Bioinformatics Solutions Inc.

PEAKS Studio 4.0
User Manual

*For use with the PEAKS Studio
de novo sequencing &
protein identification software*

www.BioinformaticsSolutions.com

BIOINFORMATICS SOLUTIONS INC

PEAKS Studio 4.0 User's Manual

Introduction

Introduction to PEAKS Studio 4.0

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 4.0 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.

How to use this user's manual

This user's manual is intended to help us get started using PEAKS Studio 4.0, 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 modification 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.

ANZ file: a PEAKS 'zip-compressed' XML-based Annotated spectrum file format. ANZ files preserve all the information from the PEAKS session.

ANN data file: within the ANZ file, a folder contains ANN data files that store the MS/MS information and peptide information of one spectrum.

ANN index file: within the ANZ file, is one compressed file used to organize the data; the ANN index file links to a directory containing multiple ANN data files.

Getting started with PEAKS Studio 4.0

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 4.0. 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 4.0 package should contain:

- This manual

- PEAKS Studio 4.0

System requirements
PEAKS Studio 4.0 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 1.5 or better.

Instrumentation
PEAKS Studio 4.0 will work with data from any type of tandem mass spectrometer designed for proteomics work.

PEAKS will accept data in the following formats:

Thermo Electron instrument's data in RAW format provided that PEAKS and XCalibur are installed on the same computer, or dta format (and concatenated dta formats), with the ability to load an entire folder full of dta's.

Waters/Micromass instrument's data in .RAW format, provided that PEAKS and MassLynx are installed on the same computer, or pkl files.

mzXML is a standard data format from the Sashimi Project. It is an XML based format.

Applied Biosystems instrument's data in .wiff format provided that PEAKS, the Infochromics converter plug-in and Analyst are installed on the same computer. PEAKS has the ability to read directly from the 4700/4800 Oracle database.

All other instrument's data as can be converted into mzXML, .pkl, dta, or .mgf.

Installation

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

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

2. Insert the PEAKS Studio 4.0 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. Click on the exe file.

4. A menu screen will appear with the title "PEAKS Studio 4.0". Select "Install Peaks Studio".

5. The installation utility will begin the install. Wait while it does so. Choose 'English' as the language for installation instructions. When the "PEAKS Studio 4.0" 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 folder/directory in which we'd like to install PEAKS Studio 4.0. Press the "Choose…" button to browse our system and make a selection, or type a folder name in the textbox. Click "Next".

8. Choose where we'd like to place icons for PEAKS Studio 4.0. The default will put these icons in the programs section of our start menu. Click "Next".

9. 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.

10. PEAKS Studio 4.0 will now install on our system. We may cancel at any time by pressing the "Cancel" button in the lower left corner.

11. When installation is complete, click "Done". The "PEAKS Studio 4.0" menu screen should still be open. One 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. Follow the on-screen instructions.

Database Configuration

In addition to *de novo* sequencing of peptides, PEAKS Studio 4.0 also has the ability to search through a database search to identify
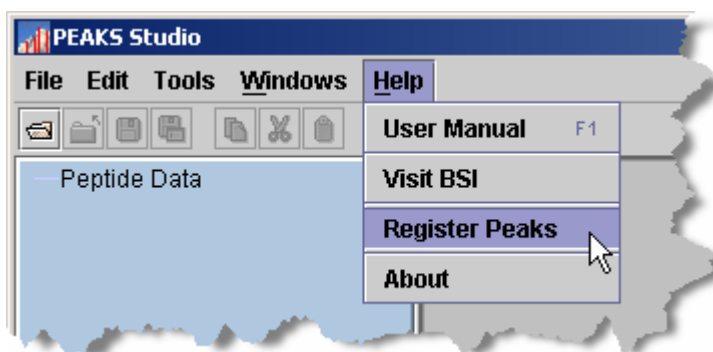
proteins. But in order to use this function, PEAKS must have access to a protein or EST database in FASTA format or an EST database of DNA sequences. We can point PEAKS to an existing database on our system, or download one. Additionally, we can associate taxonomy with certain databases. This is database 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 4.0. 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"
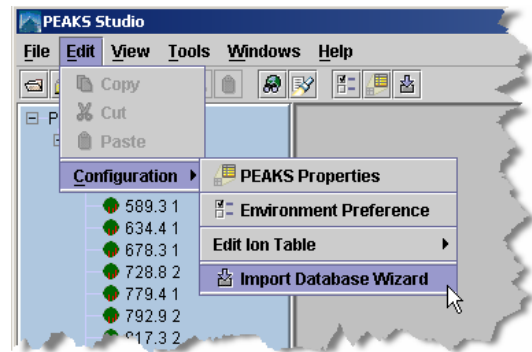
Microsoft Internet Explorer, for example, is an FTP client. We may use Internet explorer and the provided URL to download a database.

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"

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

The next section provides a walkthrough of PEAKS Studio 4.0's main functionality using the NCBI nr database.
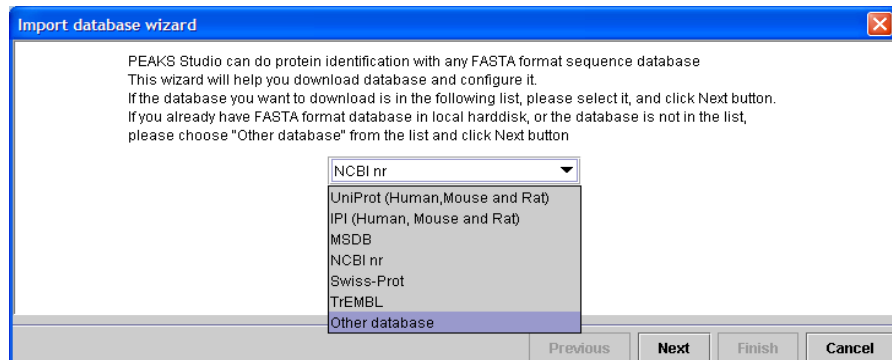
we wish to use, click the provided link to begin downloading. A dialogue box will appear with instructions on downloading using file transfer protocol (FTP). It does not matter where we put the download file, but we must remember where it is.

A note on downloading databases

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/
```

Best practices: configuring databases for use with X!Tandem

At the time of this writing, X!Tandem had trouble searching through large databases, and would crash. It is therefore suggested that X!Tandem only be used with small databases; or if used with a large database, a taxon should be specified. The NCBI nr and Swiss Prot databases are ideal for this purpose.

Best practices: configuring databases for use with OMSSA

At the time of this writing, we could not use OMSSA with databases that were not in NCBI format, or Swiss-Prot format, and have those results available to inChorus.

Also, a bug in OMSSA prevents us from easily using databases with OMSSA when they are stored in a folder that contains a space in its path. This creates problems when PEAKS creates temporary databases on our behalf. To avoid this, best practices suggest we put all our databases in a folder "c:\peaksdatabases". The folder "c:\my documents\databases" wouldn't work because it contains a space between 'my' and 'documents'. Using spaces in the database file name causes the same problem. So after we download and extract our database we should call the database file "ncbinr.fas", or "ncbi_nr.fas" rather than "ncbi nr.fas".

5. The database we downloaded may be in a compressed file, perhaps a .zip or a .gz file. We must find the file and use a decompression utility, such as WinZip,

or WinRar to extract its contents. The file inside the compressed file will be a FASTA format text file (a .fas or a .fasta file).

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.

The taxonomy options are only available if the NCBI nr database is selected and the Apply button has been pressed, or NCBI nr was selected on a previous screen.

- PEAKS will ask us to enter the database nickname. This is a nickname that we chose to represent the database we are configuring. It doesn't matter what name we enter, but 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 (see step 5)**.

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 our database is an EST database containing DNA sequences, check the 'EST database' checkbox.

- 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 the same as 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.

- If we are configuring the NCBI nr database or the Swiss-Prot database, we may choose to point PEAKS Studio 4.0 to the location of the taxonomy files associated with that database. Under "Taxonomy Options" we must type the location of the taxonomy files, or click browse to find the file on our system. If we do not specify these taxonomy files, we will not be able to limit our database search to a specific taxon. We can use the compressed (.zip or .gz) files; **no decompression is required for the taxon files**.

A note on choosing the taxonomy files for NCBI nr

At the time of printing, the gi_taxid file was called: *gi_taxid_prot.dmp.gz* and the taxdmp file was called: *taxdmp.zip*. Select these files.

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

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

Trouble shooting

Some problems with a database may not appear until we run a search. While PEAKS is quite tolerant of format errors in databases, other search engines called from the inChorus tool may not be. If there is an error in the search, it will be reported in a summary screen after the work has finished. If there is a problem, check the best practices outlined in this section. If the problem persists, it is possible that the database download was corrupted; try downloading again. Please contact technical support for help.

Features Walkthrough

Let's familiarize ourselves with PEAKS

This section of the manual will walk us through most of the basic functionality of PEAKS Studio 4.0. After completing this section we will have seen how easy it is to load and view a data file, perform *de novo* sequencing, and database search protein identification.

Begin the walkthrough

Run PEAKS Studio 4.0, then download and configure the NCBI nr database. The procedures for doing so are outlined in the previous section.

The demo sample data should load automatically on startup under the heading OrbiOrbi.pkl. If it is not loaded, open the data file 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\" folder. Load the file "OrbiOrbi.pkl" by clicking on it, then clicking "Open.".

The data file will appear in the left-hand frame. Make sure "OrbiOrbi.pkl" (i.e. the data file) is selected.

See the section entitled "Merging Spectra" for more help.

In the "Tools" menu, select "Protein Identification". The protein identification options dialogue will appear.

After entering the settings, as shown, click the "Save As" button to save these parameters for future use. When prompted, type OrbiStandard and press enter. Click "OK" to commence analysis. Analysis will be initialized (most of this time is spent creating a partial database, which only has to be done once) this takes one or two minutes.

After this, *de novo* sequencing will commence. For this sample it takes just over a 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). 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 4.0. PEAKS uses a unique sequence tag plus peptide fragment fingerprinting approach to protein identification.
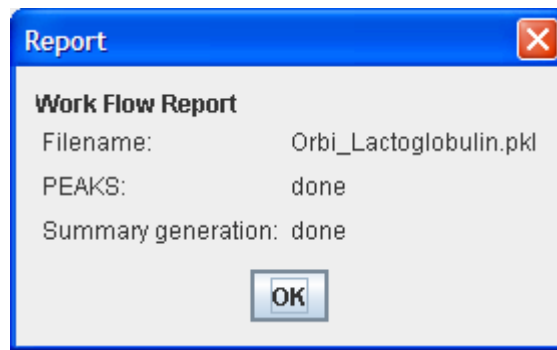
After the search is finished, the protein identification results will appear on screen. The 'Peptide View' is displayed by default. The display shows each spectrum for which PEAKS found a matching peptide. The spectra are grouped/sorted by index number. Since a spectrum may match to more than one peptide, there may be more than one entry per spectrum. The list is sort-able; click the heading on each column to experiment with sorting by score, by mass, etc.

Click the 'Protein View' tab. PEAKS Studio 4.0 presents a list of proteins that it believes to be the best match for the sample. The top section is an index, listing them by accession number, ranked in descending order from highest score on downward.

Accession	Mass	Score(%)	Coverage	Query matched	Marked	Description
PEAKS DB Search					☐	
gi|229460	18,367	100	61.73%	18	☐	lactoglobulin beta
gi|28972720	78,207	29	0.84%	1	☐	mKIAA1321 protein [Mus musculus]

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.

The correct protein, Lactoglobulin beta, is shown at the top of the list and with high score. Since one cannot distinguish between the different forms of Lactoglobulin Beta, PEAKS Studio 4.0 groups them all together, thus avoiding cluttering the report. Click the plus sign next to gi|229460 for a listing of other possible lactoglobulin. The peptides matching these homologues will be the same set or a subset of Cytochrome c matches. Collapse this list of homologues by clicking the minus sign next to gi|229460.

The listing, as shown above, is simply an index. We will find this useful, in the future, when dealing with complex mixtures. Clicking any protein's gi number will display the peptides matched to that protein in the bottom pane.

NCBI BLAST search of gi|229460
Links to retrieve entries containing this sequence from NCBI Entrez:

Accession/ID **Description**

gi|229460   lactoglobulin beta

Peptides List

Mz	Charge	Mr(calc)	Start	End	Score(%)	Search Engines	Peptide
452.28497	2	902.5589	76	83	67	p-----	TKIPAVFK
458.73856	2	915.4661	84	91	69	p-----	LDAINENK
467.2729	2	932.5366	1	8	33	p-----	LIVTQTMK
533.2925	2	1,064.5752	92	100	94	p-----	VLVLDTDYK
545.92804	3	1,634.7673	125	138	99	p-----	TPEVDDEALEKFDK
597.3386	2	1,192.6702	92	101	99	p-----	VLVLDTDYKK
623.2947	2	1,244.5771	125	135	99	p-----	TPEVDDEALEK
673.38513	1	672.3807	9	14	19	p-----	GLDIQK
674.4233	1	673.4163	78	83	74	p-----	IPAVFK
697.7167	3	2,090.1260	84	101	98	p-----	LDAINENKVLVLDTDYKK
771.75635	3	2,312.2517	41	60	99	p-----	VYVEELKPTPEGDLEILLQK
818.3876	2	1,634.7673	125	138	99	p-----	TPEVDDEALEKFDK
903.1304	3	2,706.3687	15	40	65	p-----	VAGTWYSLAMAASDISLLDAQSAPLR
903.5675	1	902.5589	76	83	3	p-----	TKIPAVFK
933.5426	1	932.5366	1	8	34	p-----	LIVTQTMK
1065.586	1	1,064.5752	92	100	5	p-----	VLVLDTDYK
1157.1289	2	2,312.2517	41	60	99	p-----	VYVEELKPTPEGDLEILLQK
1354.1898	2	2,706.3687	15	40	90	p-----	VAGTWYSLAMAASDISLLDAQSAPLR

Matched peptides shown in Red

1	LIVTQTMKGL DIQKVAGTWY SLAMAASDIS LLDAQSAPLR VYVEELKPTP
51	EGDLEILLQK WENDECAQKK IIAEKTKIPA VFKLDAINEN KVLVLDTDYK
101	KYLLFCMENS AEPEQSLVCQ CLVRTPEVDD EALEKFDKAL KALPMHIRLS
151	FNPTLQEEQC HI

Above, 18 of the original 22 spectra indicated a peptide sequence matching with Lactoglobulin Beta. Each peptide match shows a high confidence: strong evidence for having found the correct protein. We can also see exactly where the peptide fits into the protein sequence; with the matching sequences highlighted in red at the bottom.

As mentioned above, the peptide sequence results are based on a database search guided by an initial *de novo* sequencing analysis. Let's see how the *de novo* sequencing was able to help. Click on the **467.2729** hyperlink. This will bring up the main processing window for spectrum 467.2729 2. 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 partial peptide sequence LLVXXTMK, but was not as sure of the middle two

467.2729 2 [Main processing window]

1:1

Peptide Candidates
- PEAKS Auto De Novo (0.01 0.01 TEMP)
  - LLVTQTMK 49.2%
  - LLVQTTMK 49.2%
  - LLVASATMK <1%
  - LLVSAATMK <1%
  - LLVAASTMK <1%
- PEAKS DB Search (NR 0.01 0.01 TEMP)
 - LIVTQTMK 33.4%
  - IIVTQTMK 33.4%

Color Code: >90% 80-90% 60-80% <60%

residues. The PEAKS DB Search was able to confirm this result, returning the peptide: LLVTQTMK.

Selecting another spectrum from the Peptide Data frame (left), e.g. 545.928 3, will allow us to view the results from that spectrum without having to return to the protein identification result. Click on the time and date stamp beneath the filename to return to the report.

Let's try another kind of search. This time we'll use inChorus database searching – this technology, unique to PEAKS, allows us to launch other search engines that will help improve the results. The best confirmation of results comes from using two or more methods to confirm the peptide matches.

Select 'OrbiOrbi.pkl' from the Peptide Data frame (left) and choose 'inChorus protein ID' from the Tools menu. The inChorus Database search dialogue appears:
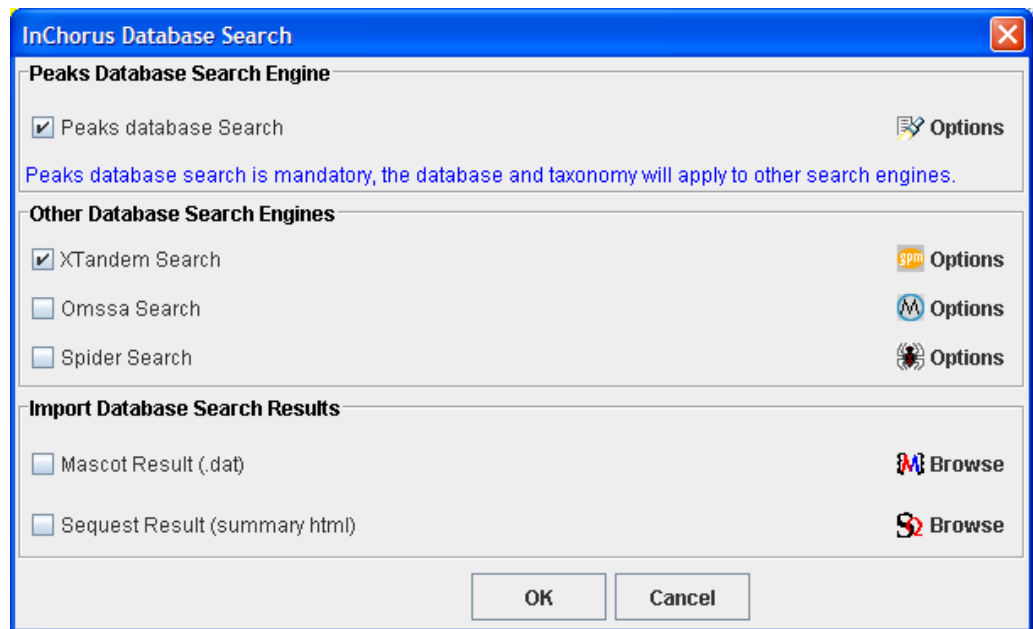
Make sure that "Peaks database Search" and "X!Tandem Search" are selected. Notice that there are three 'Options' icons on the right. They correspond to each search engine.

Click the Peaks database Search options button. The options pane is similar to the one we've seen already. The settings that we used before should still be there. If not, select "OrbiStandard" from the drop down list in the top right corner. Before pressing the "OK" button, we can make one change. Since we already have *de novo* sequencing results, we don't need to do *de novo* sequencing again. Click the option 'I have already run de novo, don't do it again', then press the OK button.

Click the X!Tandem search options button (top). This window allows us to set options for the X!Tandem search tool. This window is set up to behave almost exactly the same as the X!Tandem interface, so it may look familiar. Leave the fragment error at 0.1, and make sure there are no modifications turned on. Under "7. Predefined methods" choose FTICR. To learn more about X!Tandem settings, double-click any of the question marks. Press the OK button.

Now that we've set everything up for the inChorus search, press the OK button on the inChorus Database search dialogue. inChorus will call on each search engine, wait until they are finished, then compile their results together, ensuring the integrity of the data-results relationship.

Watch the task queue (bottom-left of PEAKS Studio). After everything is finished, new search results will appear in the Peptide Data frame (left), stamped with the date and time. The task queue will be empty, and the results will display. There's also a nice little report to tell us if there were any errors.

05-Jun-06 14:47:33_NR [inChorus protein ID results]

Peptide View | Protein View | Search parameters

Index	m/z	z	mass	Peptide	Mr(Calc)	Delta(Mass)	Score (%)	Start	End	Accession No.	Peaks	XTandem
1	452.2...	2	902.55...	TKIPAVFK	902.55...	-0.0035	68	76	83	gi|229460	✔	
2	458.7...	2	915.46...	LDAINENK	915.46...	-0.0035	71	84	91	gi|229460	✔	
2	458.7...	2	915.46...	IDALNENK	915.46...	-0.0035	71	84	91	gi|2194089	✔	
3	467.2...	2	932.53...	LIVTQTMK	932.53...	-0.0053	35	1	8	gi|229460	✔	
3	467.2...	2	932.53...	IIVTQTMK	932.53...	-0.0053	35	1	8	gi|72079	✔	
4	533.2...	2	1,064...	VLVLDTDYK	1,064...	-0.0048	85	92	100	gi|229460	✔	
5	545.9...	3	1,634...	TPEVDDEALEKFDK	1,634...	-0.0050	100	125	138	gi|229460	✔	✔
7	597.3...	2	1,192...	VLVLDTDYKK	1,192...	-0.0074	100	92	101	gi|229460	✔	✔
8	623.2...	2	1,244...	TPEVDDEALEK	1,244...	-0.0023	100	125	135	gi|229460	✔	✔
9	673.3...	1	672.37...	IGDLQK	672.38...	-0.0028	50	619	624	gi|28972720	✔	
9	673.3...	1	672.37...	LGDLQK	672.38...	-0.0028	50	136	141	gi|52782259	✔	
9	673.3...	1	672.37...	LGDIQK	672.38...	-0.0028	50	736	741	gi|63586345	✔	
9	673.3...	1	672.37...	GLDIQK	672.38...	-0.0028	19	9	14	gi|229460	✔	
10	674.4...	1	673.41...	IPAVFK	673.41...	-0.0002	76	78	83	gi|229460	✔	
10	674.4...	1	673.41...	LPAVFK	673.41...	-0.0002	76	24	29	gi|62659199	✔	
12	697.7...	3	2,090...	LDAINENKVLVLDTDYKK	2,090...	0.0022	100	84	101	gi|229460	✔	✔
12	697.7...	3	2,090...	IDALNENKVLVLDTDYKK	2,090...	0.0022	100	84	101	gi|2194089	✔	✔
13	771.7...	3	2,312...	VYVEELKPTPEGDLEILLQK	2,312...	-0.0044	100	41	60	gi|229460	✔	✔
14	818.3...	2	1,634...	TPEVDDEALEKFDK	1,634...	-0.0067	100	125	138	gi|229460	✔	✔
16	903.1...	3	2,706...	VAGTWYSLAMAASDISLLDAQSAPLR	2,706...	0.0007	99	15	40	gi|229460	✔	✔
17	903.5...	1	902.56...	TKIPAVFK	902.55...	0.0013	3	76	83	gi|229460	✔	
18	933.5...	1	932.53...	LIVTQTMK	932.53...	-0.0012	34	1	8	gi|229460	✔	
18	933.5...	1	932.53...	IIVTQTMK	932.53...	-0.0012	34	1	8	gi|72079	✔	
20	1065...	1	1,064...	VLVLDTDYK	1,064...	0.0035	5	92	100	gi|229460	✔	
21	1157...	2	2,312...	VYVEELKPTPEGDLEILLQK	2,312...	-0.0085	100	41	60	gi|229460	✔	✔
22	1354...	2	2,706...	VAGTWYSLAMAASDISLLDAQSAPLR	2,706...	-0.0037	90	15	40	gi|229460	✔	

PEAKS displays the same 'Peptide View' as before. Now, X!Tandem search results are considered as well. Where X!Tandem agrees with a PEAKS assignment, there's a checkmark in the Peaks column and a checkmark in the X!Tandem column. Also, the score on this peptide is increased – in this way we increase confidence in the assignment. Since the two tools take different approaches, we may discover that PEAKS finds some peptides that X!Tandem misses, and vise-verse. Where this is the case, only one checkmark will be displayed and the score is penalized slightly in some cases. Sometimes we can find a good hit that the other search engine would have missed. In this way we increase coverage.

Click the 'Protein View' tab to see a summary of PEAKS' and X!Tandem's results at the protein level.

Thus concludes our walkthrough of PEAKS Studio 4.0's basic features.

Graphical User Interface

*A reference section to help us find our way around.*

This chapter deals with interface elements. It is meant to be used as a reference so we can look up certain interface elements when we get stuck. For instructions on how to use PEAKS Studio to perform certain tasks, the section entitled "Using PEAKS Studio" will be more instructive.

The first part of this chapter describes **windows, dialogues, frames and reports**. This tells us what certain dialogue boxes, windows and frames do and how to read them.

The second part of this chapter deals with **toolbars**. Toolbars are a very useful way to quickly get at the functions we use most.

Windows, Dialogues, Frames and Reports

PEAKS Studio 4.0 main window
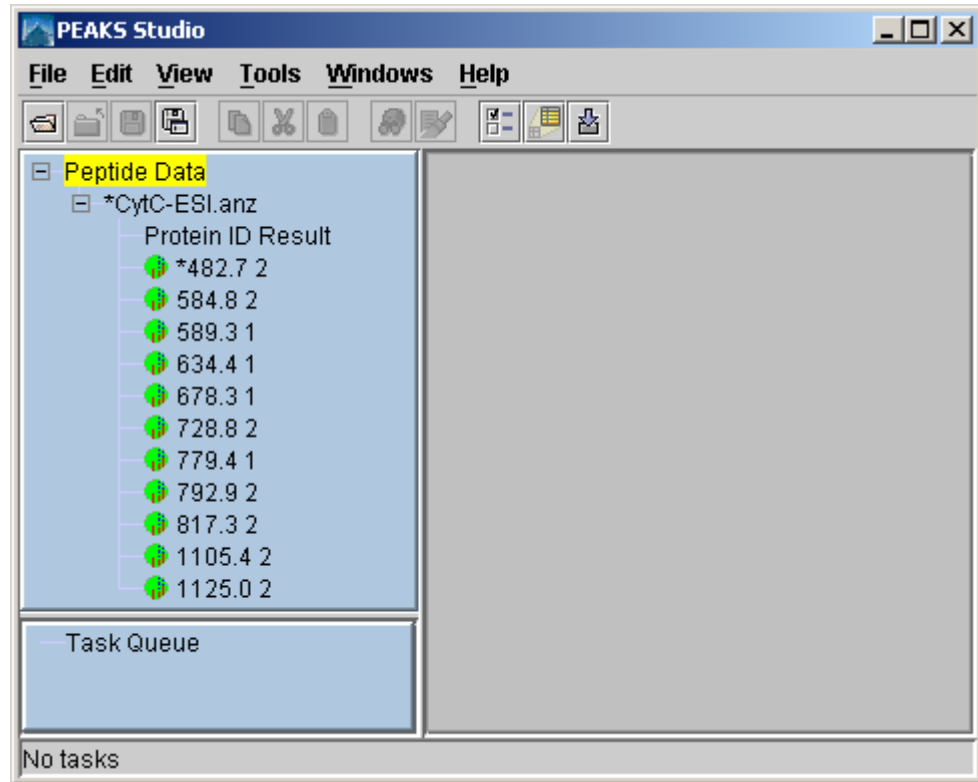
Comprises:

- **Peptide data frame (left):** This displays a listing of parent ions by m/z and charge. Clicking on one will bring up the its MS/MS spectrum. The colored dot by each spectrum shows dark green for unprocessed, or light green for sequenced (or partially sequenced). An asterisk (*) next to a spectrum shows that it contains unsaved information.

  Spectra are grouped by data files, or by nodes (which act like data files). Select a data file or node by clicking on its name (i.e. click on CytC-ESI.anz in the above example), or a spectrum within a data file by clicking on it. Use the ' + ' and ' – ' boxes to expand and collapse the view.

- **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.

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

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

Auto de Novo Parameters Dialogue

- **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.

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

- **Enzyme:** choose from a dropdown list of enzymes that we used to digest our protein sample. Click the "Edit Enzymes" button to edit the enzymes defined in this list, or to add to it.

- **Report top:** set how many de novo sequence candidates PEAKS will report. Choose from a dropdown list.

- **PTM selected for search:** this box displays the modifications currently selected for analysis, these will be considered during auto de novo sequencing. To change this, click the Add/Remove PTM button.

- **Preprocess before auto de novo:** PEAKS Studio has its own built-in preprocessor for removing noise, centroiding, and peak charge recognition from MS/MS data. Check this box to turn preprocessing on.

Protein Identification Parameters Dialogue

- **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.

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

- **Enzyme:** choose from a dropdown list of enzymes that we used to digest our protein sample. Click the "Edit Enzymes" button to edit the enzymes defined in this list, or to add to it.

- **Report top:** set how many de novo sequence candidates PEAKS will report. Choose from a dropdown list.

- **PTM selected for search:** this box displays the modifications currently selected for analysis, these will be considered during database searching. To change this, click the Add/Remove PTM button.

- **Preprocess before auto de novo:** PEAKS Studio has its own built-in preprocessor for removing noise, centroiding, and peak charge recognition from MS/MS data. Check this box to turn preprocessing on.

- **Advanced options (de novo sequencing):** The PEAKS approach to protein identification uses de novo sequences to help out in the search. This section allows you to decide how to obtain the de novo sequences required for the search.

PEAKS Properties Dialogue
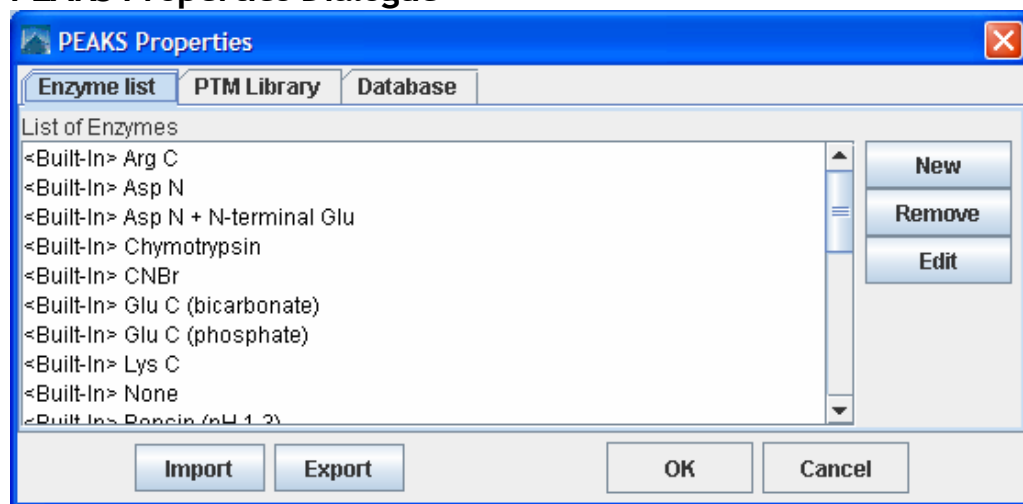
- **Enzyme list tab:** Displays a list of built-in and user-defined enzymes. We may edit and create Enzymes from here.

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

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

Enzyme Editor Dialogue

- **Digestion Rules:** This is how we specify where our enzyme will cleave the protein between two amino acids to create peptides. Use set brackets {} around a residue to denote "any amino acid except the ones enclosed in these brackets". Use X to denote "any residue". Listing several amino acids in one box means "any one of these residues".

- **Specificity Parameters:** Peptides can break down such that only one end is a cleavage site. Check the boxes to tell PEAKS to search for only for peptides that have proper cleavage sites on both ends, or to require that only one end be a proper cleavage site.

- **Shorthand notation:** Advanced users may specify their enzyme cleaveage in shorthand notation, but it is not required.

- **Saving/Loading Enzymes:** After setting up an enzyme, we can save it for future use. Click the "Save Parameters" button, and choose a name for future reference if prompted. Don't worry, we can't accidently overwrite the defaults. Any enzyme we save will be available in the drop-down list at the top of the window. To see what's inside, just select one, and the enzymes digest rules boxes will be populated.

PTM Selector Dialogue
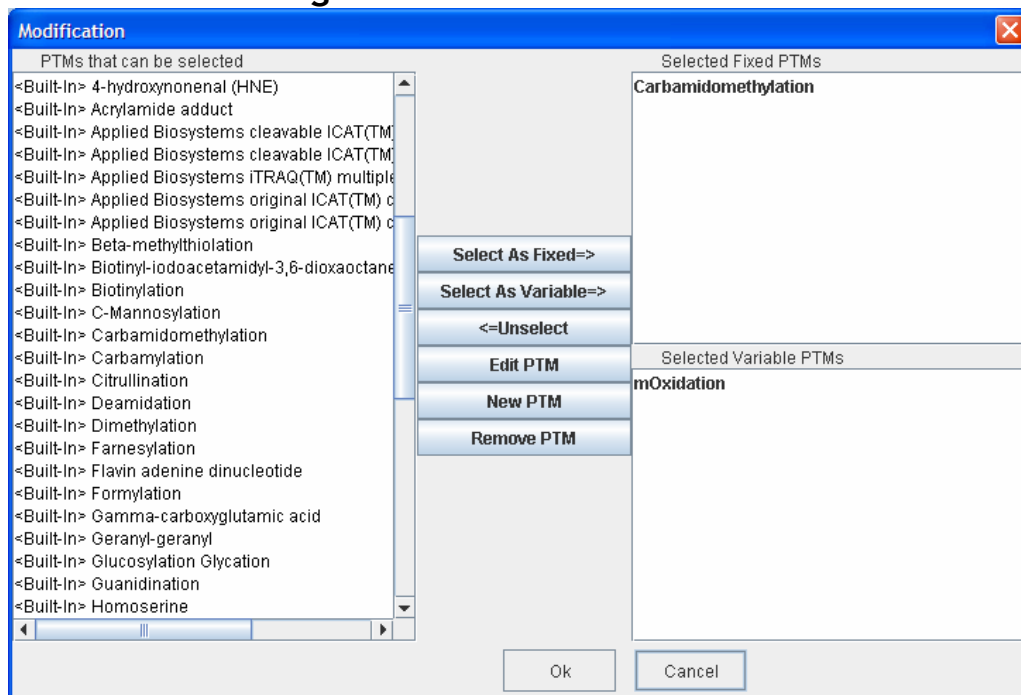
Here we can from a list of available post-translational modifications. We can choose any PTM as Fixed PTM or Varied PTM (to tell PEAKS that it may or may not occur). To make this selection, click on a PTM in the list at left, and then click the "Select As Fixed=>" or the "Select as Varied=>" button. 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. These are listed as <Built-In>. If we want to create a new PTM, we can click 'New PTM' to create a new one. The **Editing a PTM** and **Creating a New PTM** sections below describes how this is done.

PTM Editing Dialogue

- **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 if they are at the N-terminus and residues that can only be modified at the C-terminus, and residues that can only be modified if they are not on either terminus.

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

Ion Table Settings Dialogue

There are two such dialogues, one each for the 'Advanced' Ion Table, and the 'Basic' Ion table. The two dialogues are identical but for the table they effect.

See the below section entitled Main Processing Window for a description of the Ion Table itself.
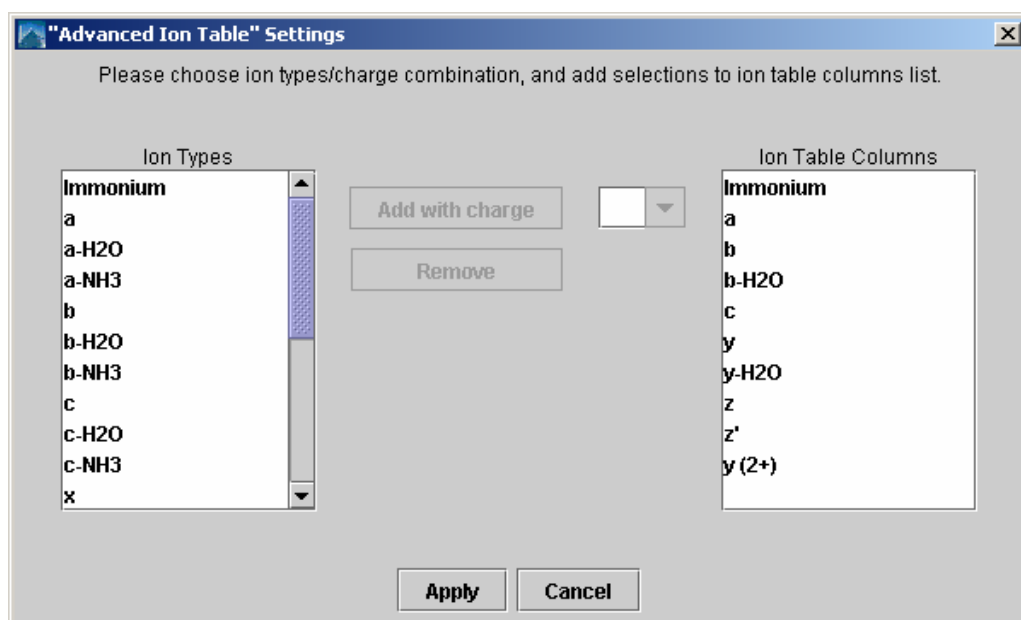
The 'Basic' and 'Advanced' Ion tables differ only in the number of ions they can display. The 'Basic' table displays up to six ions.

- **Ion Types list:** A listing of all the types of ions PEAKS Studio considers in its analysis.

- **Ion Table Columns list:** A list of the columns that will appear in the ion table, each representing a type of ion. The columns will contain the masses at which the particular type of ion was found, if at all.

- **Charge list box (unlabeled):** Each type of ion can be added up to 4 times, depending on the charge we specify.

- **Add with charge button:** After selecting an ion type from the Ion Types list and a charge from the list box, clicking this button will add that ion to the Ion Table Columns list.

Protein Identification Result Window

The protein identification result window contains the results from one protein identification run on one data set. It is organized into three tabs: **peptide view, protein view** and **search parameters**.

Peptide View

The peptide view summarizes the results for each MS/MS spectrum. All peptides that match to each spectrum are displayed. By default the peptides are grouped by spectrum, but the list is sort-able by any of the columns in the table. Selecting one or more (click and drag, or use shift+click) items in this list selects those spectra in the peptide data tree (left).

	01-Nov-05 12:33:21_NCBI nr tax									

Peptide View   Protein View   Search parameters

Index	m/z	z	mass	Peptide	Mr(Calc)	Delta(Mass)	Score (%)	Start	End	Accession No.	Peaks
1	482.7	2	963.38...	EDLLAYLK	963.52...	0.1421	99.0	130	137	gi|76690040	✔
2	584.8	2	1167.5...	TGPNLHGLFGR	1167.6...	0.0295	99.0	66	76	gi|76690040	✔
3	589.3	1	588.29...	VDVEK	588.31...	0.0192	89.47	59	63	gi|111734	✔
3	589.3	1	588.29...	ELEAK	588.31...	0.0192	0.01	315	319	gi|73975307	✔
4	634.4	1	633.39...	LFVQK	633.385	0.0078	97.22	47	51	gi|76690040	✔
5	678.3	1	677.29...	YLPGTK	677.37...	0.0821	99.0	112	117	gi|76690040	✔
5	678.3	1	677.29...	LFTGLK	677.41...	0.1185	0.0	482	487	gi|62646155	✔
6	728.8	2	1455.5...	TGQAPGFSYTDANK	1455.6...	0.0776	99.0	78	91	gi|76690040	✔
7	779.4	1	778.39...	MLFAGLK	778.44...	0.0483	98.75	118	124	gi|76690040	✔
7	779.4	1	778.39...	KLFLMK	778.47...	0.0847	0.01	668	673	gi|55628230	✔
8	792.9	2	1583.7...	KTGQAPGFSYTDANK	1583.7...	0.0275	97.96	77	91	gi|76690040	✔
8	792.9	2	1583.7...	MMALSDNLPAGVTHK	1583.78	0.0055	0.0	377	391	gi|55628230	✔
9	817.3	2	1632.5...	LLMKCSQCHTVEK	1632.7...	0.1934	29.21	11	23	gi|31540569	✔
10	1105.4	2	2208.7...	EHWQREEQQRLEEEQR	2209.0...	0.2407	7.44	404	419	gi|73975307	✔
10	1105.4	2	2208.7...	QKEFLRDTPQRHCEHTK	2209.0...	0.2957	2.06	36	52	gi|73964720	✔
11	1125.0	2	2247.9...	NSAAKHKLSVKPKNQRVSKK	2247.3...	0.6567	0.64	239	258	gi|73975307	✔

Protein View

The protein view is most useful as a summary of what proteins were present in a sample, and the peptides matched to them. It has two sections:

- **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. Show the whole group by clicking the + sign.* In the example below, lactoglobulin beta is the top ranked protein candidate.

- **Peptide Match Reports (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. At the bottom of this list, the complete protein sequence is shown with matching peptides highlighted in red.

Protein View
showing two
proteins in the
index, and
beginning the full
report

	05-Jun-06 14:47:33_NR [inChorus protein ID results]							
Peptide View	**Protein View**	**Search parameters**						
Accession			Mass	Score(%)	Coverage	Query matched	Marked	Description
PEAKS DB Search (NR 0.01 0.01 TEMP)							☐	
⊞ gi|229460			18,367	100	61.73%	18	☐	lactoglobulin beta
⊟ gi|28972720			78,207	29	0.84%	1	☐	mKIAA1321 protein [Mus mus...
gi|62648518			57,150	29	1.13%	1	☐	PREDICTED: similar to mKIA...
gi|52782259			85,124	29	0.77%	1	☐	phosphofructokinase, muscle...
gi|63586345			261,103	29	0.25%	1	☐	PREDICTED: similar to mKIA...

NCBI BLAST search of **gi|229460**

Links to retrieve entries containing this sequence from NCBI Entrez:

Accession/ID	Description
gi|229460	lactoglobulin beta

Peptides List

Mz	Charge	Mr(calc)	Start	End	Score(%)	Search Engines	Peptide
452.28497	2	902.5589	76	83	68	p-----	TKIPAVFK
458.73856	2	915.4661	84	91	71	p-----	LDAINENK
467.2729	2	932.5366	1	8	35	p-----	LIVTQTMK
533.2925	2	1,064.5752	92	100	85	p-----	VLVLDTDYK
545.92804	3	1,634.7673	125	138	100	px----	TPEVDDEALEKFDK
597.3386	2	1,192.6702	92	101	100	px----	VLVLDTDYKK
623.2947	2	1,244.5771	125	135	100	px----	TPEVDDEALEK
673.38513	1	672.3807	9	14	19	p-----	GLDIQK
674.4233	1	673.4163	78	83	76	p-----	IPAVFK

Search parameters

This tab displays the protein identification parameters that were used to guide the search that generated these results.

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
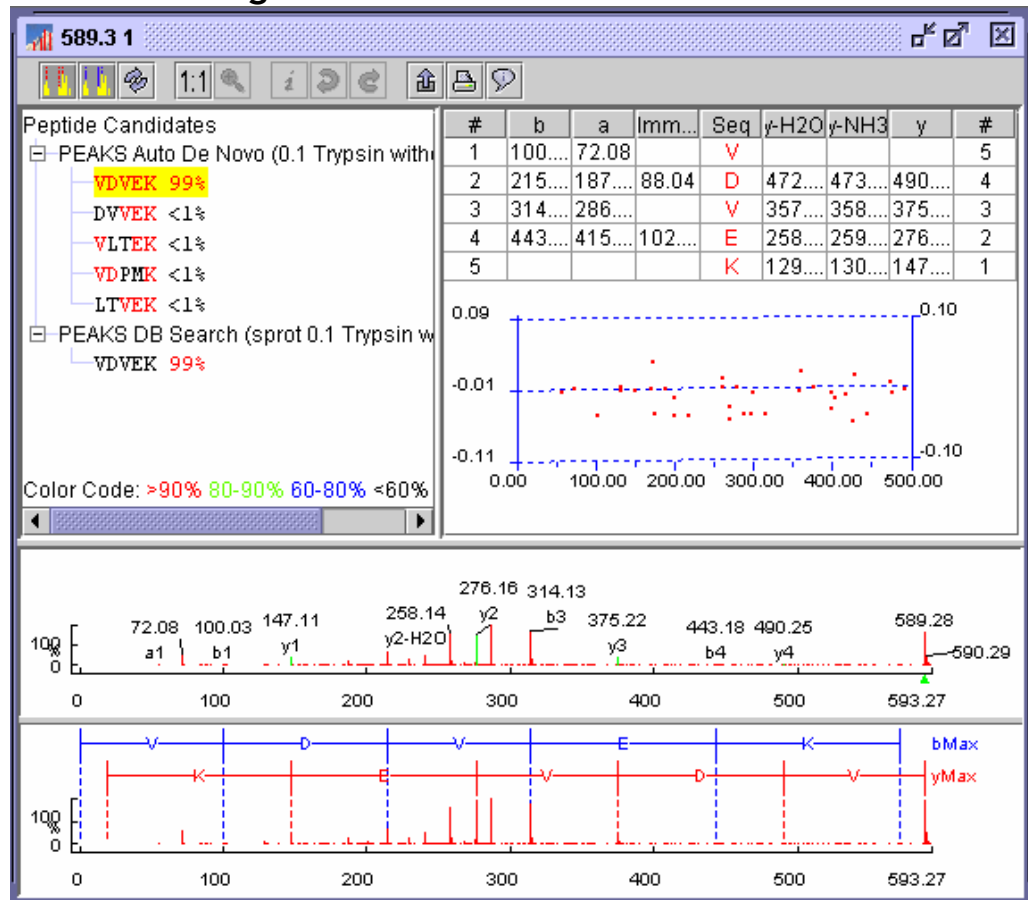
- **Main Processing Window Toolbar:** quick access to many processing functions. See "Toolbars" section below.

- **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 ions with their corresponding masses -- i.e. the mass of the b1 ion is shown in the top right corner. The default Ion Table will display b, a, immonium, yH2O, yNH3 and y ions in basic mode; it will display b, b-H2O, a, c, immonium, y, y-H2O, z, z' and y(2+) ions in advanced mode. To switch from basic mode to advanced mode, choose 'Show ion table' from the 'View' menu.

  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. We 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 the positions of major ions that delimit the proposed sequence. By default, the Spectrum Alignment Frame displays b-ion and y-ion peaks and the derived peptide sequence between them. The Spectrum Alignment Frame can also show the position of c-ion and z-ion peaks.

Ion Editor

Ion Editor

Please choose ion type:
- ⦿ C Term Ions
- ○ N Term Ions

x
y
y-H2O
y-NH3

Add
Remove

Selected peak information:
mz: 136.0692
intensity: 4.0323
current ion:

Comments:

Apply   **Cancel**

- **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

Image Export

Width: 640   Height: 480   Format: gif ▼

Filename: C:\Program Files\PEAKS Studio\Data\multiple spectra\927.4.gif

☐ Export selected area
- ⦿ Peptide candidate tree
- ○ Ion table and error map
- ○ Current spectrum window

Ok   **Cancel**

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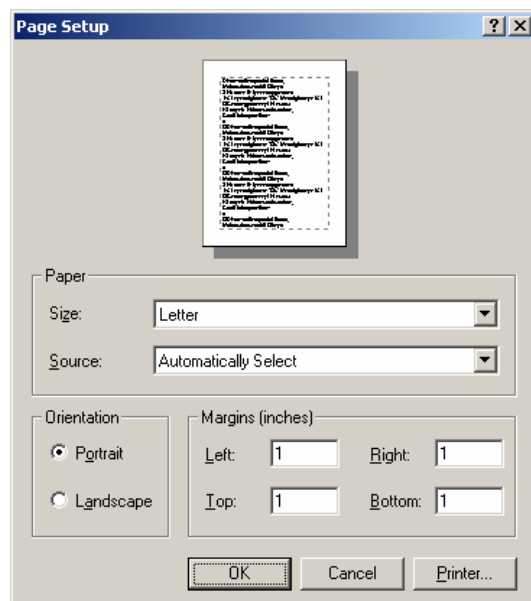
- **Orientation:** paper orientation is shown in the picture at the top. Change this by clicking the "Portrait" or "Landscape" radio buttons.

- **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 ANZ format) that also contains peptide analysis data. The file should be in PKL, DTA, MGF or ANZ 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 ANZ 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 ANZ format.

"Copy" button: Copy selected spectrum data.

"Cut" button: Cut selected spectrum data.

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

"Manual merge spectra" button: After selecting more than one spectrum in the peptide data tree, this button becomes enabled. Right click it to merge these spectra into a s ingle MS/MS spectrum, and remove the old ones.

"Data Refinement" button: Merge scans of the same peptide, remove noise spectra, preprocess within each MS/MS spectrum and recover peptide charge state. The data refinement options dialogue will allow us to choose and to set parameters for each of these refinement tools.

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

"Protein Identification" button: perform protein identification a selected data file. Press this after selecting one or more 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, and manual *de novo* parameters.

"PEAKS Properties Configuration" button: define PTM, Enzymes, 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.

PEAKS Studio Configuration

How to set up PEAKS Studio just the way we like it.

T his chapter deals with configuration. PEAKS Studio 4.0 is a versatile and flexible tool. But in order to use the software to its full extent, we must learn how to configure it to make it do what we want it to. Additionally, PEAKS Studio 4.0 allows us to set up many defaults and presets to help us be quick and precise.

We can use PEAKS Studio 4.0 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. It is recommended that we configure PEAKS Studio 4.0 before processing data files.

PEAKS Properties Configuration

One of PEAKS Studio 4.0 preferences, PEAKS Properties configuration sets the parameters that the algorithm will use in processing our data files. PEAKS properties include: Enzynme, PTM and database. PEAKS Studio 4.0 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 list**", "**PTM Library**" 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.

Creating and defining PTM

If we know that our sample protein may have been modified since translation, we need to apply this information to our analysis. To edit the list of PTM available to PEAKS Studio we use the open the PEAKS Properties dialogue and select the PTM Library tab.
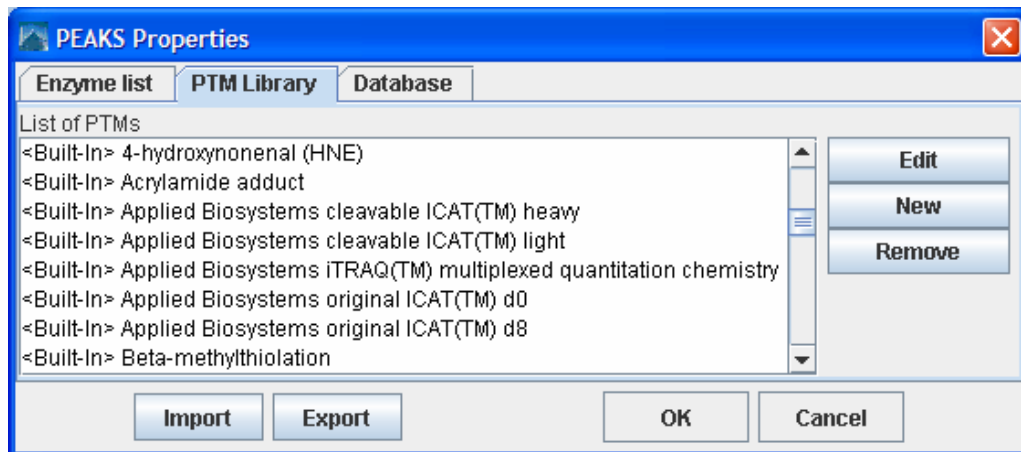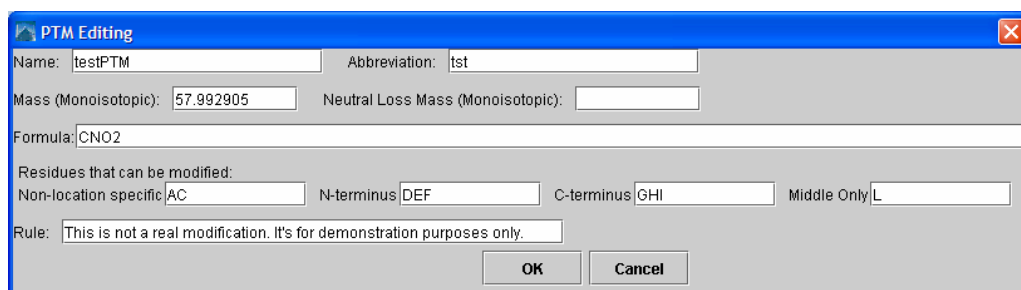
All PTM are listed here including<built-in> PTM and user defined PTM. From here, we can create a new PTM, edit an existing PTM or remove a PTM from the list. See the sections below for help with these operations. *Built-in PTM cannot be removed from the list, but can be edited.*

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.

Creating a New PTM

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 setting up PEAKS auto de novo or PEAKS Protein ID, click the "new PTM" button while selecting PTMs.* The PTM Editing dialogue will appear:
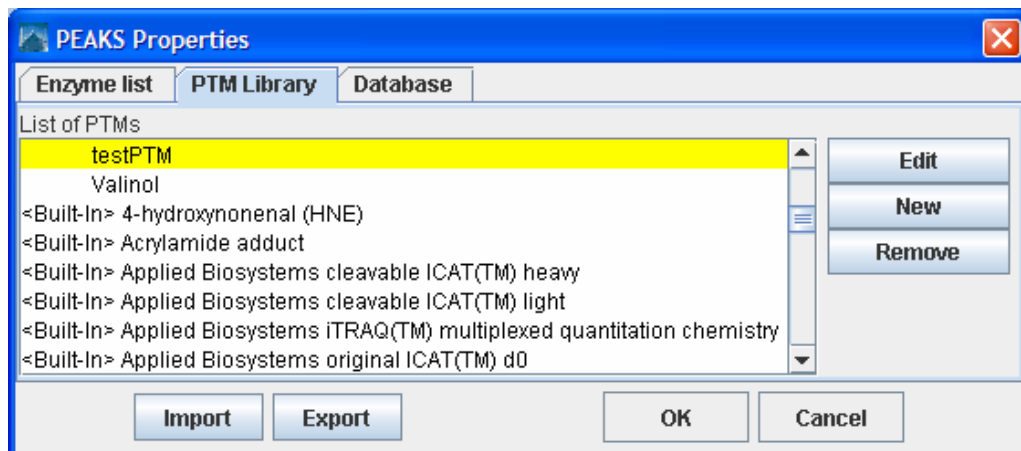
Figure 1 Create new PTM

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.


Editing a PTM

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 setting up PEAKS auto de novo or PEAKS Protein ID, click the "add/remove PTM" button, to bring up the Modification window for that search, then click the "new PTM" button.* The PTM Editing dialogue will appear. Now we follow the same procedure (see above) as we would if creating a new PTM.

Removing a 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. Built in PTM cannot be removed.

Database Manager

PEAKS Studio 4.0 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 4.0, 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.
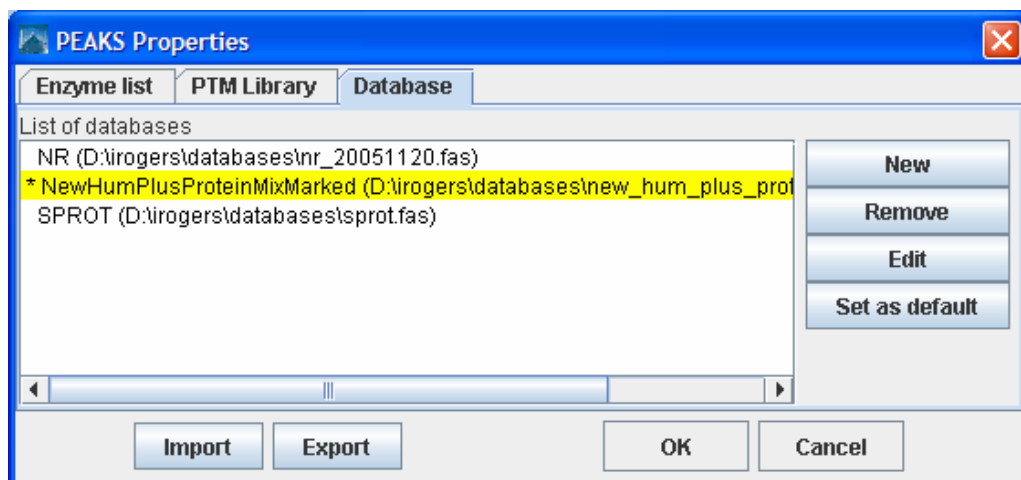
Load/Configure a new database
For an in depth look at configuring a database, see the Database Configuration section in Chapter 2.

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. If taxonomy is available for this database, download those files too. Return to PEAKS Studio 4.0 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). If taxonomy is available for the database, find those files too. 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 Studio 4.0 associates with a database, taxonomy files, and the header parsing rules for that database. 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.

Moving/Updating a Database

If we choose to move a database to another directory, or delete it entirely, we should tell PEAKS. 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.
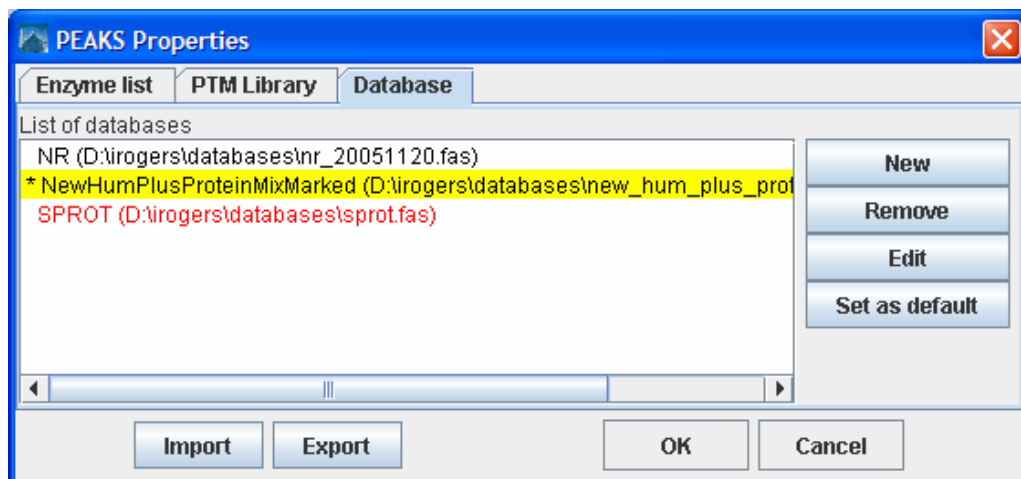

If we choose to update the database [perhaps by downloading the latest database file and overwriting the old database file], PEAKS will show the database information in light gray. *A light grey colour could also mean that the database does not have header information.*
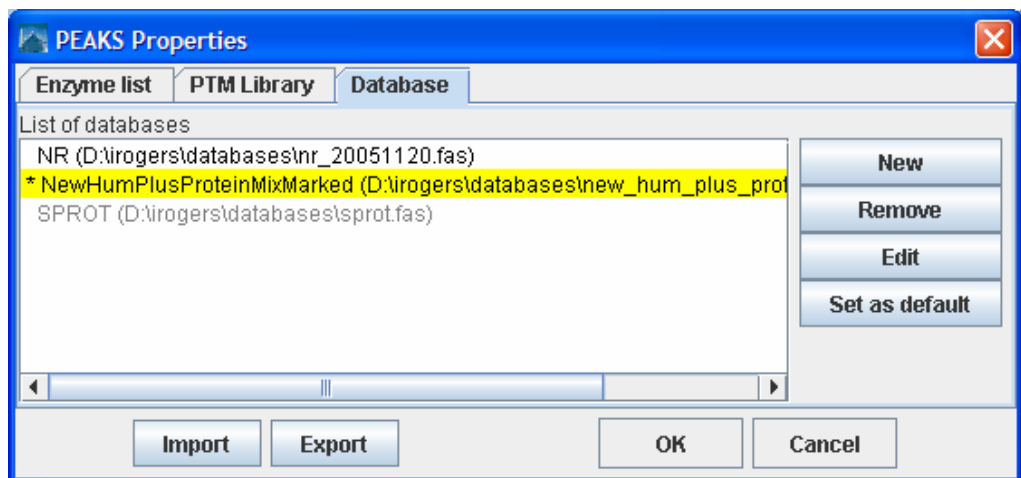
Importing and Exporting PEAKS Properties

We may wish to use PEAKS Studio 4.0 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 wish to use our PEAKS properties on a colleague's system, we must remember to export our colleague's properties to a separate file so that it will not be lost and can be imported later.
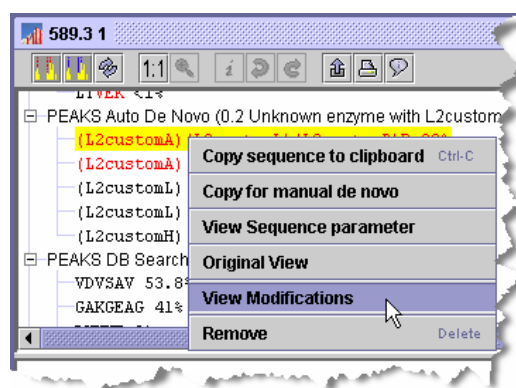

A note on sharing sequences with PTM

Sequence data and protein identification results for a given spectrum are stored in an .ANZ 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 .ANZ 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 an 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 import a user defined PTM from another user's .ANZ file, we open the .ANZ 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.
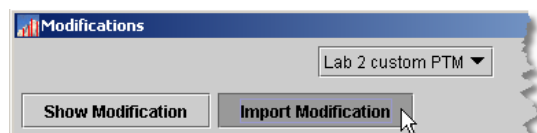
Configuring the ion table

The ion table, displayed in the top right of the main processing window, displays all the ions that were found as evidence for the selected sequence. There are two presets: the Basic Table and the Advanced Table. Select which one to display by choosing "Show Ion Table ►" from the "View" menu.
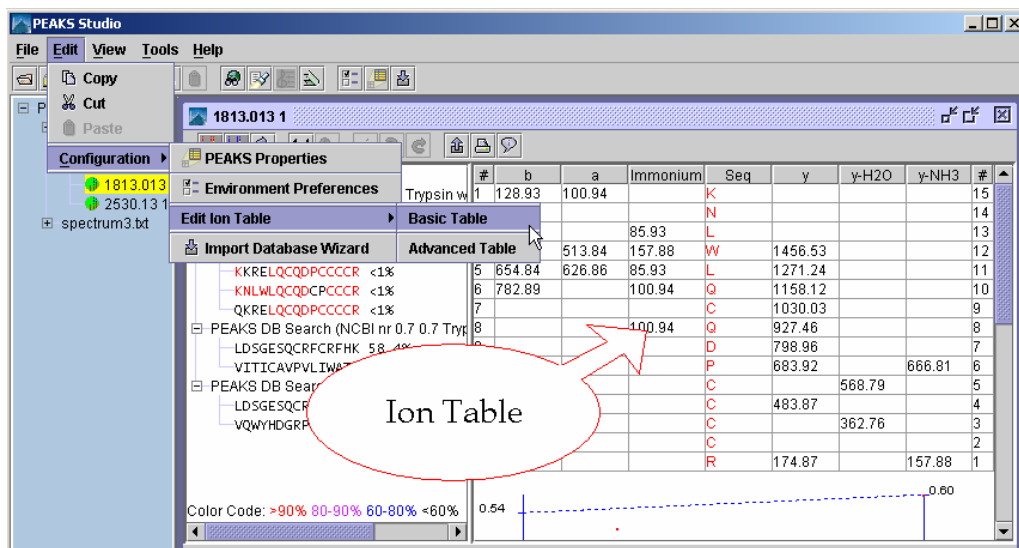
The Basic Table will display a maximum of 6 ions. The Advanced Table can be configured to display as many as are available. To configure the ion table, we choose "Edit menu" → "configuration" → "edit ion table" and then select which table we'd like to edit.

The Ion Table Settings dialogue box will appear:
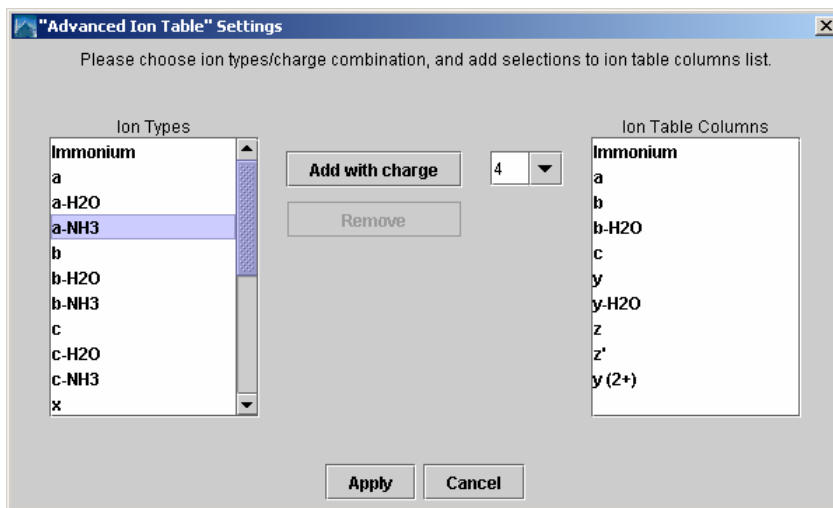
The ions types that will be displayed in the ion table are shown on the right. The complete list of ion types available is shown on the left. To add an ion type to the ion table (i.e. add a column to the ion table):

1. Select one or more ions from the list on the left. Use Shift+click and Ctrl+click to select multiple list items.

2. Select a charge (from 1-4) from the drop down list in the middle.

3. Click the 'Add with charge' button

For example, configure the ion table to display y^{2+} ions by selecting 'y' from the list on the left, and '2' from the dropdown charge list.

Remove ion types from the ion table (i.e. remove columns from the table) by selecting one or more items in the list on the right, and clicking the 'Remove' button.

PEAKS Environment Preference Configuration

One of PEAKS Studio 4.0 preferences, PEAKS Environmental Preference allows us to customize PEAKS Studio 4.0 to our needs. PEAKS Environmental Preferences include: Environment, Color, and Manual *de novo*, and Parameters. **To edit PEAKS Studio 4.0 Environment Preferences:**

- Click the ▣ icon in the toolbar;

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

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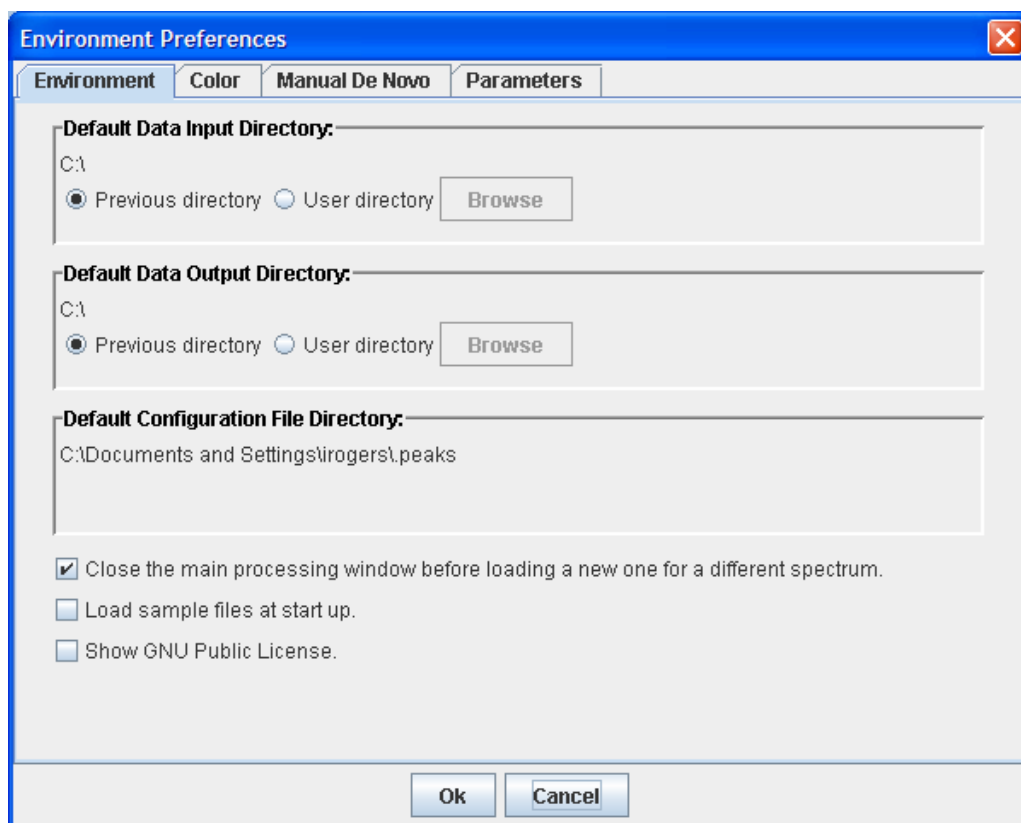
Environment

To change the working environment, we open the Environment Preferences dialogue and ensure that the "Environment" tab is selected.

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. The current working folders for data input and data output are shown. We can choose to have PEAKS Studio 4.0 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.

We can choose to load a new spectrum view window for each spectrum, or open one at a time. We can choose to show the sample data at startup. We can display the GNU license whenever GNU governed software/libraries are called. Click the appropriate 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 Preferences dialogue and ensure that the 'Color' tab is selected.
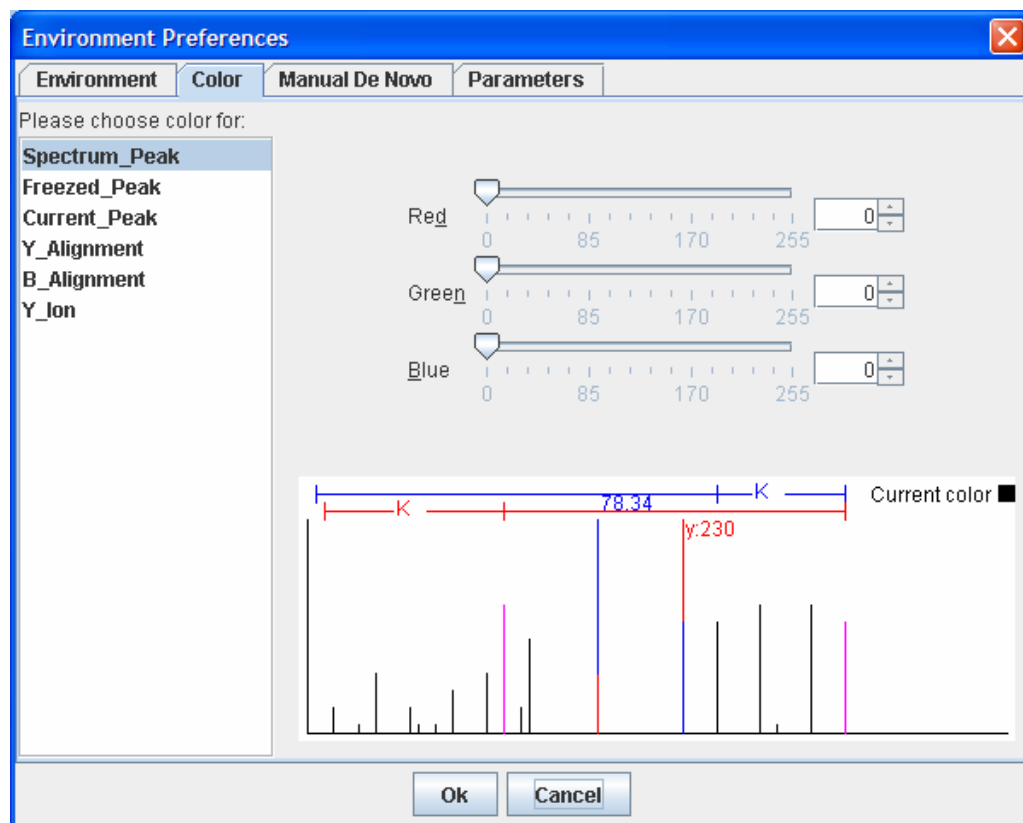

Choose the object whose color we'd like to change from the list at the left. Then we click on the slider bar, type in a number (0 to 255) in the textbox, or scroll up and down on the arrows next to the textbox to select how much of the corresponding color we'd like to apply. Choose an amount for all 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.

Manual *de novo*

We may wish to sequence a peptide manually, using spectrum data. PEAKS Studio 4.0 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.

To adjust Manual *de novo* options, open the Environment Preferences dialogue and ensure that the 'Manual *de novo*' tab is selected.
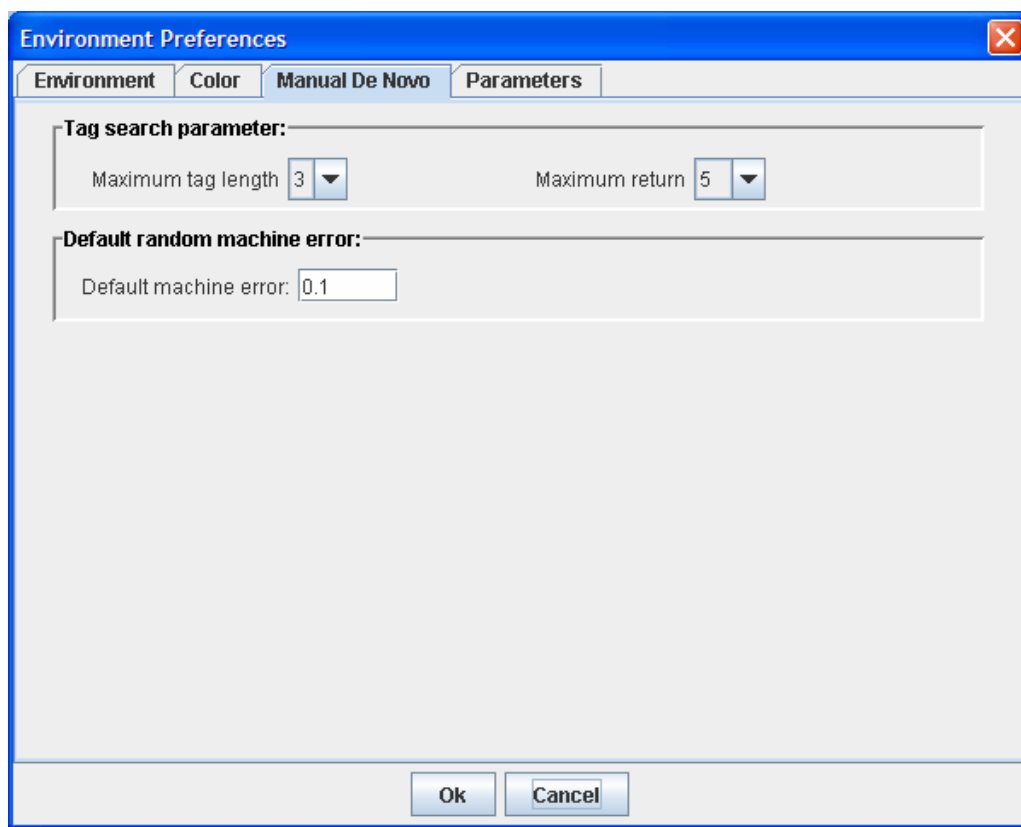
When sequencing a peptide using the manual *de novo* tools, we can get PEAKS to help us by searching to the left or right of a selected peak and returning a set of possible sequence tags (see Manual *De novo* section in the next chapter). We can choose how many search results we'd like to see, and we 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 113.19 between two y-ions that we have labeled. We are fairly confident that this should be tagged L (Leucine, with actual mass of 113.08), but PEAKS is not labeling it for us. This may be because 113.19 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.

Removing Saved Parameters

After months of use, our list of parameters saved for use with PEAKS Protein ID or PEAKS auto de novo may become cluttered with infrequently used parameter sets. It makes sense to clear them out from time to time. To do so, open the Environment Preferences dialogue and ensure that the 'Parameters' tab is selected. A list is shown for each tool that has savable parameter sets. Select one or more (using shift+click and ctrl+click) and click the adjacent "Remove" button to remove it from the list.
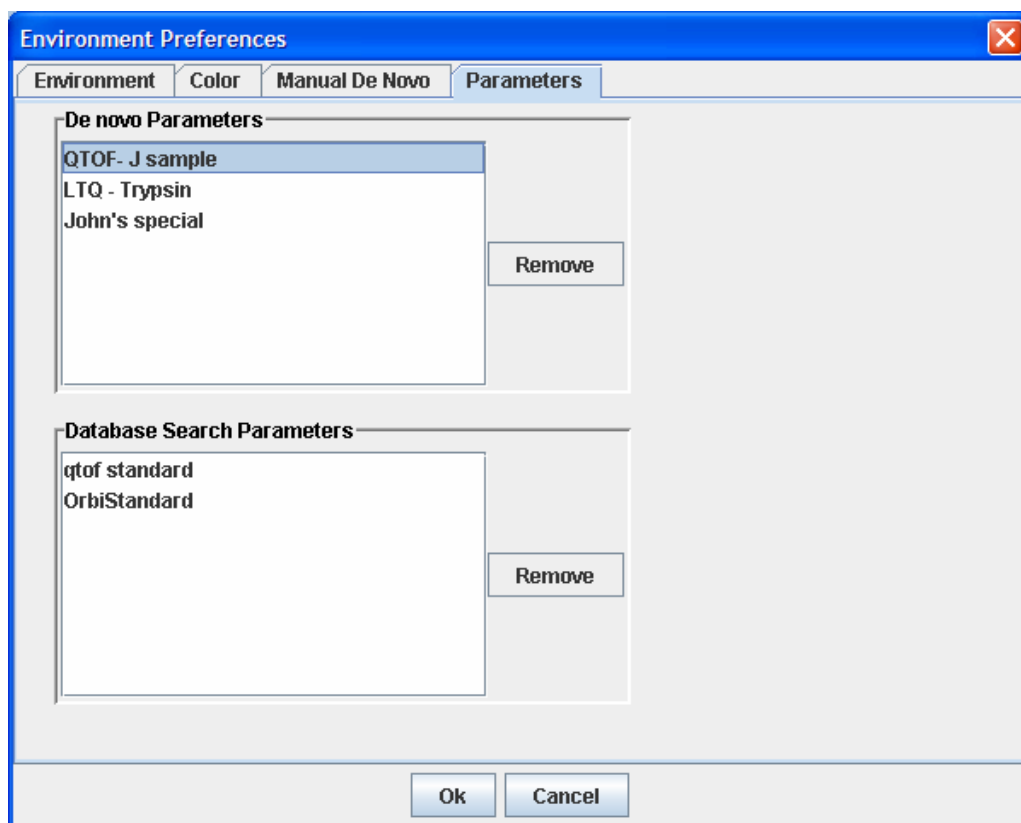
PEAKS Studio Usage

A task based guide to processing our data with PEAKS Studio 4.0.

T his chapter deals with usage. It is broken up into tasks that a typical user might perform. It assumes we can identify parts of the Graphical User Interface and that we are familiar with how PEAKS Studio 4.0 can be configured. The preceding two chapters provide in depth help on these subjects, and should be used as a reference. Such detail has occasionally been omitted from this chapter in the interest of succinctness.

The four cores of PEAKS technology are the data refinement, manual de novo sequencing, automatic de novo sequencing and protein identification tools. Surrounding Help is provided for these tasks, but also, how they may be used to perform less obvious operations.

Loading data into PEAKS Studio 4.0

PEAKS Studio 4.0 can be used to process data from any MS/MS instrument, provided the data is accessible, or can be converted to an accessible format. PEAKS handles data files in the following formats:

- PKL

- DTA

- MGF

- ANZ – the zip compressed XML based file format associated with PEAKS

- XML format files using the mzXML schema

- RAW files from Thermo Electron instruments

- RAW files from Waters QTOF instruments

- XML format files from Waters ProteinLynx software

- DAT files created by BSI's ABI converter software

Opening data files

PEAKS demo data can be found in the DATA sub-directory, located in the PEAKS directory.

In order to do any data processing we must first load our spectrum data into PEAKS Studio 4.0. To open a data file, click the 🖾 icon on the toolbar in the upper left corner of the PEAKS window or select "Open" from the "File" menu.
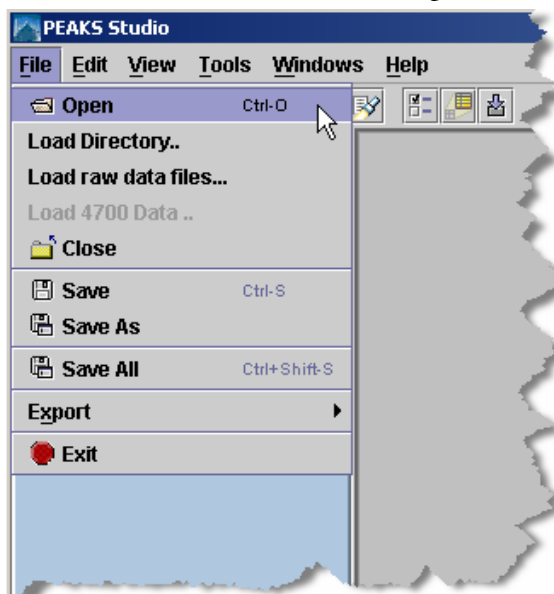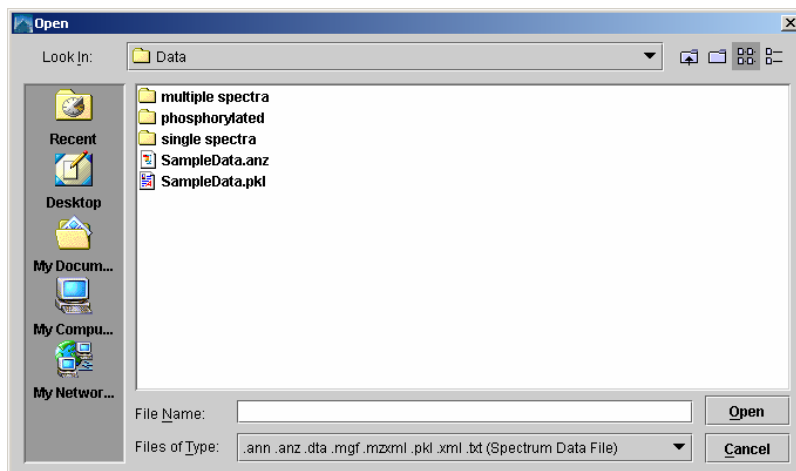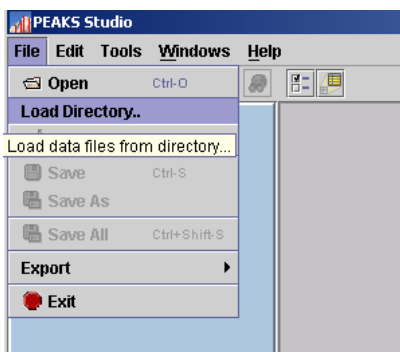
Select a file (in .ANZ, .pkl, .mgf, .dta, or .txt format). Click the Open button.

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 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 DTA 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 run) at once, and consider it as one MS run. PEAKS Studio 4.0 provides a tool for doing so.

Under the "File" menu, click "Load Directory..". Now browse to the directory we wish to load.

Do not select a file within the directory; rather select the directory itself. Press the Open button.

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.

The data file we just opened appears in the Peptide Data Frame on the left. It is represented by the folder name. Each spectrum contained in the data

file is represented by its precursor ion information (m/z value followed by the charge of the precursor ion that generated the spectrum) and file name.

Loading Thermo RAW data

PEAKS Studio 4.0 can load RAW data from our Thermo Electron mass spectrometer; provided that the masslynx software is installed on the same computer as is PEAKS Studio 4.0. To load Thermo RAW data, simply choose File → Open, and browse to the file.

Importing Masslynx RAW data

PEAKS Studio 4.0 can import RAW data from our Waters/MicroMass QTOF instrument. To do so, we choose "Import RAW data" from the File menu. As above, the file browser appears. Choose the .RAW data, and click the Open button. Again, we have the option to merge spectra, or not. For this to work, PEAKS Studio 4.0 must have access to the following libraries, which are part of MassLynx:

Some versions of MassLynx may differ.

- DACServer.dll

- Genutil.dll

- MetaGD32.dll

- raw.dll

- securityAccess.dll

- securitySettings.dll

- securitySignature.dll

They should be stored in the folder C:\MassLynx\ as part of the MassLynx software. If they are not stored here, or MassLynx is installed on another computer, the automatic loading will not work. If the automatic loading is not working for either reason try this:

1. We should be able to find the listed files on our computer, or another computer in our lab. If you can copy them, do so.

2. We can then create a folder called "C:\MassLynx\" on our computer, and place the files we copied here. But we're not finished, we must also register these files with Windows.

3. We can find a file on our system called regsvr32.exe using the 'Find' or 'Search' tool in our windows start menu. It is probably in "C:\WINDOWS\System32\". If it's not there, substitute the correct location in step 4.

4. Open a command prompt, or the Run tool from the start menu and type the following:

```
C:\WINDOWS\System32\regsvr32 C:\MassLynx\DACServer.dll
```

All on one line, with one space in the middle as shown. Press the enter key. If successful, windows will pop up a success message.

Please check the license

The libraries mentioned in this section are part of the MassLynx software, distributed by Waters corp. Please check the MassLynx license agreement, or contact a Waters representative to make sure it is okay to copy and use the libraries in this way.

Importing Data from the ABI 4700 or ABI 4800

BSI has created a converter to extract the data from an ABI-Oracle database. If we require this separate, free tool, we must ask a BSI representative. Once installed, we can start up the ABI 4700 Data Extractor from the Start menu.

System Requirements

This extractor can be installed on the same machine as ABI 4700 Explorer and the Oracle database (we will call this machine the 4700 SERVER in the following instructions) or another machine that has direct network access (no firewall, no proxy required) to the 4700 SERVER. Windows 2000 or Windows XP is recommended for use with this tool.

Configuration:

Before using the ABI 4700 Data Extractor, we should configure it. To do so, we can choose "Settings" from the "File" menu. Configuration needs the following:

- **4700 SERVER Name or IP Address:** input "localhost" if the Extractor is running on the 4700 SERVER (this is the default value), otherwise enter the IP address of the 4700 SERVER.

- **The socket used by the 4700 SERVER:** this is the port that the Oracle database listens to (the default is 1521).

- **Username to access the Oracle database:** most likely we do not need to change this (the default is "tsquared").

- **Password to access the Oracle database:** mostly likely we do not need to change this one either.

Data extraction procedure:

1. Load Spot Set List from the database (Do it via menu File | Load Spot Set List) The extractor will export the peak list of a spot set into a PKL file.

2. Open a Spot Set (menu File | Open Spot Set) Spot Set Chooser will help the user to choose a spot set. After selecting a spot set, click 'OK' to open it. The job run information of a spot set will be shown.

3. Select a job run There is a radio button before each Job Run, only the MS/MS job run can be selected for export, because we need the precursor information. Select a Job Run and click 'Convert' to do the extraction.

4. Choose a filename to save After clicking the 'Convert' button, the user needs to input a file name. And the peak lists of the selected job run will be exported.

Refining data before analysis

Since mass spectrometry data often contains noise and redundant data, it makes sense to purify the data before analysis. This will increase the quality of the results, while saving time spent on database searching and/or *de novo* sequencing. MS/MS spectra that are purely noise can be removed from the data; peptide charge information can be verified and recovered; multiple low quality scans of the same peptide can be merged into one scan with intense signal peaks; and the MS/MS scans themselves can be centroided, filtered for noise, and deconvoluted.

To begin refinement of data from a whole MS/MS run, we:

1. In the *Peptide Data* Frame, select the data file(s) containing the data that we wish to refine.

2. Click the Data Refine toolbar icon

   Or

   Select "Data Refine" from the "Tools" menu.

   Or

   Right click on the selected data file and select "Data Refine" from the popup menu

The Data refinement options dialogue appears:

3. Choose the data refinement tools we wish to use by clicking the "yes" radio button next to each one.

- **Instrument Type:** choose the type of spectrometer that produced the data to be analyzed.

- **Merge Scans of the same peptide:** in DDA mode, a mass spectrometer will often produce several tandem ms (i.e. ms/ms) scans of the same peptide. To increase the intensity of real signal peaks within these scans, and to reduce the size of the whole data set, it makes sense to merge ms/ms scans of the same peptide together. To avoid improper merging (of ms/ms scans of different peptides) we make sure that the measured parent ion masses of these peptides are very close, and that they have similar retention times in the LC column.

- **Correct precursor charges:** Since a mass spectrometer measures mass-to-charge ratios, we must know the charge on a peptide before we can determine its mass. The standard method of finding the charge is to look at the spacing of the isotope ladder in the survey scan. However many Ion-Trap instruments don't have enough resolution for this. So PEAKS will look at the MS/MS data

to determine if it's charge 1+, 2+ or 3+. This tool need only be used on ion trap data.

- **Remove Low Quality MS/MS scans:** Scans of contaminants and electrical noise should not be included in analysis. Removing them from the data set will save time, and reduce the risk of random matches to the database. PEAKS presents an effective tool for removing these low quality ms/ms scans. This tool has been designed for use on ion trap data.

- **Preprocess MS/MS scans:** Deconvolution (de-isotoping), centroiding and noise filtering within the MS/MS data. *Data is always saved in the ANZ file along with the PEAKS results. Preprocessing can save hard disk space or upload time. But make sure to have the original data available in case we need to refer to it later.*

A note on preserving data-results integrity

Protein ID and *de novo* sequencing results obtained for a given dataset prior to use of this tool may become invalid, since some spectra are removed/merged/corrected, and the data-results relationship may be broken. PEAKS Studio 4.0 will warn us when this may occur and prompt us to save a separate copy.

Using Peaks Studio with modifications (PTM)

PEAKS Studio 4.0 provides the most flexible handing of post-translational modifications of any software built for de novo sequencing and protein ID. Users are free to create their own modifications (see the **Creating a New PTM** section), and search for any combination and any number of modifications. Modifications can be considered as part of auto *de novo* sequencing or protein identification. The search is set up the same way for both tools. The options screen for each tool has an area titled "PTM selected for search". Any modifications to be considered during the search will be shown here (and labeled as 'Fixed' or 'Var'). When we first load PEAKS Studio, the box will be blank, meaning no PTM are selected.

PTM selected for search

Fixed	Carbamidomethylation
Var	Deamidation
Var	Oxidation
Var	Phosphorylation

Add/Remove PTM

To add modifications to this list, click the Add/Remove PTM button. The Modification dialogue appears. The entire PTM library – i.e. all <built in> and user defined modifications that are available to PEAKS – are displayed in the list on the left. We'll choose the modifications we need from this list. To add a new modification to the PTM library, click the 'New PTM' button. To edit a modification in the PTM library, select it from the list on the left and press the 'Edit PTM' button. To remove a modification from the PTM library, select it from the list on the left and press the 'Remove PTM' library.

> If you remove/edit a PTM, it will be removed or changed from/in any saved parameter set that refers to it.

The lists on the right show what PTM will be enabled for the search. Use the 'Select as Fixed=>', 'Select as Variable=>' and '<=Unselect' buttons to move them in and out of these lists. Press the 'Ok' button when finished, and the changes we made will be reflected on the protein ID options dialogue.
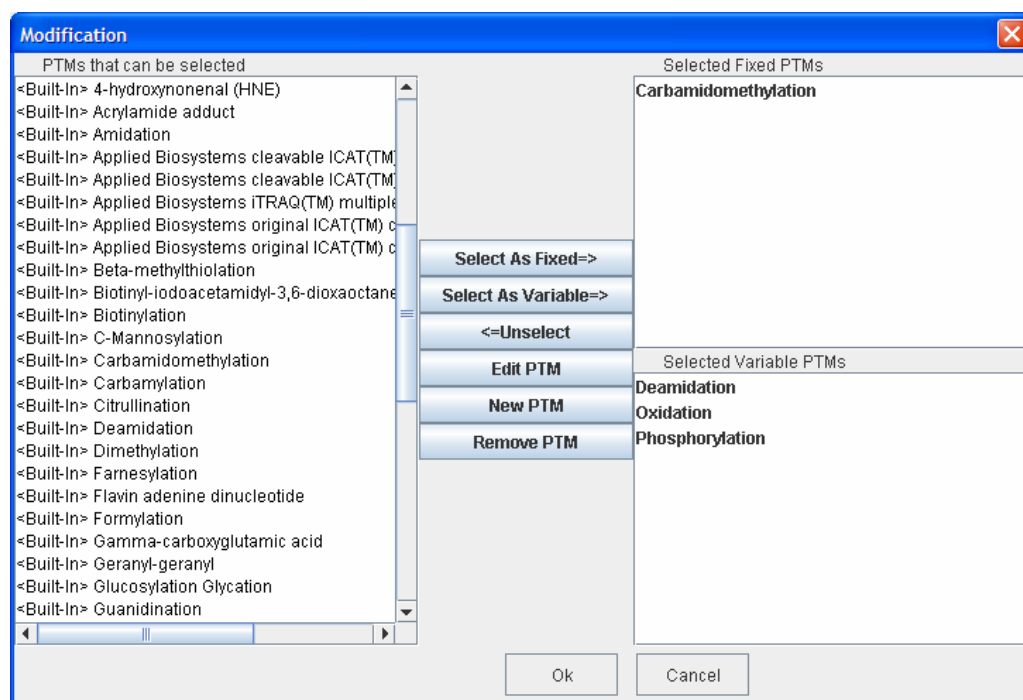
Generally, the more variable PTM we turn on, the more ambiguous will be the results.

Remember, when doing auto *de novo* sequencing or PEAKS Protein ID on a complex mixture, we will get best results if we choose the correct fixed PTM and a few variable PTM. When using PEAKS Protein ID to characterize a protein, it is best to search against a small database that contains only a few proteins, and turn on all modifications. Furthermore, to limit spurious hits, we can assume that it is less likely that a tryptic length peptide will not be modified more than a few times

and as such, limit the number of variable modifications that can occur on each peptide.

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 ![icon] 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:

We should begin by using the suggested error 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 mass error tolerance:** determine how much random and systematic experimental error on the parent/precursor ion PEAKS will allow for in its analysis. Select a tolerance from the dropdown list, or type in a value. New PEAKS users should try setting this a little higher than past experience may suggest.

▪ **Fragment mass error tolerance:** determine how much random and systematic experimental error on the fragment/daughter ion PEAKS will allow for in its analysis. Select a tolerance from the dropdown or type in a value. Again, new PEAKS users should try setting this a little higher than past experience may suggest.

▪ **Instrument Type:** choose the type of spectrometer that produced the data to be analyzed. If we are using a hybrid instrument, choose a setting that matches our fragment ion mass analyser. For example, if we measure the parent ion in an FT, and the fragment ions in an ion trap, choose the ion trap instrument setting. Fragmentation type can also be chosen from this drop down.

- **Enzyme:** Tell PEAKS what kind of enzyme was used to digest the sample. Choose from a dropdown list of enzymes, or if our enzyme (or combination of enzymes) is not in the list, click the "Edit Enzymes" button.

- **Report top:** set how many peptide sequences PEAKS will report from its de novo sequencing analysis.

- **Max missed cleavages:** determine the most missed cleavages to allow, internal to the peptide, in a de novo sequence. For instance, if we set this to 2, and Trypsin is the enzyme, then PEAKS will return de novo sequences with up to 2 R's or K's internally.

- **PTM selected for search:** this list tells PEAKS what kind of post-translational modifications to include in it's analysis. Each is marked Fixed or Variable. To edit this list, click the "Add/Remove PTM" button.

- **Max variable PTM per peptide:** To reduce uncertainty, we can limit PEAKS' de novo sequencing 'vocabulary', by restricting the number of variable PTM we can find on a peptide. Specify a number by typing it into the box. To lift such restrictions, type a very large number (longer than the length of the peptide).

- **Saving/Loading Parameters:** After setting up parameters, we can save them for future use. Click the "Save Parameters" button, and choose a name for future reference when prompted. Don't worry, we can't accidently overwrite the defaults. Any parameters we save will be available in the drop-down list at the top of the window. To see what's inside, just select one, and the parameters boxes will be populated.

- **Preprocess before auto de novo:** PEAKS Studio has its own built-in preprocessor for removing noise, centroiding, and peak charge recognition from MS/MS data. Check this box to turn preprocessing on.

> ### Notes on pre-processing
>
> BSI highly recommends using PEAKS to preprocess all data, as opposed to using instrument vendor software, if the data is to be used by PEAKS. PEAKS preprocessor should not be used on data that has already been pre-processed as this will have adverse effects on the results (unless it is ion-trap data).

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

Once a job is submitted to PEAKS Studio 4.0, it is added to the Task Queue for processing. After processing, the job is removed from the task queue list, and the ● icon beside the spectrum (in the Peptide Data Frame) changes to light green ● and/or an asterisk (*) appears.

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.*

Confidence scores are probability based, on a scale of 0 to 100%

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%), **purple** 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
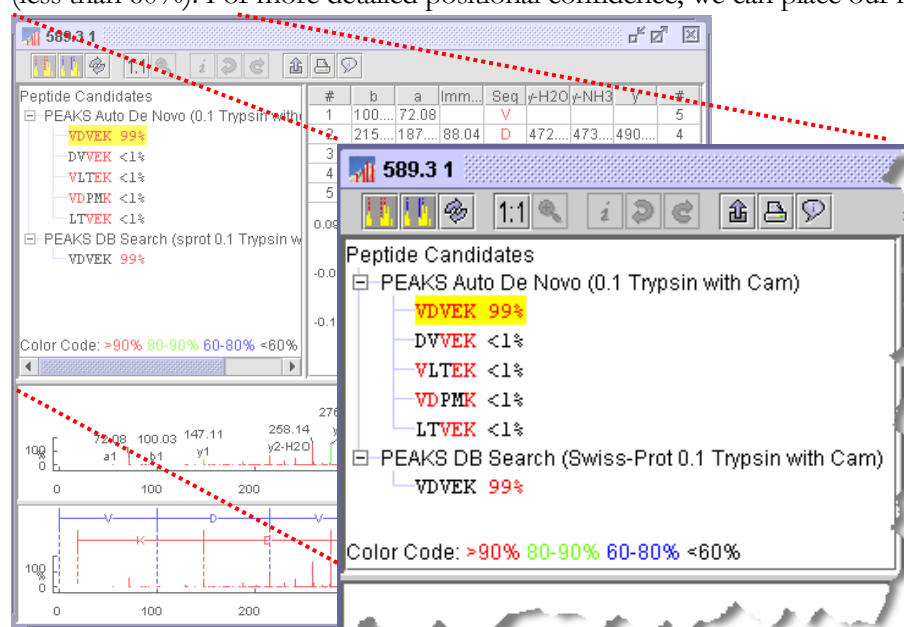

the sequence of interest. A Position Confidence Table will appear, showing the confidence that each tag/subsequence is correct.

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. The types of ions displayed in the ion table can be configured, choose Configuration → Edit Ion Table from the Edit menu. FTMS users might find this particularly useful when sequencing data acquired using ECD.

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 4.0 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 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.

Editing sequencing results (preparation)

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 4.0 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.
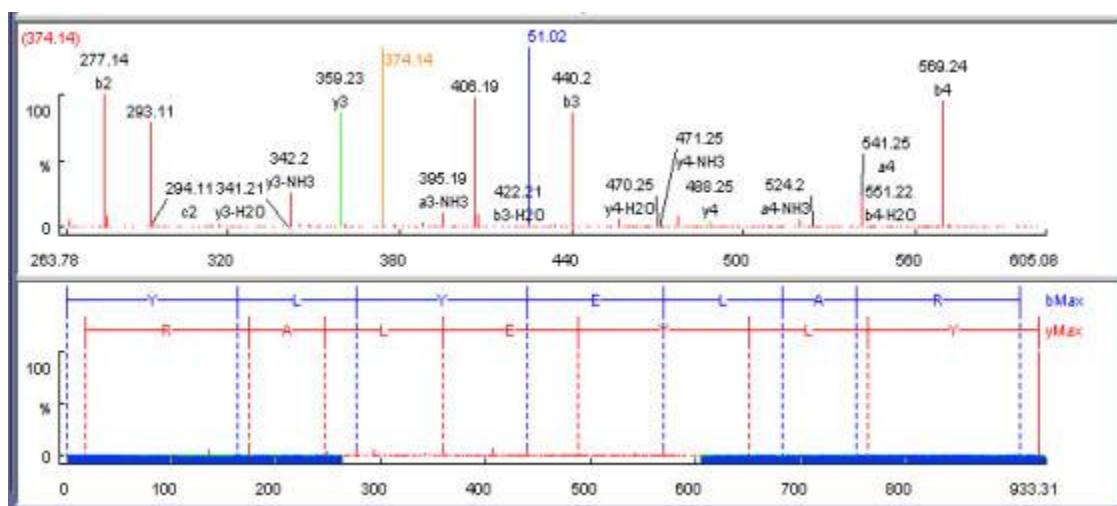
Measure the m/z difference between two PEAKS

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
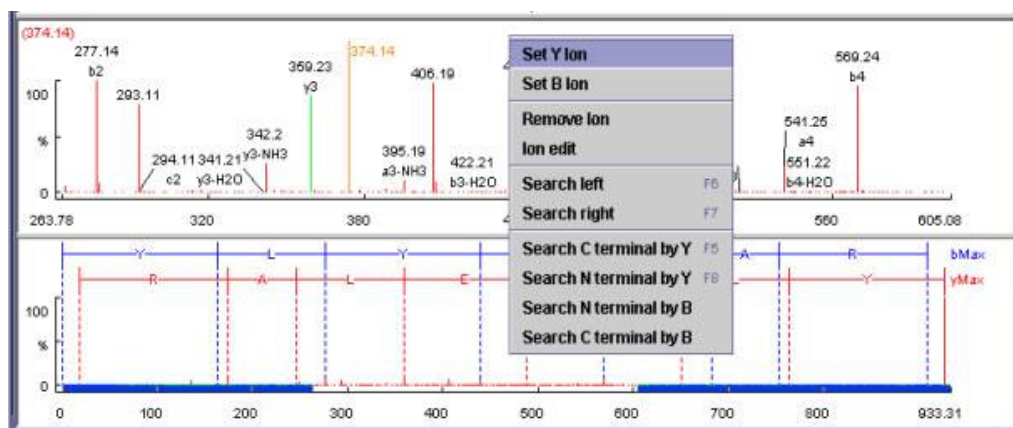
Deselect a peak

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.

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 4.0 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.

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.

# Suggesting a sequence to see how it fits the data

If the data is ambiguous, PEAKS Studio 4.0 may not have displayed a particular candidate that we wish to evaluate, after auto de novo, or protein ID. We may enter this sequence and have PEAKS Studio 4.0 find if there is any evidence for it in the data.

For instance, PEAKS may give the sequence RMYNVHGC(phosphorylationS)K for a particular spectrum, and we may wish to see if there's any evidence for the phosphorylation being on the Tyrosine. As such, we may type in our own version of the sequence and have PEAKS find ions that might support our hypothesis.

To do so, open the spectra in the Main Processing Window, and right click on "Peptide Candidates" (in the Peptide Candidates Frame). Then, from the pop-up menu that appears, choose "New Candidate for Manual De Novo".
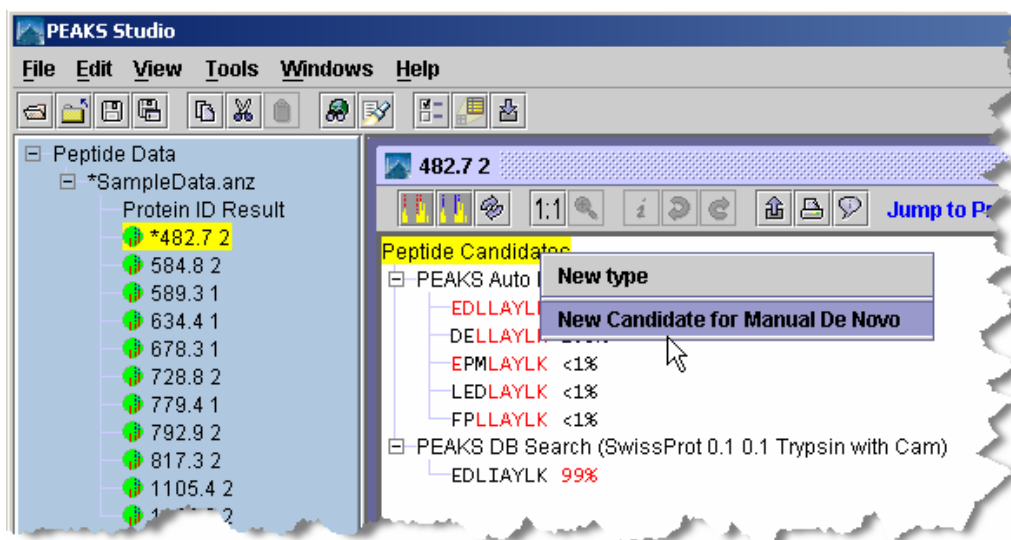
A new node will appear with the heading "Manual De Novo" and beneath it will be the mass of the residues yet to be sequenced (in square brackets). Right-click on this heading. In the pop-up menu that appears, choose "New Candidate with user input sequence", and the Sequence Input dialogue box will appear.
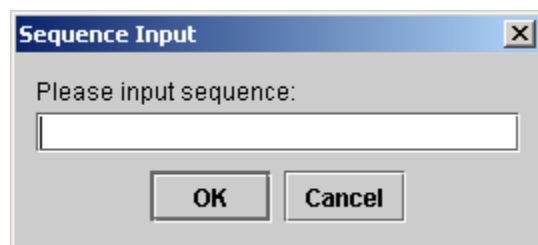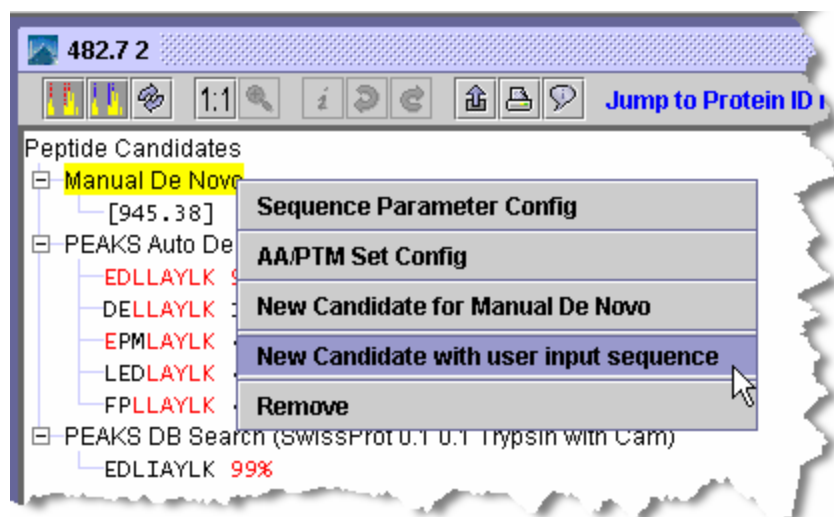
We can now enter our proposed sequence. The total mass of the residues, modifications, and un-sequenced masses should equal the total mass of the peptide (minus water). We might find the mass calculator tool (Tools menu), useful in this regard.

Short forms for the modifications may also be used.

Enter sequences in the format:
MPELAYLK
[228.09]ELAYLK
DE[226.168]AYLK
EDLLA(phosphorylationY)LK
DE[226.168]A(phosphorylationY)LK

Then press the OK button. The sequence we just entered will appear under the "Manual De Novo" heading, and when selected, the ions that PEAKS has found to match the proposed sequence appear on the spectrum, spectrum alignment view, and ion table.

Protein Identification

PEAKS introduces an amalgamative approach to protein identification called **inChorus**. With inChorus protein identification technology, we can use PEAKS together with several other protein identification methods. This will deliver more protein coverage and more confidence in results than any one method on its own.

An integral part of the inChorus search, PEAKS own protein identification method is unique – an improvement on and the ideal compliment to existing tools. The unique approach is a combination of sequence tag searching and fragment ion mass matching.

The following two sections deal with usage of PEAKS protein identification on its own, and usage of inChorus protein identification.

PEAKS protein identification

PEAKS Protein ID search engine is a hybrid approach that uses sequence tag information to filter the protein or EST database before fragment ion fingerprinting. 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 ![icon].

   Or

   Select "PEAKS Protein ID" from the tools menu.

The Protein Identification Parameters dialogue window will appear

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

- **Parent mass error tolerance:** determine how much random and systematic experimental error on the parent/precursor ion PEAKS will allow for in its

analysis. Select a tolerance from the dropdown list, or type in a value. New PEAKS users should try setting this a little higher than past experience would suggest.

- **Fragment mass error tolerance:** determine how much random and systematic experimental error on the fragment/daughter ion PEAKS will allow for in its analysis. Select a tolerance from the dropdown or type in a value. Again, new PEAKS users should try setting this a little higher than past experience would suggest.

- **Instrument Type:** choose the type of spectrometer that produced the data to be analyzed. If we are using a hybrid instrument, choose a setting that matches our fragment ion mass analyser. For example, if we measure the parent ion in an FT, and the fragment ions in an ion trap, choose the ion trap instrument setting. Fragmentation type can also be chosen from this drop down.

- **Enzyme:** Tell PEAKS what kind of enzyme was used to digest the sample. Choose from a dropdown list of enzymes, or if our enzyme (or combination of enzymes) is not in the list, click the "Edit Enzymes" button.

- **Report top:** set how many peptide sequences PEAKS will report from its de novo sequencing analysis.

- **Max missed cleavages:** determine the most missed cleavages to allow, internal to the peptide, in a de novo sequence. For instance, if we set this to 2, and Trypsin is the enzyme, then PEAKS will return de novo sequences with up to 2 R's or K's internally.

- **PTM selected for search:** this list tells PEAKS what kind of post-translational modifications to include in it's analysis. Each is marked Fixed or Variable. To edit this list, click the "Add/Remove PTM" button.

- **Max variable PTM per peptide:** To reduce uncertainty, we can limit PEAKS' de novo sequencing 'vocabulary', by restricting the number of variable PTM we can find on a peptide. Specify a number by typing it into the box. To lift such restrictions, type a very large number (longer than the length of the peptide).

Best practices for setting modifications (PTM)

The developers have discovered that database searching often returns better results if the auto *de novo* analysis 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.

- **Report up to:** Type a number here to specify the number of proteins (including all homologs but counting them as one entry) to include in the result report. PEAKS will report up to this many.

- **Database to search:** Select, from this dropdown list, one of the FASTA databases that we've set up in PEAKS. If the database we'd like to search is not in this list, click the "Load new database" button.

- **Taxonomy selection:** This list displays the taxa we've chosen for our search. If the database selected has taxon information available, we can click the aptly labeled "Add/Remove Taxa" button. Otherwise, the whole database will be searched. The selections correspond to established hierarchy -- i.e. selecting 'Mamalia' will search all of 'horse, cow, rat, mouse, human, etc.'

- **Preprocess before auto de novo:** PEAKS Studio has its own built-in preprocessor for removing noise, centroiding, and peak charge recognition from MS/MS data. Check this box to turn preprocessing on.

Notes on pre-processing

BSI highly recommends using PEAKS to preprocess all data, as opposed to using instrument vendor software, if the data is to be used by PEAKS. PEAKS preprocessor should not be used on data that has already been de-convolved by instrument software as this will have adverse effects on the results (unless it is ion-trap data).

PEAKS preserves the original data, and does not save the results of its preprocessing. As such, the decision to preprocess or not should be independent of what we've already done with PEAKS.

- **Advanced Options -** De novo: We must have some de novo sequences before database searching since PEAKS sequence tags to help in database searching. As such the option of doing de novo prior to protein ID is presented here. In most cases, the same values for instrument, error, enzyme and PTM can be used in de novo and in protein ID, but we have the option of using one of our saved de novo parameter sets for the de novo portion. Select from the drop down list.

- **Saving/Loading Parameters:** After setting up parameters, we can save them for future use. Click the "Save Parameters" button, and choose a name for

future reference when prompted. Don't worry, we can't accidently overwrite the defaults. Any parameters we save will be available in the drop-down list at the top of the window. To see what's inside, just select one, and the parameters boxes will be populated. Note: the Advanced Options selections will not be saved.

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:
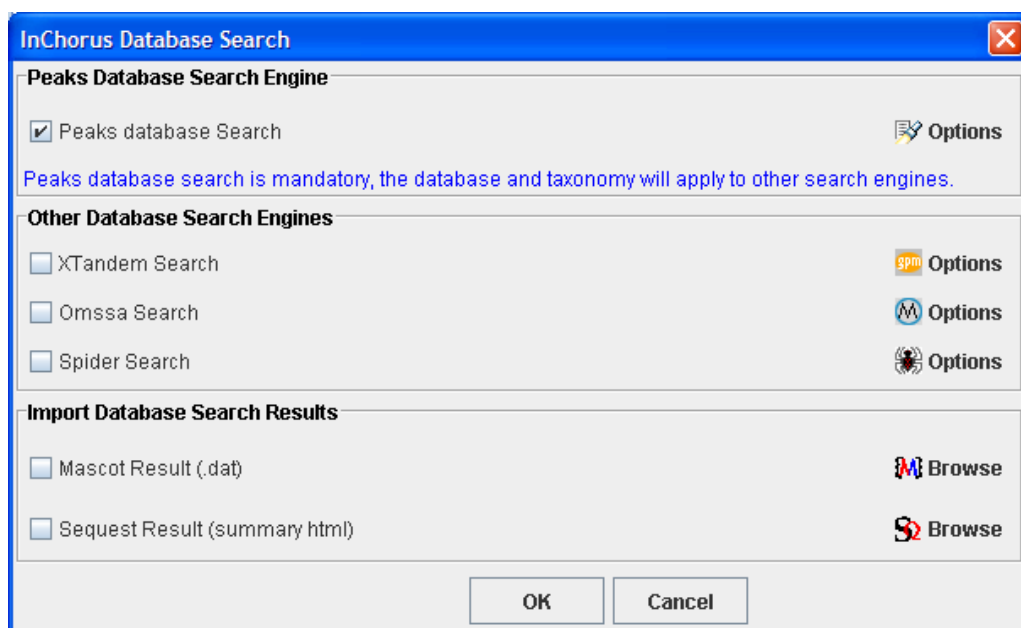
inChorus protein identification

inChorus protein identification will call upon several search engines for protein identification. Once we load our data into PEAKS, we can invoke (start searches running on) several search engines at once. When all the results are returned, PEAKS Studio 4.0 will compare the answers and summarize everything in one simple report.

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 "inChorus Protein ID" from the tools menu

The inChorus database search launch window will appear:

3. First select each of the protein identification tools we would like to use by putting a checkmark in their respective checkboxes. Search parameters for each program can be set by clicking the corresponding Options icon. PEAKS database search engine is mandatory.

4. Then set search parameters for each search engine. Options screens each of the programs available to inChorus are designed to work in the same way as options screens for the original programs. For help in setting search parameters for each program, please refer to that program's user manual. In the case of PEAKS database search, please refer to the above section.

Viewing Protein identification results

To view Protein identification results for a data file, we must have performed PEAKS protein identification or inChorus protein identification on that data file. The result from each protein identification search is represented by the time stamp and database searched (just under a data file's "Protein ID Result"). Click on one to display the results report. We can view results by peptide or by protein, and check on the search parameters we used to generate these results.

Search parameters that we used when generating this report are preserved for future reference, and are available by clicking on the search parameters tab.

Peptide view is available by clicking on the peptide view tab. In tabular format, it displays relevant information about each peptide found. Since two peptides may match to one MS/MS spectrum, they are visually grouped together (using colour) by MS/MS scan.

Scores associated with each peptide represent the quality of the match and the number of programs in agreement on the peptide.

SampleData.anz													
Peptide View	Protein View	Search parameters											
Index	m/z	z	mass	Peptide	Mr(Calc)	Delta(M..	Score (..	Start	End	Accession ...	Peaks	XTandem	Omssa
1	482.7	2	963.4	EDLLAYLK	963.5276	0.1276	71.67	91	99	CYC_BOVIN	✔	☐	☐
1	482.7	2	963.4	AEKNFDLK	963.5025	0.1025	62.22	1379	1387	ACFD_ECO...	✔	☐	☐
2	584.8	2	1167.6	TGPNLHGLFGR	1167.615	0.015	99.44	27	38	CYC_BOVIN	✔	✔	☐
2	584.8	2	1167.6	EAPPGNAKAGEK	1167.5...	0.0116	60.0	5	17	CYC2_ARA...	✔	☐	☐
2	584.8	2	1167.6	TGPNLHGIFGR	1167.615	0.015	71.67	28	39	CYC_APTPA	☐	✔	☐
3	589.3	1	588.3	VDVEK	588.3119	0.0119	71.11	1048	1053	ACFD_ECO...	✔	☐	☐
3	589.3	1	588.3	EKGKK	588.3595	0.0595	62.78	3	8	CYC_BOVIN	✔	☐	☐
4	634.4	1	633.4	LFVQK	633.385	0.015	71.11	8	13	CYC_BOVIN	✔	☐	☐
4	634.4	1	633.4	KEMVK	633.352	0.048	63.89	180	185	CTN2_MO...	✔	☐	☐
5	678.3	1	677.3	YLPGTK	677.3748	0.0748	71.11	73	79	CYC_BOVIN	✔	☐	☐
5	678.3	1	677.3	EARFR	677.3609	0.0609	61.67	295	300	CTN2_MO...	✔	☐	☐
6	728.8	2	1455.6	TGQAPGFSYTDANK	1455.6..	0.0631	99.72	39	53	CYC_BOVIN	✔	✔	☐
6	728.8	2	1455.6	DAYAGGTKCSADLK	1455.6..	0.0665	60.56	986	1000	ACFD_ECO...	✔	☐	☐
7	779.4	1	778.4	MLFAGLK	778.4411	0.0411	71.11	79	86	CYC_BOVIN	✔	☐	☐
7	779.4	1	778.4	KLFLMK	778.4775	0.0775	63.33	7	13	CYC_HUM...	✔	☐	☐
8	792.9	2	1583.8	KTGQAPGFSYTDANK	1583.758	0.042	100.0	38	53	CYC_BOVIN	✔	✔	☐
8	792.9	2	1583.8	SEDTLYEYLLNPK	1583.772	0.0281	61.11	67	80	CYC_SPIOL	✔	☐	☐
9	817.3	2	1632.6	AEMENYEAGVYTEK	1632.6..	0.0978	70.56	555	569	CTN2_MO...	✔	☐	☐
9	817.3	2	1632.6	ASWLEDVDGKQETR	1632.7..	0.1744	65.0	820	834	ACFD_ECO...	✔	☐	☐
10	1105...2		2208.8	CREWFWLRLYYLYLK	2208.1..	0.6658	70.0	427	442	GRT1_SCH...	✔	☐	☐

When we first load the report, it is sorted by MS/MS index. This is analogous to scan number or DTA file name (unless spectra have been merged). In the example above, there were two matches returned for MS/MS spectrum #1, EDLLAYLK and AEKNFDLK. For spectrum #2, there were three possible peptides matching, for number 3 there were two, and so on. Each peptide is given a score, and the protein it matches is displayed for reference, under "Accession No." If the report was generated after inChorus protein ID, each search program that agreed on the peptide is given a checkmark under its column.

This list is sort-able. Click on a column header to sort the list using that columns values. For example, sort by score. Clicking again on the same column header will toggle ascending/descending sort. When sorting, the grouping by MS/MS scan number will be retained except when sorting by Accession No.

Clicking on this report will highlight the spectrum in the Peptide Data tree on the left. Select multiple spectra by clicking and dragging, using shift+click or ctrl+click. In this way, the highest scoring peptides may be selected and isolated for further analysis.

Protein view is accessed by clicking on the Protein View tab. It collects all the peptide identifications together, summarizes which proteins were present in the sample, and groups homologous proteins together. The same information is displayed in the peptide view as in this protein view, however the results are organized to best enable us to evaluate at the protein level.
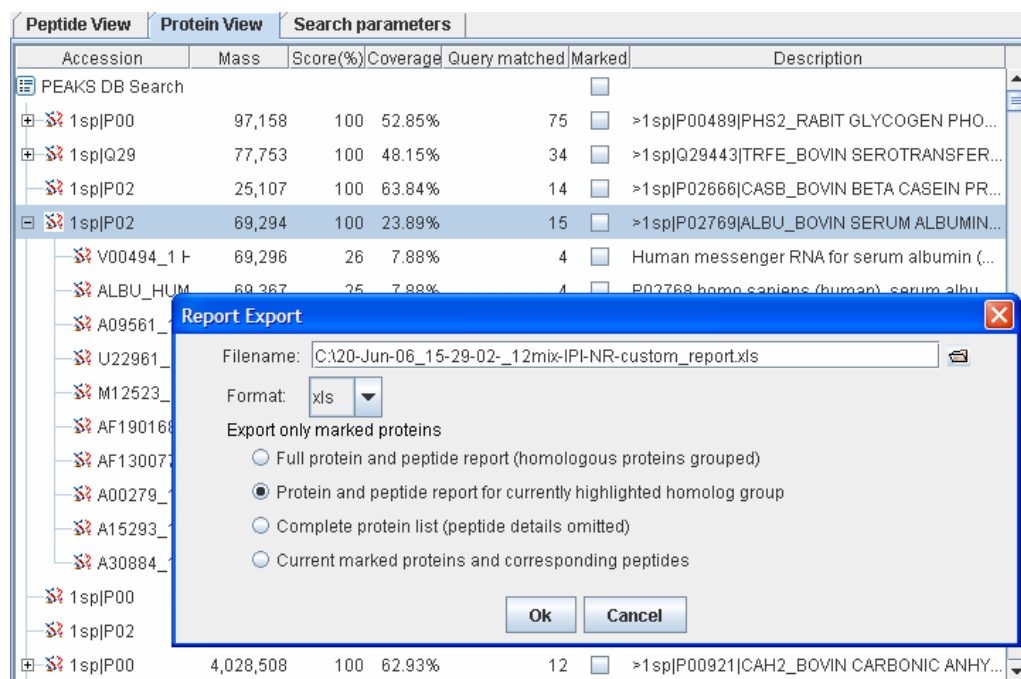
This section of the report behaves like an index, listing each protein found in the sample. Very similar proteins, containing the same set or a subset of the matched peptides, are grouped together. To see the full list of proteins within each grouping, click the '+' sign. In the example above, the Bovine Serum Albumin node has been expanded to reveal several similar proteins. Clicking in the same place (now a '—' sign) will collapse the list.

This view is helpful when building a summary that can be sent to a customer/collaborator. Simply right click to export to an MS Excel file. We can export interesting parts of the report, or a whole summary. Mark proteins of interest by clicking their checkboxes, and export protein and peptide information for those. Or highlight a homologue group and export proteins and peptides in that group. Or just export the whole report.

Try the features walkthrough in Chapter 3 for more help with viewing Protein ID results

Whenever we click on a protein, the Matched Peptides panel (bottom) changes to display the spectra (and peptides) that were found to be supporting evidence for that protein. It may be necessary to scroll this panel down to see the complete list. Clicking a peptide in this list brings up the Main Processing Window for the corresponding spectrum, and displays the ions that were found in support of this peptide.

Scrolling the Matched Peptides panel further down reveals the complete sequence of the highlighted protein, with the matched peptides highlighted in red. Where an EST database was used, the translated sequence is shown (with all six reading frames concatenated).

Manually Manipulating Data Files

## A note on preserving data-results integrity

Protein ID and *de novo* sequencing results obtained for a given dataset prior to use of the tools in this section may become invalid, since some spectra are removed/merged/corrected, and the data-results relationship may be broken. PEAKS Studio 4.0 will warn us when this may occur and prompt us to save a separate copy.

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 ANZ 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.
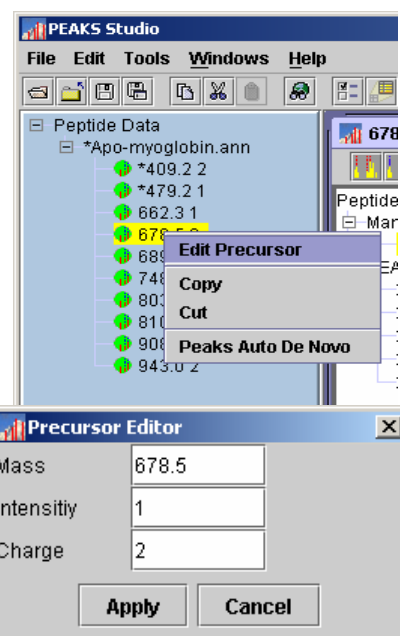
The precursor information will be updated, reflected by a change in the name of the spectrum in the Peptide Data Frame. A * will also appear in front of that name, indicating that there is unsaved information pertaining to that spectrum.

Manually merging MS/MS scans of the same peptide

If we've done several MS/MS scans of the same peptide, we may want to reduce the amount of data to process, and at the same time, improve the data quality, by merging all of a peptide's MS/MS scans together. Often, we choose to automatically merge appropriate spectra from the whole data file using the Data Refine tool (see above). But this can also be done manually.

To manually merge spectra after opening a data file:

1. Select those spectra we wish to merge together from the Peptide Data Tree (left) using shift+click and ctrl+click.

2. Next, right click in the Peptide Data panel, and choose Merge Spectra from the popup window that appears, *OR* click the ⇥ 'Manual Merge Spectra' toolbar button.

3. A dialogue will appear, asking what should be the correct value for the precursor mass and charge. After reviewing and/or correcting the value, press ok. The spectra will be merged.

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

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

   OR

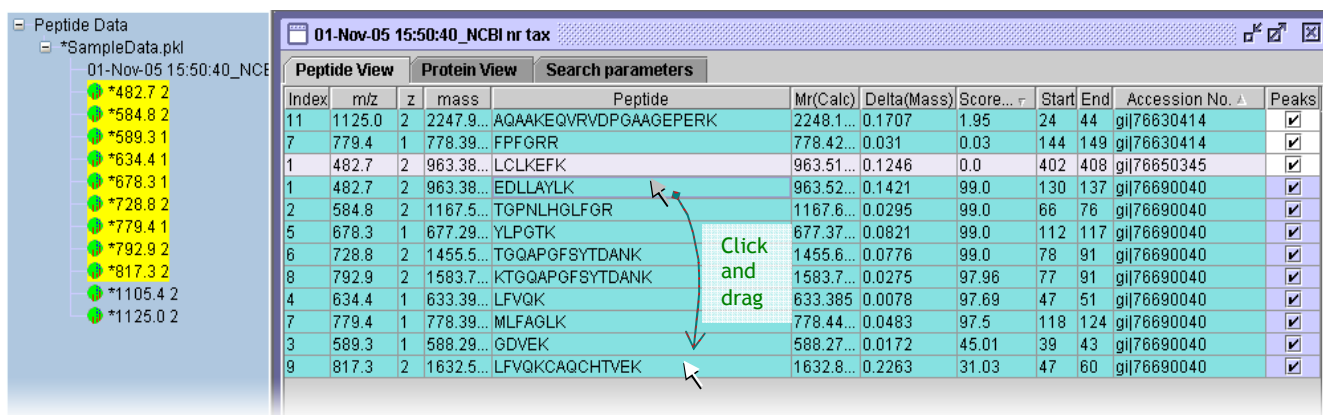   Click the "Paste" button 🗐 in the main toolbar.

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

Selecting unmatched de novo results

If we know which proteins to look for, we can create our own FASTA database to search against.

When working with unknown organisms, abundant but uninteresting proteins (like keratin) can get in the way. We may find it more convenient to eliminate them from the analysis. To do so, we must first identify which peptides belong to those proteins. To do so, run PEAKS Studio 4.0's protein identification tool. The proteins identified in the sample will be shown in the protein ID report.

In the Protein ID Result report (peptide view), first sort by accession number by clicking on the header for that column. Next scroll down to an abundant, but uninteresting protein. Click and drag downwards to select those spectra matching to that protein.


They appear highlighted in the peptide data tree on the left. We may then chose run auto de novo, or protein id on these spectra, or on everything but these spectra. Read on…

## 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:
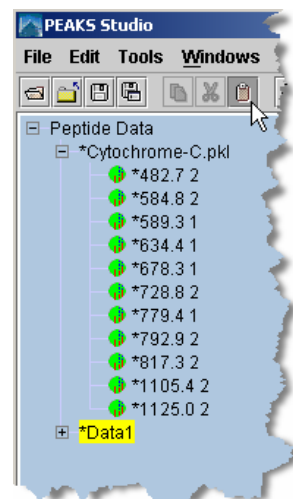
1. Make a new Data node by right clicking on the peptide data node. The new node appears as "Data1".
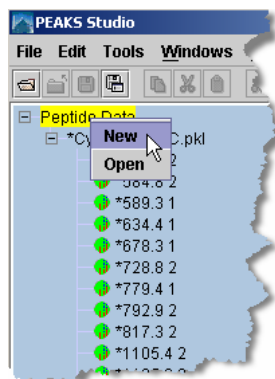
2. Select the relevant spectra using <shift>+click and <ctrl>+click. Then press the cut button..
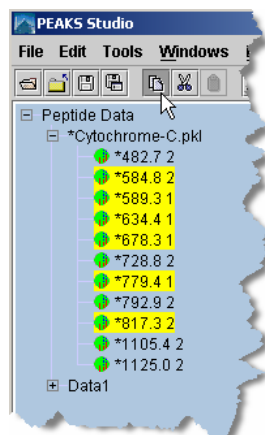
3. Click the new node and press the paste button. Pressing the + next to Data1 will expand it and reveal the pasted spectra


Now we've essentially removed the already matched peptides from our dataset. We can now run protein identification on Data1, or on the remaining spectra in our original dataset. 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.
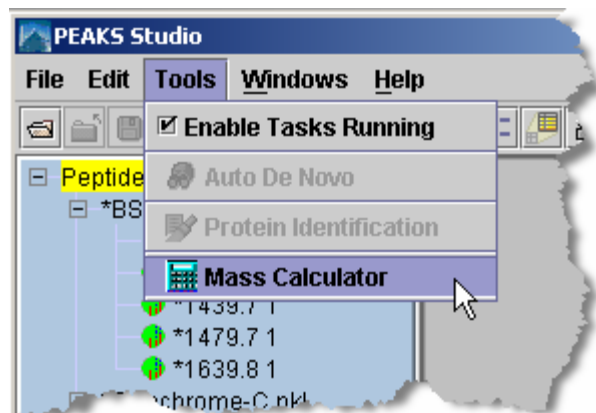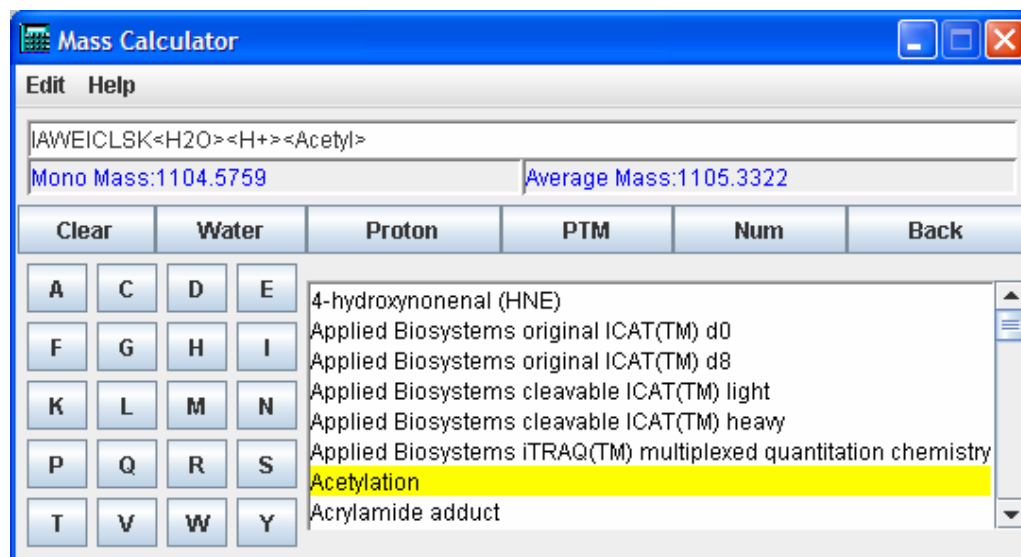
- Amino acids are represented by their single letter symbols. Clicking on an amino acid's button will add it to the 'sequence' above, and add its mass to the mass of the peptide. Note that the peptide's monoisotopic and average masses are both computed.

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

Creating a high-throughput workflow

In some situations, we may have many data sets that we wish to process all at once, and in the same way. PEAKS Studio 4.0 allows us to do this kind of work, and with minimal effort on our part. By setting up a workflow, we can start a batch process of several data files, and not worry about it until it is finished.

It is important to note that all the files we load will be processed in exactly the same way, using exactly the same parameters. If we want to do some differently than others, we must set up another workflow.
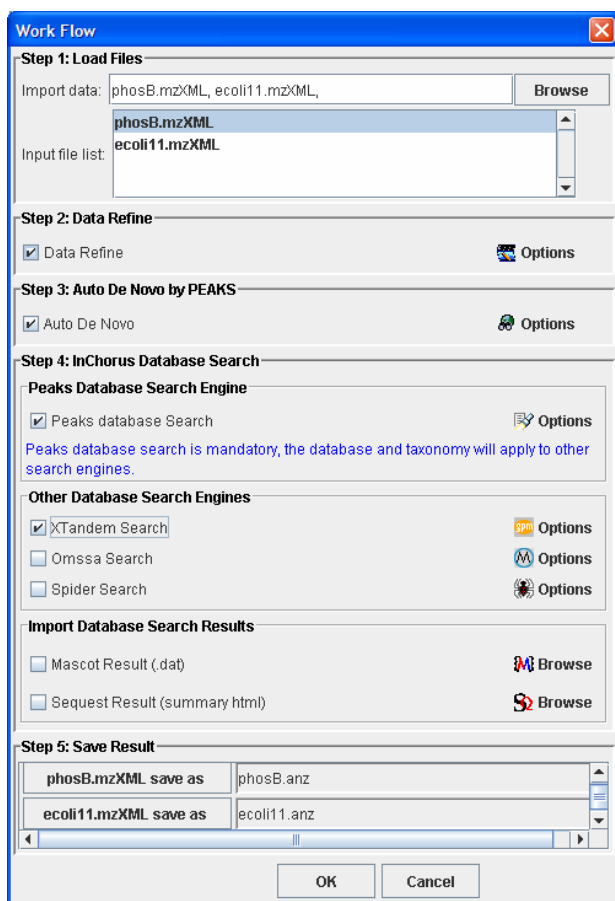
Step1: Load files – click the browse button to open a file chooser. From the chooser, select several files by shift+click or ctrl+click, and pressing the "OK" button. Load more files by pressing browse again, or remove them from the list by right clicking on them.

Step2: Data Refine – choose how to filter and correct data for maximum utility.

Step3: Auto de novo – choose whether or not to do auto de novo sequencing. Note that PEAKS database search requires some de novo sequencing results.

Step4: inChorus protein identification – choose which protein identification programs with which to run the data. PEAKS database seasrch is

Step5: Save results – saved automatically into an .ANZ file with the same name as the data file. All files will be placed in one folder. Typing in the textboxes or clicking each file's button changes the name and/or save location.

Saving Results

Saving results will preserve our work for later use. Saving files in PEAKS's .ANZ 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 ANZ format and of the same name as the data file we opened.

To change the name of the ANZ file, choose "Save as" from the "File" menu, or right click on the data file and select "Save as" from the popup menu. We 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 peptide sequencing 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.

To export protein identification results (including protein and corresponding peptide information) to an MS Excell file, right click in the protein view. PEAKS will then ask us which portion of the results we'd like to export. The resulting .xls file contains a collapsible list.

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.

PEAKS Software License

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

If we are evaluating a time limited trial version of PEAKS, and we wish to update the software to the full version, we 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 its 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.

Reference: PEAKS Paper

Please use the following reference when publishing a study that involved the use of PEAKS.

Bin Ma, Kaizhong Zhang, Christopher Hendrie, Chengzhi Liang, Ming Li, Amanda Doherty-Kirby, and Gilles Lajoie. PEAKS: Powerful Software for Peptide De Novo Sequencing by Tandem Mass Spectrometry. *Rapid Communication in Mass Spectrometry* 17(20): 2337-2342. 2003