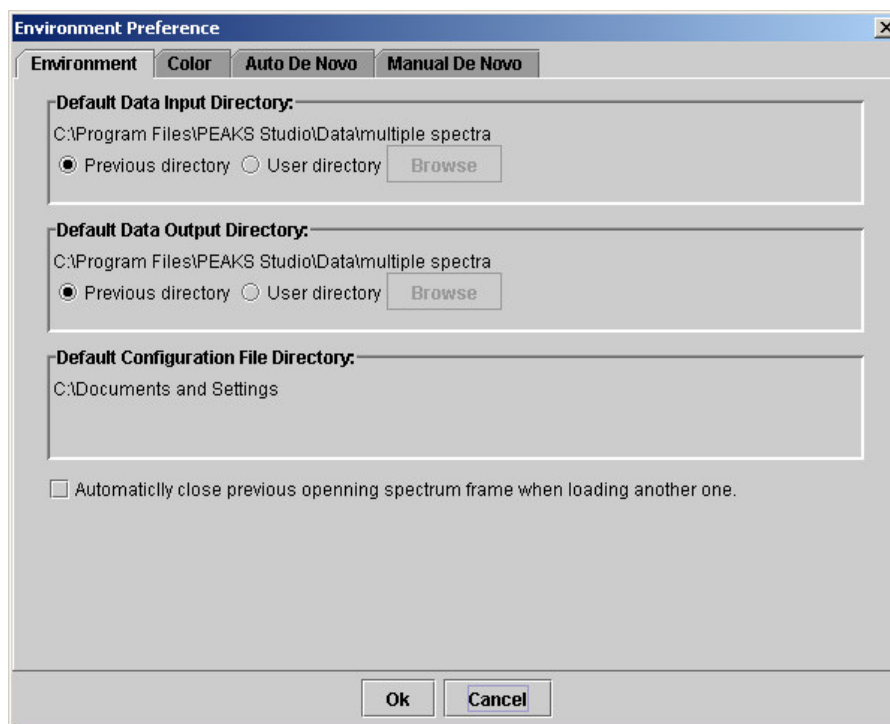
Once we've chosen from these options, pressing the "Ok" button will exit, saving changes. The "Cancel" button will exit discarding changes.

Colors

For ease of viewing, we can choose which colors we would like to represent which items on the spectrum view. To change the color of an object on the spectrum, we open the Environment Preference dialogue and ensure that the 'Color' tab is selected.
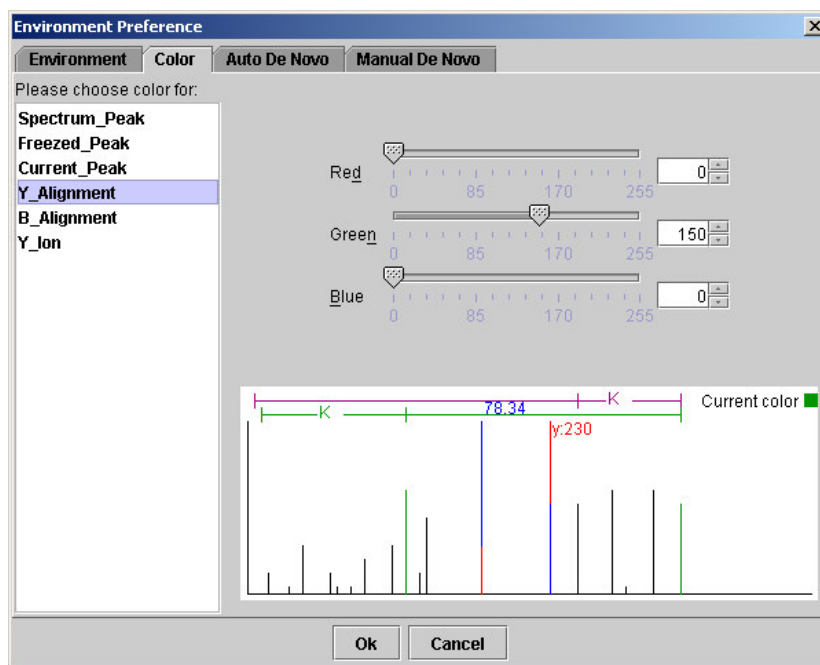
Now we choose the object whose color we'd like to change from the list at the left. Then we click on the slider bar, type in a number (0 to 255) in the textbox, or scroll up and down on the arrows next to the textbox to select how much of the corresponding color we'd like to apply. Choose an amount for all three colors. In the example above, we've chosen pure red (255) to represent a spectrum peak. After we've chosen colors we may click the "Ok" button to exit and save changes.

Auto *de novo*

Before performing an auto *de novo* analysis of our spectrum data, we must specify certain parameters to the PEAKS algorithm. These include error tolerance, Enzyme/PTMs set to use, database to search and instrument used. Normally we do this before each auto *de novo* run, but here we can set defaults.

To set parameters for auto *de novo*, we open the Environment Preference dialogue and ensure that the 'Auto *de novo*' tab is selected. We then select a value for each of the parameters to use by clicking on the dropdown list and selecting one of the options. Clicking the Residue and PTM dropdown list brings up the list of our defined Enzyme/PTMs. Please refer to the Peaks Studio v2.4 Interface section for information on what these values mean.

To see how to change the list of Residue and PTMs, refer to the above section: PEAKS Properties Configuration

These values can also be set from the Auto *de novo* options dialogue that appears before each Auto *de novo* operation.

If we always use the same instrument and the same experimental conditions, we may wish to avoid having to confirm these options every time we perform auto *de novo* sequencing. To do so, check the box next to "use this configuration as default". These settings will be used for each auto *de novo* sequencing operation, without confirmation. Un-checking the box will make the Auto *de novo* options dialogue appear before each auto *de novo* once again.

Protein Identification

Settings for the database search portion of the analysis can only be set immediately before each protein identification operation. That is to say, to set parameters for database search and protein identification, we click on the Protein Identification button.

To change the default database, see the above section: PEAKS Properties Configuration

Next, we choose which database to search through. This is done by selecting a database from the dropdown list next to "Search results in database". The default database is already selected, but we may choose another. Choose, from the dropdown lists, the amount of error (m/z tolerance) that the database search will tolerate.

> ### A note on error tolerance and Enzyme/PTMs
>
> The error tolerance and Enzyme/PTMs can be set separately for the *de novo* and database search portions of the analysis. We will get better protein identification results if we follow best practices by setting them differently. Auto *de novo* should be run with no variable PTM, but with the correct enzyme and any fixed PTM. Modifications should be then turned on for the database search function. m/z tolerance can also be adjusted separately for each phase to allow us to tweak the results.

Manual *de novo*

We may wish to sequence a peptide manually, using spectrum data. Peaks Studio v2.4 provides us with a set of tools to help us do so. We may need to tweak these tools to adjust for error tolerance, and to customize the working environment. Adjusting error tolerance is important if, for example, we know that tagging a peak should label a certain residue, but PEAKS will not do so – this may be because it is tolerating enough error.

To adjust Manual *de novo* options, open the Environment Preference dialogue and ensure that the 'Manual *de novo*' tab is selected.
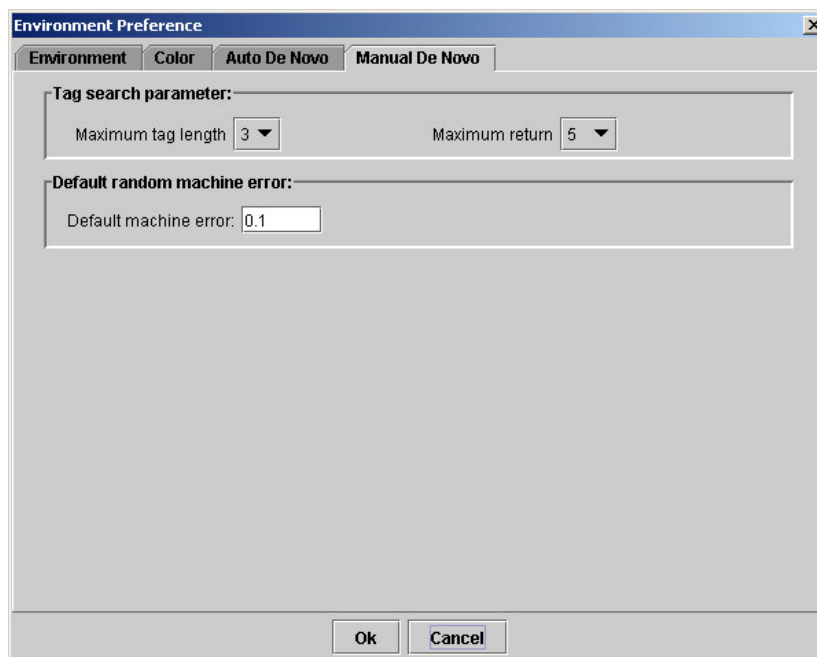
When sequencing a peptide using the manual *de novo* tools, we can get PEAKS to help us by searching to the left or right of a selected peak and returning a set of possible

sequence tags (see Manual *De novo* section later in this chapter). We can choose how many search results we'd like to see, and we can choose how long (number of amino acid residues) we'd like these tags to be at a maximum. To choose how long tags will be, we click on the "Maximum tag length" dropdown list box, and making a selection. Choose the number of search results displayed by clicking on the "Maximum return" dropdown list and making a selection.

Changing the default machine error sets the amount of error PEAKS will tolerate when tagging a residue. *For example: we have a mass difference of 113.14 between two y-ions that we have labeled. We are fairly confident that this should be tagged L (Leucine), but PEAKS is not labeling it for us. This may be because 113.14 is too far out of PEAKS error tolerance for the mass of L.* We can tweak the settings until we get the desired result. To do so, type a value for error (larger numbers indicate greater tolerance) into the "Default machine error" textbox.

After having made all desired changes, click the "Ok" button to save changes and exit the dialogue box. Click the "Cancel" button to exit, discarding changes.

Using Peaks Studio with modifications (PTM)

Peaks Studio v2.4 ships with a library of 30 or so commonly used post-translational modifications (PTM). But Peaks will not automatically look for these modifications. We must tell Peaks which ones to look for. To do so, we must create an "Enzyme/ PTMs set". For help creating Enzyme/PTMs sets please refer to the **Peaks Properties Configuration** section of the user manual – look for the **Enzyme/PTMs Set editor** heading.

When creating our set, we should choose a name that refers to the enzyme we used to digest and reminds us of the PTM Peaks must look for. We must select the enzyme that we used to digest our sample. We must also select which PTMs Peaks should look for. Select PTMs as Varied or Fixed. Save this set and it will appear in the list of Enzyme/PTMs sets. For our convenience, this set will be saved within Peaks since we will probably process many samples this same way.

Now that we have created the Enzyme/PTMs set we can select it before running auto *de novo* or protein ID.
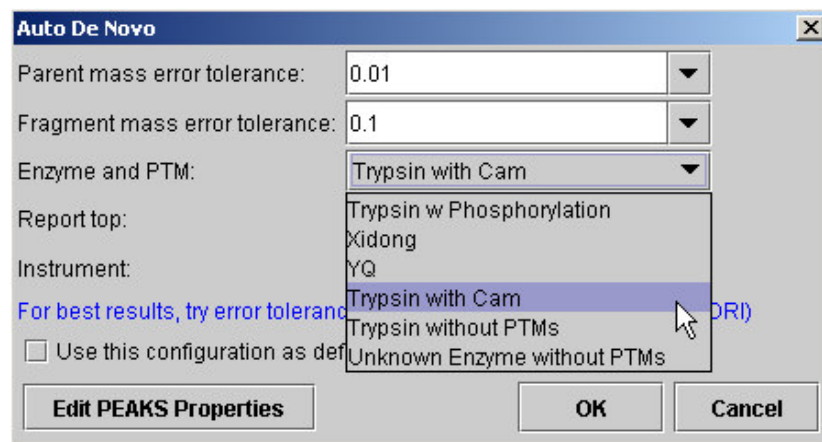
Remember, when doing protein identification, we will get best results if we:

- For Protein Identification database search portion, choose an Enzyme and PTM set that contains the correct enzyme and correct PTMs. If we don't know which PTM to expect, turn them all on.

- For auto *de novo*, choose an Enzyme and PTM set that contains the correct enzyme, correct fixed PTM, but no variable PTM.

The more variable PTM we turn on, the more ambiguous will be the results.

Remember, when doing auto *de novo* sequencing, we will get best results if we:

- Choose an Enzyme and PTM set that contains the correct enzyme, correct fixed PTM and a few variable PTM.

Opening spectrum data files

In order to perform *de novo* sequencing and protein identification we must first load our spectrum data into Peaks Studio v2.4. To open a spectrum data file, click the ⬚ icon on the toolbar in the upper left corner of the PEAKS window or select "Open" from the "File" menu.
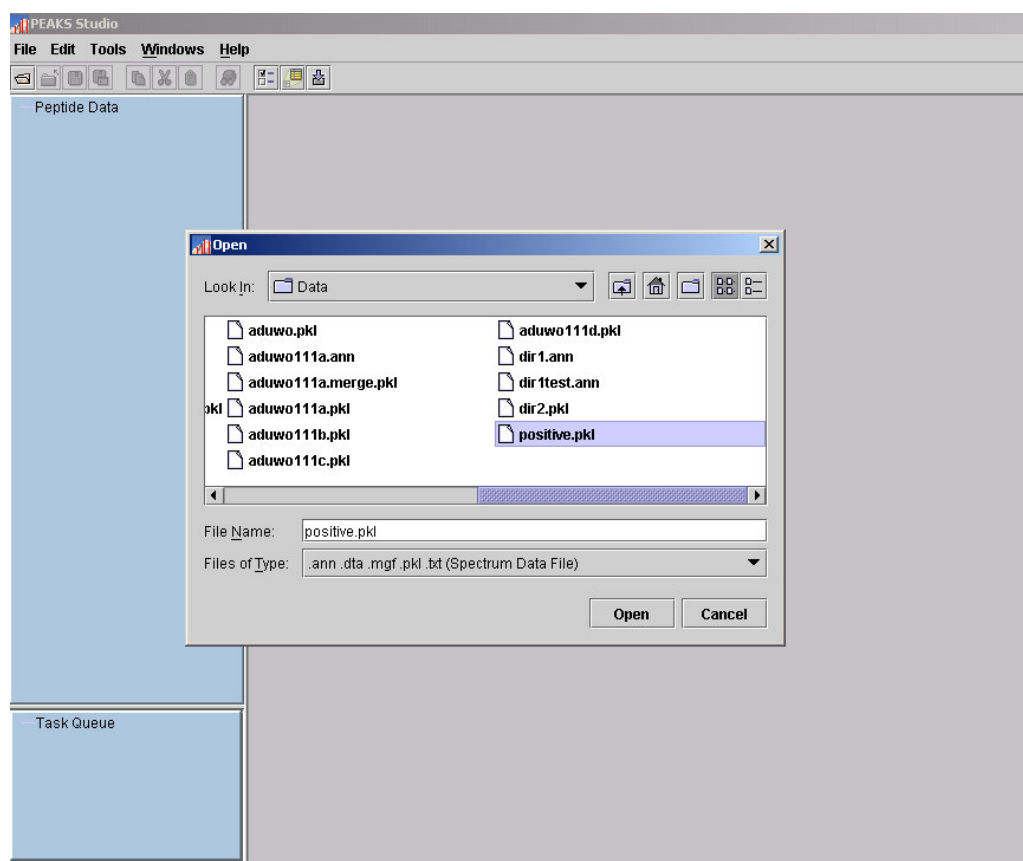
Select a file (in .ann, .pkl, .mgf, .dta, or .txt format). ANN data files for a whole spectrum are stored in a separate directory and are linked to by an ANN index file with the same name as the directory. Open the ANN index file in the root directory to open the whole spectrum, each spectrum can be accessed from within here.

The data file we just opened appears in the Peptide Data Frame on the left. It is represented by its file name. Each spectrum contained in the spectrum data file is represented by its precursor ion information (m/z value followed by the charge of the precursor ion that generated the spectrum).

Loading a directory full of data files

DTA spectrum data files can be opened by the same procedure as listed above. However, as we know, some DTA files contain the data for only one spectrum. As such, we may find it useful to import a whole directory (containing DTA MS/MS spectrum data files for a whole MS spectrum) at once. Peaks Studio v2.4 provides a tool for doing so.

Under the "File" menu, click "Load Directory.." Now we browse to the directory we wish to load.
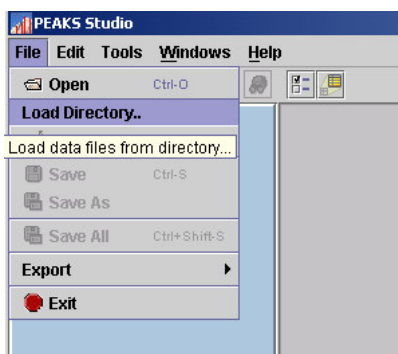
Do not select a file within the directory; rather select the directory itself.

After loading the spectra, we can choose sort the spectrum by the source filename or by the precursor m/z value of spectrum. To do so, right click the parent node on the Peptide Data, and choose to sort.

Auto De novo Sequencing

To begin auto *de novo* sequence derivation, we:

1. In the *Peptide Data* Frame, select the data file(s) containing the spectra that we wish to sequence by Auto *de novo*. We can also select an individual spectrum, or a few spectra within a data file; auto *de novo* will proceed on only the spectra selected.

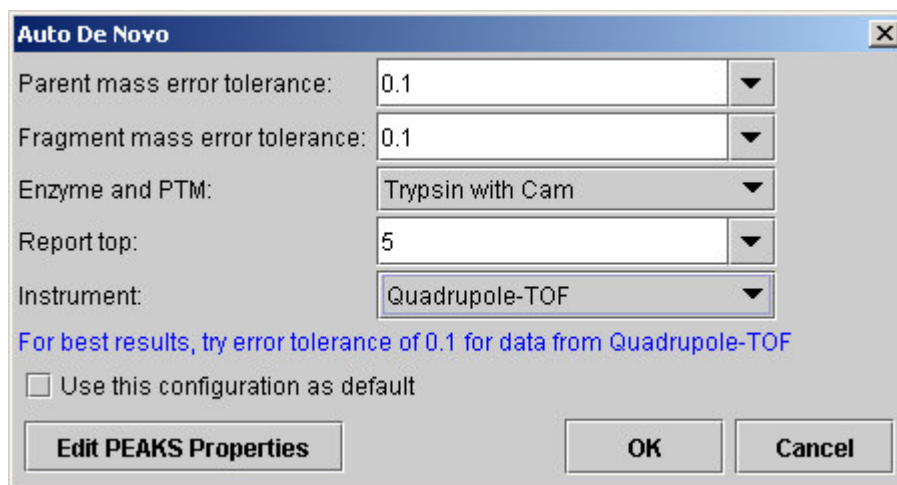2. Click the 🐞 Automatic *De novo* toolbar icon

   Or

   Select "Auto *De novo*" from the "Tools" menu.

   Or

   Right click on the selected spectra or data files and select "Auto *De novo*" from the popup menu

The Auto *de novo* Parameters dialogue window will appear:

Suggested values for error tolerance appear on-screen. In most cases, we can use this number for both parent and fragment error. We should begin by using the suggested values, then try some slightly higher or lower ones to find the best result.

3. If we wish to change any of these parameters, we do so now.

   - Parent error tolerance: how much random and systematic experimental error on the parent/precursor ion mass PEAKS will account for in its analysis.

   - Fragment error tolerance: how much random and systematic experimental error on the parent/precursor ion mass PEAKS will account for in its analysis.

- Enzyme and PTM: enzyme that we used to digest our protein sample, and PTM that have acted on it.

- Report top: how many peptide sequences PEAKS will report.

- Instrument: the type of spectrometer that produced our data.

4. Press the "Ok" button to commence Auto *de novo* sequencing.

Once a job is submitted to Peaks Studio v2.4, it is added to the Task Queue for processing. Once a spectrum has been sequenced, the job is removed from the task queue list, and the ● icon beside the spectrum (in the Peptide Data Frame) changes to light green ◐.
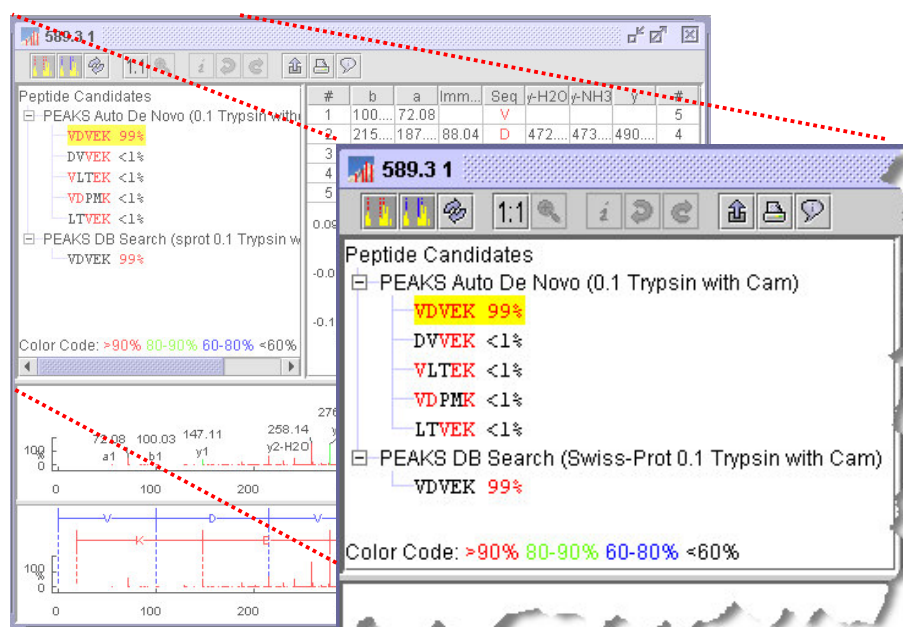
Viewing Auto de novo Results

After performing auto *de novo* on a spectrum, we may wish to see what the algorithm determined the peptide sequence to be, and review the results for ourselves. To do so, we click on the spectrum of interest in the Peptide Data Frame. This brings up the Main Processing Window for that spectrum.

The most likely peptide sequence candidate, as determined by auto *de novo*, will be automatically selected. This is found in the Peptide Candidates Frame, as the top listed candidate under 'PEAKS Auto *De novo*'. In the example above, this is the highlighted sequence VDVEK. Any modifications that have been found will be shown

abbreviated and in sequence before the amino acid residue they are associated with. *If the PTM was defined/created by another PEAKS user on another system, the PTM will still be shown and it can be imported into the local PEAKS configuration as desired.*

Right next to the proposed sequence, the auto *de novo* confidence score is shown. Positional confidences (that is, confidence that the correct residue in each position has been identified) are readily available by color coding. **Red** represents a very high confidence (greater than 90%), **Green** represents a high confidence (80 to 90%) **blue** represents a medium confidence (60 to 80%) and **black** represents a low confidence (less than 60%). For more detailed positional confidence, we can place our mouse over the sequence of interest. A Position Confidence Table will appear, showing the confidence that each subsequence is correct.

In the Ion Table frame, select a cell from the Ion Table – each cell represents an ion. This will highlight its position on an error plot (scroll the Ion Table frame down if the error plot is not visible). A point close to the centerline indicates a more confident result. We can also notice that the peak corresponding to the Ion we selected is highlighted on the Spectrum View. Select a whole column to highlight all the points for that type of ion.

Zoom in far enough and we may resolve the isotopic ladder (depending on our instrument/data).

By looking at the Spectrum View Frame, we can see the strength of the MS/MS peaks that Peaks Studio v2.4 has set as ions. The view also displays the mass of the ions at that peak and the type of ion. Click on a peak to mark it and display its information at the top left corner of the Spectrum View Frame.

Zoom in by clicking and dragging horizontally on an area of the Spectrum view. The area over which we dragged will now take up the whole spectrum view. To un-zoom, press the 🔍 undo zoom icon; or press the 1:1 1:1 icon to return to the full spectrum view.

We may also zoom in on the spectrum using the Spectrum Alignment Frame. Again, click and drag horizontally on an area of the Spectrum view. The area over which we dragged will now take up the whole spectrum view. The blue bar beneath the Spectrum Alignment view shows where we are zoomed in. The white portion of the bar represents the area that we are zoomed in on.

We can toggle whether or not we'd like to see the positions of the y-ions and b-ions (and the proposed residues in sequence between them) on the alignment view by pressing the y-ion alignment and b-ion alignment icons in the main processing window toolbar.

To view another peptide candidate, as determined by auto *de novo*, click on another peptide in the Peptide Candidates Frame and under 'PEAKS Auto *De novo*'. The information in the Ion Table will change, as will the tags on the spectrum, to reflect the selected peptide candidate's sequence.

Preparing to Edit sequencing results

We cannot change the results provided by PEAKS auto do novo or PEAKS database search. However, we can make a copy of any sequence and edit it using manual *de novo* techniques. To copy a sequence for editing:

1. Select a peptide sequence candidate from within the Peptide Candidates Frame. We can only select one peptide sequence candidate at a time.

2. Right click the mouse button while holding the mouse over that sequence. A popup menu will appear.

3. We can select the popup menu item "Copy for manual *de novo*". In this case, the sequence will be automatically placed under the 'Manual *De novo*' heading. A 'Manual *De novo*' heading will be created if there wasn't one there already.

4. Now we select our newly copied sequence, under the 'Manual *De novo*' heading, to display this sequence in the Ion Table Frame, Spectrum View Frame, and Spectrum Alignment Frame.

Now we are ready to edit the sequence, using manual *de novo* techniques.

Manual De Novo Sequencing

We can use manual *de novo* sequencing to fine tune the results of an auto *de novo* analysis, or to perform our own sequencing analysis from scratch. Peaks Studio v2.4 provides a set of tools to help us sequence a peptide, using graphic cues from the spectrum.

Creating a fresh spectrum for sequencing

We cannot change the results provided by PEAKS auto do novo or PEAKS database search. Thus, to begin manual *de novo* sequencing, we must either copy a sequenced peptide (see above section: Preparing to edit sequence results) or create a new peptide candidate for sequencing.

To create a new peptide candidate for sequencing:

1. Right click on the 'Peptide Candidates' heading, the 'Manual *De novo*', or any 'user defined type' heading. This will bring up a popup menu.

2. Select "New candidate for manual *de novo*" from the popup menu

A new candidate will be created under the 'Manual *De novo*' heading, or under the 'user defined type' heading if we selected a user defined type. The new candidate will not have been sequenced, so it will be represented by the mass difference across the spectrum (e.g. [945.15]).

Manual De novo Operations
All operations occur in the Spectrum View Frame of the Main Processing Window.

When the mouse is placed in the Spectrum View Frame, a blue (by default) bar follows the movement of the mouse. This is the *Position Bar* and it is used as a cursor for all manual *de novo* operations. The cursor's position on the m/z scale is enumerated on the top of the *Position Bar*.

Selecting a peak
To select a peak, click on it. An orange (by default) bar, called *Freeze Bar*, indicates the selected peak.

Alternatively an ion peak can be selected by clicking on its corresponding cell in the Ion Table.

Measuring distance along the m/z scale
Once a peak is selected with the *Freeze Bar*, moving the mouse left or right will display the *Position Bar* along with a value that represents the m/z difference (as an absolute value) between the selected peak (orange) and the *Position Bar* (blue). In the example below, the distance between the selected peak and the position bar is 51.02 Daltons.
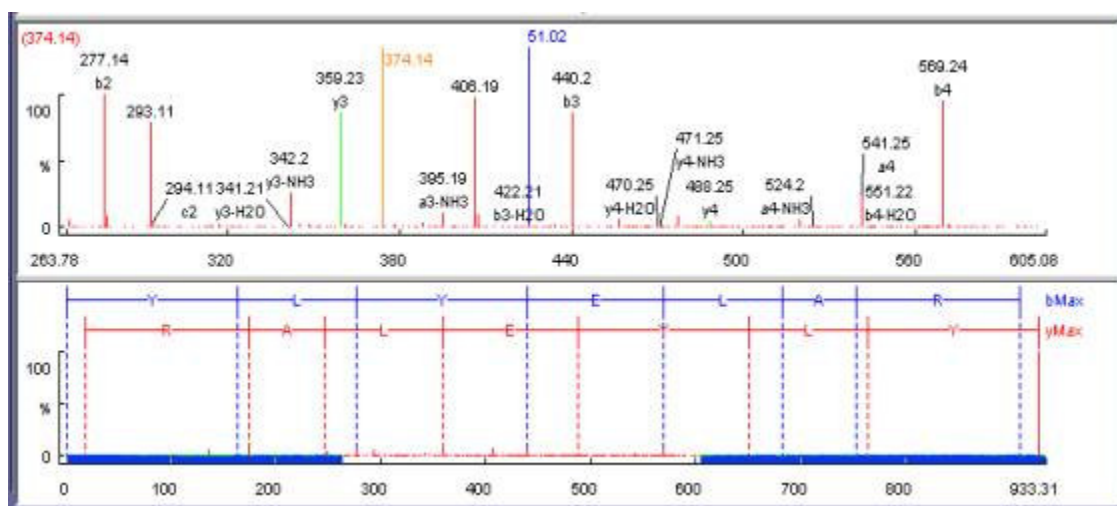
Measure the m/z difference between two PEAKS

Select a peak (orange line by default) with the *Freeze Bar*, and move the mouse to the left or right. Hold the *Position Bar* above another peak. The number above the *Position Bar* is the difference between the two PEAKS

Deselect a peak

Double click anywhere in the Spectrum View Frame

Zoom in on part of the spectrum

In the Spectrum View Frame, or the Spectrum Alignment Frame, click and drag the mouse horizontally. The selected area will be shown in the Spectrum View Frame.

Add (remove) ions to (from) a peak

Select a peak, then right click the mouse anywhere in the Spectrum View Frame. Select "Set Y Ion" from the popup menu to designate the peak as a y ion, "Set B Ion" from the popup menu to designate the peak as a b ion, Select "Ion Edit" from the popup menu to view the Ion Editor dialog box and designate the peak as another ion.
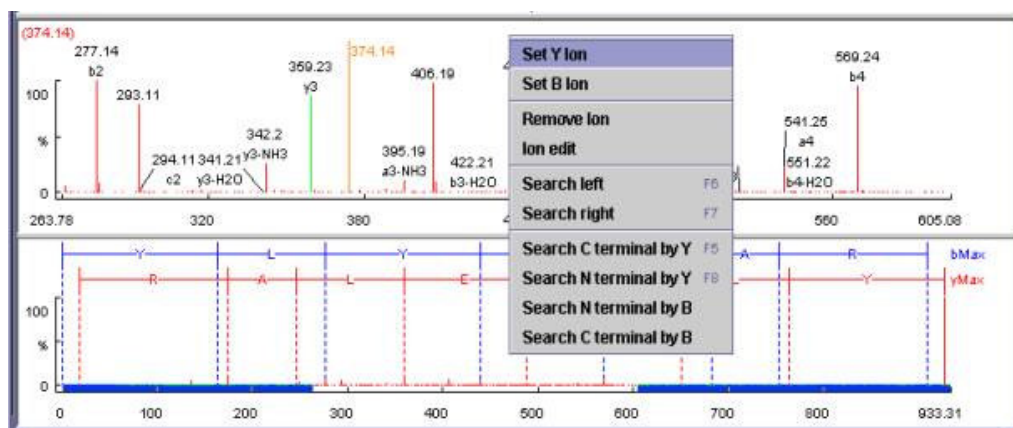
The Ion Editor dialogue allows us to add or remove ion designations to/from a peak. Select an ion from the ion choice list and press the "Add" button to add it to the selected ion list. Remove an ion from the selected ion list by selecting it and pressing the "Remove" button. We can type any comments we wish to make about the ion/peak, then press the "Apply" button to apply the changes to the selected peak.

See PEAKS Environment Preference Configuration to find out how to change the sensitivity of the residue estimate.

Two short-cut keys may also be used to label a peak. Select a peak, then hit the 'y' key to add a y-ion and or the 'b' to add a b-ion to the peak.

After setting an ion, both the alignment view and the peptide sequence candidate name (as displayed in the peptide candidate frame) will change to reflect the mass remaining to be sequenced on either side of the ion. After setting two ions, Peaks Studio v2.4 will estimate the residue found between them (if a residue corresponds closely to the mass difference). The peptide sequence candidate name (as displayed in the peptide candidate frame) will change to show the residue and the mass remaining to be sequenced on either side of the residue.

Using sequence tags
Searching the C/N terminal by Y/B: right click anywhere in the Spectrum View Frame to trigger the popup menu. From the menu, select the terminal search of interest. PEAKS will select the appropriate terminal tags and show them in the Ion Table Frame. We may test the suitability of a tag by clicking on its radio button; the tag will be shown, in position, on the Spectrum View. We may insert one or more tags by clicking on their checkboxes, then clicking the "Apply" button. Press the "Cancel" button at any time to exit the search discarding changes.

Two short cut keys can be used: 'F6' for searching the left side and 'F7' for searching the right side.

Search a sequence tag: select a peak with a defined ion (i.e. an ion that has been labeled with a peptide). Right click to trigger the popup menu, then select "Search Right" or "Search Left" to search peptide tags either to the right or left of the selected peak. PEAKS will select the appropriate terminal tags and show them in the Ion Table Frame. We may test the suitability of a tag, by clicking on its radio button; the tag will be shown, in position, on the Spectrum View. We may insert one or more tags by clicking on their checkboxes, then clicking the "Apply" button. Press the "Cancel" button at any time to exit the search discarding changes.

Undoing an edit
If we have made an error in our sequencing it is possible to undo the change. With the Peptide candidate still selected in the Peptide Candidates Frame, click the previous peptide button, to return to the previous peptide sequence. We can click this button multiple times to return to successively earlier stages in our edit.

Redoing an edit
If we have undone one too many changes, we can redo that change by clicking the next peptide button. We can click this button multiple times to proceed to successively later stages in our edit.

Protein Identification

Peaks Studio v2.4 uses peptide sequence information to filter the protein or EST database. So, to get useful protein identification results, we must first perform *de novo* sequencing on the spectrum data.
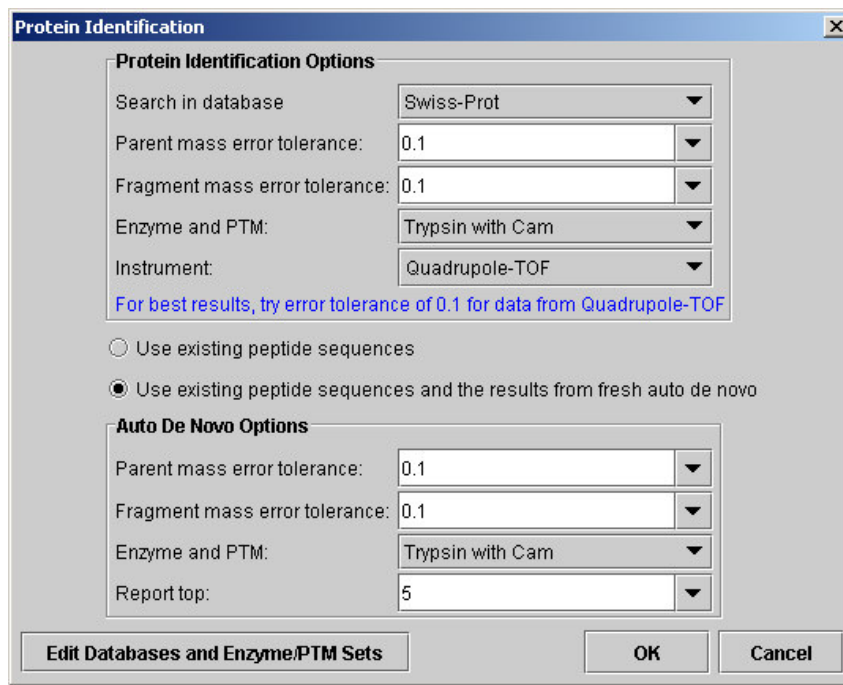
If we already have sequence information for this data, we may use this existing sequence information (manual or auto *de novo* sequences) to filter the database. If we do not have existing sequence information, or if we wish to refine our database search by providing brand new sequence information, we can ask PEAKS to perform auto *de novo* before searching the database. Brand new results will **not overwrite** any existing sequence data that we have.

1. In the *Peptide Data* Frame, we select the data file(s) that we wish PEAKS to use to identify our protein(s). This can be done by clicking on a data file's name in, the peptide data frame.

2. Click the "Protein identification" toolbar icon .

   Or

   Select "Database Search" from the tools menu

The Protein Identification Parameters dialogue window will appear:

3. If we wish to change any of the protein identification options, we do so now.

Suggested values for error tolerance appear on-screen. In most cases, we can use this number for both parent and fragment error. We should begin by using the suggested values, then try some slightly higher or lower ones to find the best result.

- Parent mass error tolerance: how much random and systematic experimental error on the parent/precursor ion mass PEAKS will account for in its database search.

- Fragment mass error tolerance: how much random and systematic experimental error on the daughter/fragmentation ion mass PEAKS will account for in its database search.

- Enzyme and PTM: enzyme that we used to digest our protein sample, and PTM that have acted on it. Protein or EST databases contain unmodified peptide sequences, so PEAKS will mathematically remove the PTM for comparison with the database.

- Search results in database: The default database is shown, but can be changed. See configuration section to find out how to change the default database.

- Use existing peptide sequences and the results from fresh Auto *de novo*: Select this radio button if we wish to generate brand new sequence information to use in filtering the database. Best practices dictate that we perform this auto *de novo* sequencing with parameters that differ slightly from the database search parameters.

Best practices for setting modifications (PTM)

The developers have discovered that database searching returns better results if the initial [auto *de novo*] portion is run with no variable PTM (perhaps one or two if necessary), but with the correct enzyme and fixed PTM. Modifications should be then turned on for the database search function. m/z tolerance can also be adjusted separately for each phase to allow us to tweak the results.

4. Press the "Ok" button to commence Auto *de novo* (if we have so chosen) and subsequent protein identification.

If we have chosen to perform auto *de novo* prior to our database search, the Auto *De novo* process will appear first in the task queue. Once this is finished the database search will begin. If PEAKS finds protein candidates after searching the database, a Protein Identification results window will appear.

Viewing Protein identification results

To view Protein identification results for a data file, we must have performed Protein Identification on that data file. If we have done so, selecting "Protein ID Result" beneath the data file from the Peptide data frame will present us with the results for that data file in the Protein Identification result window. We may also click the **jump to protein ID** link in the toolbar of the main processing window for any spectrum.

We may have performed database searching twice using the same data file. If this is the case, two sets of protein candidates will be available. They are both listed in the Database frame and can be identified by the name of the database we searched and the parameters we used when searching the database. In the example below, we searched through the Swiss-Prot database twice, once using error of 0.1, tryptic digest and Cam as our PTM, once using error of 0.1, an unknown enzyme and no PTM. Clicking on either one brings up its results. If we've only searched once, there will only be one database to choose from and its results will be automatically displayed.
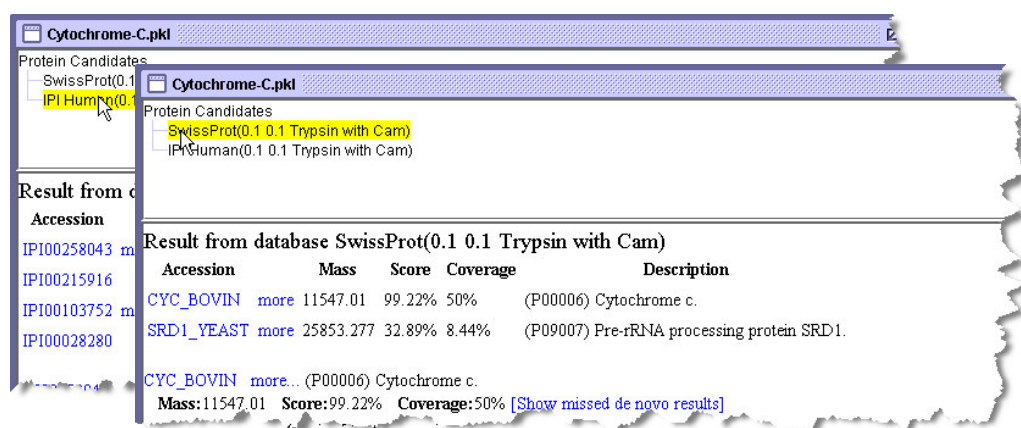
A note on differentiating between results

Peaks Studio v2.4 differentiates between protein identification runs by the database search parameters that were used, not the auto *de novo* parameters. As such, if we do not change database search parameters before searching, we will overwrite whatever previous search results we had – no matter how much we alter the auto *de novo* parameters. If this presents a problem, please contact technical support.

After selecting the results to view, the protein candidates Peaks has identified will be shown in the index section at the top of the report. Very similar proteins are grouped together. To see the full list of proteins within each grouping, click the **more** hyperlink. Another window will appear, showing all the related proteins:

This window is equivalent to, and behaves the same way as, the Search Result Report Frame (of the Protein Identification Result Window) but lists only those proteins identified within that group. The following paragraphs describe the use of both.

Clicking on the Accession number hyperlink (in the index at the top) will scroll down to section of the peptide match report that corresponds with that Accession number. In this section of the report, the spectra that PEAKS found to match that protein are displayed, each with a score, and ordered by mass. It is possible that not all spectra from a data file are shown. Only the ones that matched the protein show up. **More spectra matching with higher confidence scores indicate better probability of having identified the correct protein.**

CYC_BOVIN (P00006) Cytochrome c.
  Mass: 11547.01   **Score:** 99.22%   **Coverage:** 50% [Show missed de novo results]

Mz	Charge	Mr(calc)	Start	End	Score	Peptide
482.7	2	945.5171	92	99	99%	EDLIAYLK
584.8	2	1149.6042	28	38	99%	TGPNLHGLFGR
634.4	1	615.3744	9	13	93.58%	IFVQK
678.3	1	659.36426	74	79	84.58%	YIPGTK
728.8	2	1437.6525	40	53	99%	TGQAPGFSYTDANK
779.4	1	760.43054	80	86	86.4%	MIFAGIK
792.9	2	1565.7474	39	53	97.83%	KTGQAPGFSYTDANK

[Return to top]

While viewing this section of the report, click on a Peptide sequence hyperlink to bring up the Main Processing window associated with that spectrum and peptide sequence.

Or, click on the **more** hyperlink to again bring up the list of related proteins (see above). Or, click on the protein's accession number to access the protein view:

This protein view window shows the peptide matches for this protein in red. Clicking on the accession number in this window links to the NCBI Blast page and inputs search parameters for the protein. If we click the "View with web browser" button at the top of this window, PEAKS loads this same report into our default web browser in html format for print or incorporation into reports.

Please see the walkthrough in chapter 3 for an example of viewing protein identification results.

Manipulating Data Files

Editing Precursor information

It is possible that the precursor information, as listed in the Peptide Data Frame, is incorrect. If the charge listed is wrong, or if the m/z listed is even slightly incorrect (more than 0.1 Daltons, depending on the accuracy selected) it could really affect the quality of the results. In this case it is imperative that we change the precursor information. The change will only affect the ANN file we are working on.

To edit precursor information, select a spectrum by clicking on its name, then right click the mouse while holding it in position over the name. A small menu will appear. Click on "Edit Precursor".

In the dialogue that follows, type the new precursor information into the appropriate textboxes. Click the "Apply" button when finished to apply the changes. Click the "Cancel" button to exit, discarding changes.
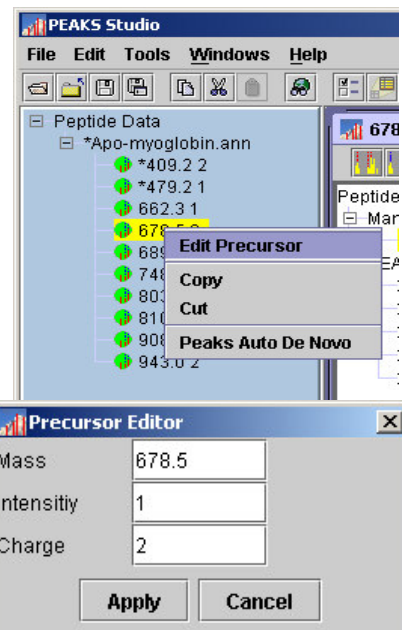
The precursor information will be updated, reflected by a change in the name of the spectrum in the Peptide Data Frame. A * will also appear in front of that name, indicating that there is unsaved information pertaining to that spectrum.

Cutting and Copying Spectrum Data

Changes made to the original spectrum after duplication will not affect the duplicated spectrum.

If we wish to move spectrum data from one data file to another we may do so by copying and pasting it (see below for pasting instructions). Also, we may wish to make a copy of the spectrum in the same data file in order to re-sequence an individual spectrum using different preferences. Cutting spectrum data will remove it completely until pasted. Copying spectrum data will duplicate the spectra when pasted. To cut/copy spectrum data:

Select a spectrum by clicking on its name (select multiple spectra by holding down the 'control' key and clicking on any number of spectrum) in the Peptide Data Frame.

Right click on one of the selected spectra. A small pop-menu will appear. Select "Cut" or "Copy"

OR

Click the "Copy" button or "Cut" button in the main toolbar.

Copied/Cut items will remain on the clipboard until replaced by another copied/cut item. **Warning: unless pasted, a cut item will be lost, as subsequent cut/copied items will displace it from the clipboard.**

Pasting Spectrum Data

After having copied or cut spectrum data, we would like to paste it into another data file, or the same data file. To paste spectrum data:

1. Select the spectrum data file into which we wish to paste the spectrum (or spectra) by clicking on its name in the Peptide Data Frame. *You may only choose to paste into one spectrum data file at a time.*

2. Right click on one of the selected spectra. A small popup menu will appear. Select "Cut" or "Copy" from the popup menu.

    OR

    Click the "Paste" button in the main toolbar.

The pasted spectra will appear in the Peptide Data, under the spectrum data file into which we pasted.

Running protein identification on select spectra

When searching our dataset against a particular database, Peaks may not have found a hit for certain spectra. If these are good data, we may wish to try searching them against a more general database. Before we do so, we must **create a new data set** with these "good spectra that did not match". This is essential so that we can organize our data well, and because Peaks will only run Protein ID on all the spectra in a data node. To create a new data node:

1. *Make a new Data node by right clicking on the peptide data node. The new node appears as "Data1".*

2. *Select the relevant spectra using <shift>+click and <ctrl>+click, then press the copy button.*

3. *Click the new node and press the paste button. Pressing the + next to Data1 will expand it and reveal the pasted spectra*
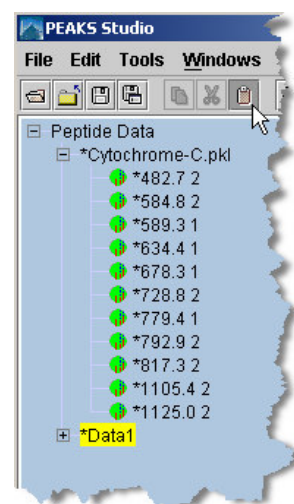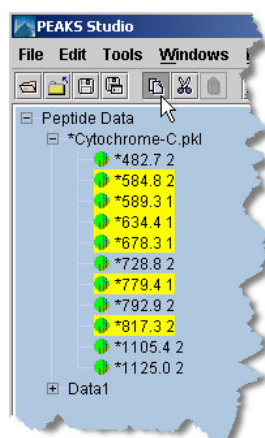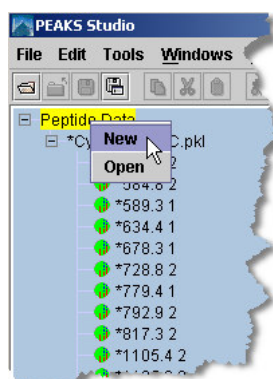
Now we can run protein identification on the new data set. We can save that dataset in a new file, or any of the other functions that apply to regular nodes. Make sure the new node is selected before running protein ID (or any other function) on it.

Using the Mass Calculator

The mass calculator is a simple tool to help us determine the molecular weight of a peptide. To access the mass calculator, open the "Tools" menu and click "Mass Calculator". The mass calculator will appear .
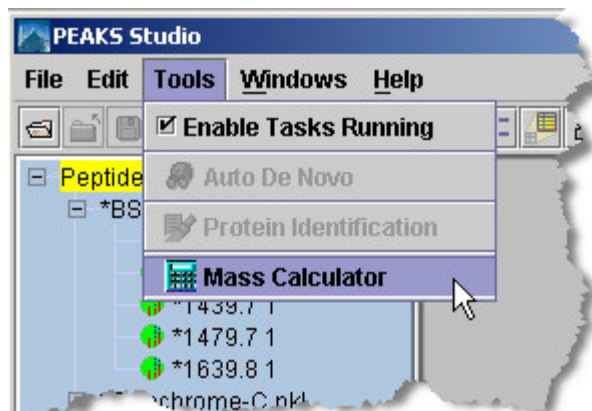
We can also load the mass calculator outside of Peaks, and separately. To access the mass calculator without having to load PEAKS, click on the mass calculator's icon in the start menu. It will appear in the same program group as PEAKS Studio.

When using the mass calculator, remember to start with water.
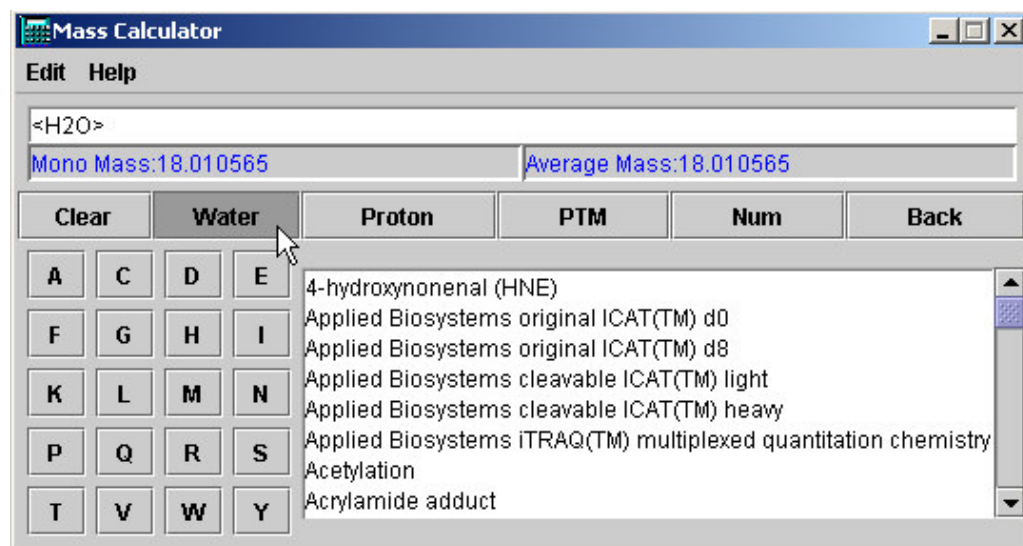
We can click any of these buttons multiple times to repeatedly add that mass.

- Amino acids are represented by their single letter symbols. Clicking on an amino acid's button will add it to the 'sequence' above, and add its mass to the mass of the peptide. Note that the peptide's monoisotopic and average masses are both computed.

- Add a Proton by clicking the "Proton" button. It will be represented by an H in the 'sequence' above.

- To compute the mass of the peptide as if it had been modified, select a PTM from the list, and press the "PTM" button to apply them to the peptide.

- If the PTM we wish to add does not appear in the list, we may wish to enter it's mass manually. To add a mass numerically, click the "Num" button and enter a numeric value in the dialogue box that appears. Press the "OK" button on the dialogue and the mass will be added to the sequence.

- To remove a mass that we've just added to the peptide, press the "Undo" button.

Saving Results

Saving results will preserve our work for later use. Saving files in PEAKS's .ANN format will preserve spectrum data, manual *de novo* sequence information, automatic *de novo* sequence information, protein identification results, and information about any PTM that were found in sequence.

To save the results of our analysis, we first select the data file we wish to save in the Peptide Data Frame.

To save, click the icon in the main window toolbar, select "Save" from the "File" menu, or right click on the data file and select "save" from the popup menu. This will save the processed spectra in ANN format and of the same name as the data file we opened.

To change the name of the ANN file, choose "Save as" from the "File" menu, or right click on the data file and select "Save as" from the popup menu. You may then change the file name.

To save all currently opened data files, select "Save all" from the "File" menu.

To export data to a PKL file, we select the data file (not an individual spectrum) to export. Then, from the "File" menu, select "Export", then "Export PKL File". The spectrum data will be saved in PKL format, but all sequencing and protein data will be lost.

To export peptide sequencing results to a FASTA format file, select the data file (not an individual spectrum) to export. Then, from the "File" menu, select "Export", then "Export Peptide Sequence". The sequencing data will be saved in FASTA format, but will not retain any spectrum data.

To export results to an HTML file, select the data file (not an individual spectrum) to export. Then, from the "File" menu, select "Export", then "Export HTML File". Peaks will then ask us which results we would like to export. We can choose from any *de novo* sequencing or protein ID run we have done. Each will be listed with the parameter set we used.

Frequently Asked Questions

The PEAKS FAQ online may contain more answers. Check:
www.bioinformaticssolutions.com/faqs/peaks.php

1. **What is .ANN format?**

The ANN format is annotated data file in XML format. PEAKS uses it to save MS/MS information and peptide information. It supports both single MS/MS spectrum data and multiple spectra data. PEAKS accepts and saves file(s) in ANN format.

Saving files in PEAKS's .ANN format will preserve spectrum data, manual *de novo* sequence information, automatic *de novo* sequence information, protein identification results, and information about any PTM that were found in sequence.

ANN uses an ANN data file and an ANN index file (both of them use the same extension 'ann'). Each ANN data file contains the MS/MS information and peptide information from one spectrum. An ANN index file can be used to organize multiple spectra. The ANN index file links to a directory that contains multiple ANN data files. The directory has the same name as the ANN index file but with 'dir' as the extension instead of 'ann'.

2. **What is the required system configuration for PEAKS?**

PEAKS can be run on any computer that supports Sun's Java Runtime Environment (JRE) 1.4. On installation, PEAKS will install a dedicated JRE for its use, so it can co-exist with another version of Java on our machine. The system should also have: 1024MB RAM, 1GB free space on hard drive.

3. **I started an auto *de novo* process, but nothing happened. What did I do wrong?**

First, check the Tools menu and verify that Enable Tasks Running is checked. If it is disabled, the 'Task Queue' will be shown in red and no tasks in the queue will be processed.

It is possible that the process is running but you just don't notice. PEAKS processes the auto *de novo* task in the background so you can continue to work. This process takes several seconds. You will find the job is still in the queue if you check the Task Queue Frame. After the process is done, it will disappear from the queue. By selecting the spectrum in the Peptide Data Frame, you can find the peptide candidates, the spectrum image and ion alignment in the main process window.

4. Can I edit the data file manually in PEAKS?

You can edit the precursor information (m/z and charge) by right clicking on the precursor in the Peptide Data Frame. However, you cannot edit the spectrum data itself from within PEAKS.

5. What enzymes should I use to digest the protein, in order to use PEAKS to interpret the MS/MS data?

The most popular enzyme for digesting proteins for MS/MS analysis is Trypsin. PEAKS comes with an Enzyme/PTMs set, predefined for unmodified tryptic digests, to handle this common case. Tryptic peptides typically show excellent MS/MS spectra, and produce good sequences. If you wish to use a different enzyme, or sequence small peptides in entirety, you can create a new Enzyme/PTMs set and select another enzyme. Selecting an 'unknown enzyme' places no restrictions on the residues appearing at the C-terminal end. You may set them yourself.

6. Can I edit/modify the result of PEAKS Auto *De novo*?

You cannot modify the sequences returned by PEAKS Auto *De novo* search. However, you can copy the sequence for manual *de novo* to achieve the same goal. Right-click on the desired sequence and select Copy for manual *de novo*. You can now edit the sequence and ion assignments in the Manual *De novo* section of the Peptide Candidates tree.

7. Why can I not find the Freeze bar to indicate the position of the peak in the ion edit window when I select an ion in the ion table?

The spectrum is zoomed in to far, or not in the right area. Adjust your spectrum view's zoom to 1:1.

8. How can I save the sequences resulting from auto *de novo*?

You can copy the predicted sequences by right clicking on the sequences, and select copy. This will copy the sequence of letter symbols to the clipboard. You may then paste them into a word processor or text editor.

You can print images by clicking the right-most 'Print result image' button at the top of the main process window. Or you can save images by clicking the 'Export results as

image' button. The printed or saved images look slightly different to those displayed on the screen.

9. **What is the difference between save data file and save all files?**

Save data file: If current data file or the sequencing result of the selected data file(s) has been changed, PEAKS will save the data file and the sequence information in .ann format.

Save all files: If there is more than one data file loaded, using "save all files" will save all the data files and their sequence information in their respective .ann format files. There is no need to select any files.

10. **We saw a green triangle on the m/z axis of the spectrum. What does that represent?**

This green triangle is used to shown the location of the one charged precursor.

11. **Why I cannot delete a task from the Task Queue?**

There are 2 steps to remove a task from the Task Queue:

Disable the task running by clicking on "Enable task running" from the "Tools" menu to deselect it.

Right-click the task in the *Task Queue*. From the popup menu, choose "remove".

12. **Where can I get the demo version of PEAKS?**

The demo version of PEAKS can be found at BSI's web site. You can either download the trial version or try the on-line version from our web site.

13. **Where can I get the log file for PEAKS?**

The log file can offer us some information for bug tracking. PEAKS sets the location of the log and configuration file, users can not change it. Its location is listed in the Environment Configuration Dialoged, under the Environment tab.

14. **What are the differences between PEAKS and Pepseq?**

PEAKS and Micromass Pepseq (part of MassLynx, ProteinLynx) both do *de novo* sequencing. The difference lies in the quality of the results.

A poster in ASMS this year, by Terry Cyr, has vividly described the superior quality of PEAKS results: he used 5 spectra from Micromass machines. Pepseq was able to

identify just 1 spectrum (the shortest, 5 amino acids), and badly failed on all other 4. PEAKS completely identified 3 peptides fully (every single amino acid residue was correctly identified in these sequences), and partially identified the other two peptides.

Program Defaults

These are the default settings for a few of PEAKS parameters. Many default settings can be changed.

Environment

Default Data Input Directory: last used directory

Default Data Output Directory: last used directory

Automatically close previous open spectrum when opening another one: no

Spectrum Colors

Spectrum Peak: Red

Frozen Peak: Orange

Current Peak: Blue

Y_Alignment: Red

B_Alignment: Blue

Y_Ion: Green

Auto De Novo

m/z tolerance: 0.1

Residue and PTM: Trypsin without PTM

Report top: 5 candidates

Source: ESI

Instrument: Quadrupole Time of Flight (QTOF)

Use Auto *de novo* and database search configuration as default: no

Database Search phase of Protein ID

m/z tolerance: 0.1

Residue and PTM: Trypsin with Cam

Instrument: Quadrupole Time of Flight (QTOF)

Manual de Novo

Maximum tag length: 3

Maximum return: 5

Default machine error: 0.1

File types

Default save file type: ANN index file linked to a directory of the same name, containing ANN data files.

About Bioinformatics Solutions Inc.

BSI *provides advanced software tools for analysis of biological data.*

Bioinformatics Solutions Inc. develops advanced algorithms based on innovative ideas and research, providing solutions to fundamental bioinformatics problems. This small, adaptable group is committed to serving the needs of pharmaceutical, biotechnological and academic scientists; and to the progression of drug discovery research. The company, founded in 2000 in Waterloo, Canada, comprises a select group of talented, award-winning, and intelligent developers, scientists and sales people.

At BSI, groundbreaking research and customer focus go hand in hand on our journey towards excellent software solutions. We value an intellectual space that fosters learning and an understanding of current scientific knowledge. With an understanding of theory, we can focus our talents on providing solutions to difficult, otherwise unsolved problems that have resulted in research bottlenecks. At BSI, we are not satisfied with a solution that goes only partway to solving these problems; our solutions must offer something more than existing software.

The BSI team recognizes that real people will use our software tools. As such, we hold in principle that it is not enough to develop solely on theory; we must develop with customer needs in mind. We believe the only solution is one that incorporates quality and timely results, a satisfying product experience, customer support and two-way communication. So then, we value market research, development flexibility and company-wide collaboration, evolving our offerings to match the market/user's needs.

Efficient and concentrated research, development, customer focus and market analysis have produced: PEAKS software for protein and peptide identification from tandem mass spectrometry data, RAPTOR and PROSPECT Pro software for threading based 3D protein structure prediction, and PatternHunter software for all types of homology search sequence comparison.

Chapter

8

PEAKS Software License

This is the same agreement presented on installation. It is provided here for reference only.

If you are evaluating a time limited trial version of PEAKS, and you wish to update the software to the full version, you must purchase PEAKS and obtain a full version registration key.

1. License. Subject to the terms and conditions of this Agreement, Bioinformatics Solutions (BSI) grants to you (Licensee) a non-exclusive, perpetual, non-transferable, personal license to install, execute and use one copy of PEAKS (Software) on one single CPU at any one time. Licensee may use the Software for it's internal business purposes only.

2. Ownership. The Software is a proprietary product of BSI and is protected by copyright laws and international copyright treaties, as well as other intellectual property laws and treaties. BSI shall at all times own all right, title and interest in and to the Software, including all intellectual property rights therein. You shall not remove any copyright notice or other proprietary or restrictive notice or legend contained or included in the Software and you shall reproduce and copy all such information on all copies made hereunder, including such copies as may be necessary for archival or backup purposes.

3. Restrictions. Licensee may not use, reproduce, transmit, modify, adapt or translate the Software, in whole or in part, to others, except as otherwise permitted by this Agreement. Licensee may not reverse engineer, decompile, disassemble, or create derivative works based on the Software. Licensee may not use the Software in any manner whatsoever with the result that access to the Software may be obtained through the Internet including, without limitation, any web page. Licensee may not rent, lease, license, transfer, assign, sell or otherwise provide access to the Software, in whole or in part, on a temporary or permanent basis, except as otherwise permitted by this Agreement. Licensee may not alter, remove or cover proprietary notices in or on the Licensed Software, or storage media; or use the Licensed Software in any unlawful manner whatsoever.

4. Limitation of Warranty. THE LICENSED SOFTWARE IS PROVIDED AS IS WITHOUT ANY WARRANTIES OR CONDITIONS OF ANY KIND, INCLUDING BUT NOT LIMITED TO WARRANTIES OR CONDITIONS OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. LICENSEE ASSUMES THE ENTIRE RISK AS TO THE RESULTS AND PERFORMANCE OF THE LICENSED SOFTWARE.

5. Limitation of Liability. IN NO EVENT WILL LICENSOR OR ITS SUPPLIERS BE LIABLE TO LICENSEE FOR ANY INDIRECT, INCIDENTAL, SPECIAL, OR CONSEQUENTIAL DAMAGES WHATSOEVER, EVEN IF THE LICENSOR OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE OR CLAIM, OR IT IS FORESEEABLE. LICENSOR'S MAXIMUM AGGREGATE LIABILITY TO LICENSEE SHALL NOT EXCEED THE AMOUNT PAID BY LICENSEE FOR THE SOFTWARE. THE LIMITATIONS OF THIS SECTION SHALL APPLY WHETHER OR NOT THE ALLEGED BREACH OR DEFAULT IS A BREACH OF A FUNDAMENTAL CONDITION OR TERM.

6. Termination. This Agreement is effective until terminated. This Agreement will terminate immediately without notice if you fail to comply with any provision of this Agreement. Upon termination, you must destroy all copies of the Software. Provisions 2,5,6,7 and 10 shall survive any termination of this Agreement.

7. Export Controls. The Software is subject at all times to all applicable export control laws and regulations in force from time to time. You agree to comply strictly with all such laws and regulations and acknowledge that you have the responsibility to obtain all necessary licenses to export, re-export, or import as may be required.

8. Assignment. Customer may assign Customer's rights under this Agreement to another party if the other party agrees to accept the terms of this Agreement, and Customer either transfer all copies of the Program and the Documentation, whether in printed or machine-readable form (including the original), to the other party, or Customer destroy any copies not transferred. Before such a transfer, Customer must deliver a hard copy of this Agreement to the recipient.

9. Maintenance and Support. BSI will provide technical support for a period of thirty (30) days from the date the Software is shipped to Licensee. Further maintenance and support is available to subscribers of BSI's Maintenance plan at BSI's then current rates. Technical support is available by phone, fax and email between the hours of 9 am and 5 pm, Eastern time, excluding statutory holidays.

10. Governing Law. This Agreement shall be governed by and construed in accordance with the laws in force in the Province of Ontario and the laws of Canada applicable therein, without giving effect to conflict of law provisions, and without giving effect to United Nations Convention on contracts for the International Sale of Goods.