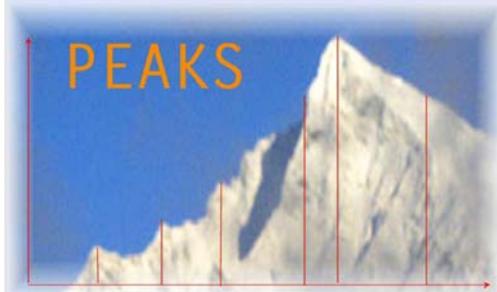


PEAKS Studio User's Manual



BIOINFORMATICS SOLUTIONS INC

PEAKS Studio User's Manual

© Bioinformatics Solutions Inc.
145 Columbia St. West Suite 2B
Waterloo, Ontario Canada N2L 3L2
Phone 519-885-8288 • Fax 519-885-9075

INTRODUCTION OF PEAKS 2.0.....	2
PEAKS BASICS.....	5
POST-TRANSLATIONAL MODIFICATIONS (PTMS).....	5
<i>Built-in PTM.....</i>	5
<i>Customized PTM.....</i>	5
AMINO ACIDS (AAS)/PTMS SET.....	6
FASTA DATABASE.....	6
FILE FORMATS.....	6
FILE SAVE/EXPORT.....	7
PEAKS PROPERTIES.....	7
PEAKS ENVIRONMENT PREFERENCE.....	7
CONFIGURATION FILE.....	8
LOGGING FILE.....	8
PEAKS STUDIO QUICK START.....	9
CONFIGURE FASTA DATABASE.....	9
OPEN A SPECTRUM FILE.....	10
AUTO DE NOVO SEQUENCING.....	11
INSPECTING THE SPECTRUM.....	13
PROTEIN IDENTIFICATION.....	14
VIEWING PROTEIN IDENTIFICATION RESULTS.....	15
SAVING RESULTS.....	16
CONFIGURATION.....	17
PEAKS PROPERTIES CONFIGURATION.....	17
<i>AAs/PTMs Set editor.....</i>	18
<i>PTM editor.....</i>	21
<i>Database Manager.....</i>	23
<i>Import and Export PEAKS Properties.....</i>	26
PEAKS ENVIRONMENT PREFERENCE.....	26
<i>Environment.....</i>	26
<i>Color.....</i>	26
<i>Manual De Novo Parameters.....</i>	27
<i>Auto De Novo Parameters.....</i>	27
PEAKS STUDIO USER INTERFACE.....	29
TOOL BAR.....	29
LOAD DIRECTORY.....	30
DATA TREE.....	30
PROTEIN IDENTIFICATION RESULT WINDOW.....	32
<i>Protein Candidate Tree.....</i>	32
<i>Protein Candidate Panel.....</i>	32
Index Panel.....	33
Information Panel.....	33
<i>Main Process Window.....</i>	34
Peptide Candidates panel.....	35
Ion Table Panel.....	35
Spectrum View and Alignment View panels.....	36
Tool Bar of Main Process Window.....	37
MANUAL DE NOVO SEQUENCING.....	38
HOW TO START AUTO DE NOVO SEQUENCING AND PROTEIN IDENTIFICATION.....	41
FREQUENTLY ASKED QUESTIONS.....	43

Introduction of PEAKS 2.0

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

PEAKS is an innovative software system designed to derive amino acid sequences and identify proteins from MS/MS experimental peptide data. Two versions of PEAKS are available: **PEAKS Studio** and **PEAKS Batch**.

PEAKS Studio is the standalone module that provides peptide sequence and protein identification results via an intuitive interface, allowing for rapid visual interpretation. PEAKS Studio provides both auto and manual de novo tools for detailed examination of MS/MS spectra, providing the flexibility to manually modify auto de novo results when searching for additional sequence possibilities. The y-ion/b-ion searching algorithms can provide potential ion candidates previously not considered.

PEAKS Batch is a scriptable, high throughput version, designed for large quantities of MS/MS data. PEAKS Batch is a powerful part of the protein identification pipeline, rapidly providing peptide sequences to facilitate data mining or statistical analysis. Additionally, the Batch version is a valuable tool when used in conjunction with database search methods. Spectra that provide poor or negative database search results may yield valuable information when sequenced by PEAKS - this information no longer needs to be ignored or discarded.

Both **PEAKS Studio** and **PEAKS Batch** share the following features:

- **Supreme Accuracy.** PEAKS correctly derives more peptide sequences from tandem MS/MS data from any mass spectrometer than any other software currently available.
 - **Post-Translational Modification Detection.** PEAKS automatically detects user-defined post-translational modifications.
-

INTRODUCTION

- **Full Automation.** PEAKS accepts a peak list in Micromass (.PKL), Sequest (.DTA) formats, or Mascot Generic Format (.MGF), and derives the peptide without the need of human interaction.
- **Convenient Output.** Exports to both FASTA and XML formats.
- **Superior Efficiency and Speed.** PEAKS typically derives one sequence in under five seconds on a standard PC.
- **No Database Required for deriving the amino acids sequence.** PEAKS efficiently searches all possible combinations of amino acids using a global optimization algorithm. No protein/DNA database is required for deriving the amino acids sequence from MS/MS experimental data.
- **Great Error-Tolerance.** PEAKS automatically estimates the noise level and the accuracy of the instrument that produced the data, adjusts the calibration errors caused by temperature changes and accounts for most types of ions into the computation.
- **Instrument Optimization.** PEAKS is optimized for TOF-TOF and Q-TOF instruments.
- **Platform Independent.** PEAKS runs in any Java 1.3.1 environment with reasonable hardware and memory requirements:

Pentium 500 or better, or any architecture where Java 1.3.1 runtime environment is supported. 1024MB memory required.

PEAKS Studio - Additional Features

- Comprehensive manual sequencing tools, including algorithms for y/b ion search, C/N terminal search.
- Easy-to-use informative interface provides access to spectra, sequence, confidence and y/b ion information.
- Zooming, labeling, annotation and peak distance tools.
- HTML Report Editor provides easy access to spectra, fragment and sequence information for fast document creation.

PEAKS Batch - Additional Features

- Scriptable, command line driven interface.
- Rapid de novo sequencing - less than 5 seconds per spectrum average on a single, 1 GHz CPU with 1 GB RAM.

I N T R O D U C T I O N

- Multi-threaded for higher throughput on multi-CPU systems.
- Both XML and FASTA output containing sequence confidence score results.

PEAKS Basics

Terminologies and concepts used in PEAKS

Terminologies and concepts used in PEAKS 2.0 will be introduced in this chapter. We will use these terminologies and concepts in the following chapters.

Post-translational Modifications (PTMs)

A PTM is the modification of a newly formed protein. This may involve amino acid deletion, chemical modification of certain amino acids, or the addition of small molecules (e.g., phosphate groups or sugars) to certain amino acids.

During the automated de novo and protein identification process, PEAKS automatically detects user specified post-translational modifications. PTMs used in PEAKS can be grouped into two categories:

Built-in PTM

PEAKS integrates a number of typical PTMs as built-in PTMs. These built-in PTMs cannot be deleted, but they can be edited. Once changed, PEAKS no longer consider them as “built-in PTM”. However deleting this edited PTM will result in the restoration of the original built-in PTM.

Customized PTM

You can add a new, custom post-translational modification by specifying the modification either by monoisotopic mass difference or by entering its empirical formula.

Note that some PTMs, (such as phosphorylation) can induce specific neutral losses (e.g. beta-elimination). Accordingly it be necessary to simulate the possibility of these neutral losses through two separate PTMs (e.g. one with beta elimination, one without).

Amino Acids (AAs)/PTMs Set

An AAs/PTMs set contains enzyme and modification information. When considering PTMs, PEAKS requires this information for both auto de novo search and protein identification. Both PEAKS Studio and PEAKS Batch provide tools with a graphical user interface to create custom AAs/PTMs set. PEAKS provides two options when setting up a PTM profile:

Fixed PTM

PEAKS considers a Fixed PTM to be a modification present universally, to every instance of the specified residue(s) or terminus.

Variable PTM

PEAKS considers a Variable PTM to be a modification that may or may not be present on a specified residue.

FASTA Database

PEAKS 2.0 supports protein identification from FASTA format databases. Before using the protein identification functionality in PEAKS, the appropriate database(s) must be acquired and configured for use in PEAKS. This process involves navigating to the location of the database file and specification of a database a name. As this name will be saved in a result file, it is recommended organizations standardize database names to ensure continuity.

File Formats

PEAKS accepts four kinds of data formats for reading spectra: Micromass (PKL), Sequest (DTA), Mascot Generic Format (MGF) and BSI's XML-based ANN file.

The PKL format supports multiple MS/MS spectrum datasets in a single file. The first line of a PKL dataset contains the observed m/z , intensity, and charge state of the precursor peptide as a triplet of space separated values. Subsequent lines contain space-separated pairs of fragment ion m/z and intensity values. Multiple MS/MS spectrum datasets are delimited by at least one blank line.

The DTA format is very simple. It only supports one MS/MS spectrum dataset in a file. The first line contains the singly protonated peptide mass (MH^+) and the peptide charge state as a pair of space separated values. Subsequent lines contain space-separated pairs of fragment ion m/z and intensity values.

PEAKS BASICS

Mascot Generic Format is described on Matrix Science's website. It supports multiple spectra per file.

The ANN format is an annotated data file in XML format. It supports both a single MS/MS spectrum and multiple spectra. The ANN file format is consisted of ANN data file and ANN index file.

One ANN data file contains the MS/MS information and peptide information of one spectrum. If the user's data file contains only one spectrum, the annotated data file will be saved as an ANN data file. The ANN data file includes the peak list of the spectrum, peptide results and protein id results.

The ANN index file provides a way to organize multiple spectra as a whole. It contains ANN data file information and protein ID results of multiple spectra. While the user's annotated data file contains multiple spectra, each spectrum will be saved as an ANN data file that contains the peak list and peptide results of this spectrum. These ANN data file will be saved in one directory. One ANN index file will be created when the annotated data file contains multiple spectra. This index file contains the following information: the directory name where the corresponding ANN data files located; the filename of ANN data files; and the protein id results of these spectra.

Note: The directory name saved in ANN index file is just the NAME of the directory, not the absolute path of the directory. So the ANN index file and the directory MUST be saved in same location. This is designed to avoid problems during annotated data files sharing.

File Save/Export

Both PEAKS Studio and PEAKS Batch save result information in an ANN file. Data can be exported into the PKL file format and FASTA file format.

PEAKS Properties

PEAKS Properties contains the value of a PTM, AAs/PTMs set and FASTA database configuration. This information is essential for PEAKS auto de novo and PEAKS protein identification.

PEAKS Environment Preference

Options for default Data Input, Output and Configuration file directories can be changed to increase efficiency. (In the current version of PEAKS, the Default Configuration File Directory cannot be changed.)

User can also customize the options for color of spectra viewing, auto de novo and manual de novo in PEAKS Environment Preference.

Configuration File

PEAKS software requires a configuration file to store the value of PEAKS Properties (and PEAKS Environment Preference for Studio). The configuration file is in XML format.

For PEAKS Studio, the configuration file will be saved in the configuration file directory. User can check the PEAKS Environment Preference to know the location of the configuration file directory. The filename of the configuration file is 'peaksconf.xml'.

For PEAKS Batch, user can use the configuration file created by PEAKS Studio, or create a customized configuration file with PEAKS Batch, or write one with any text editor.

Logging File

PEAKS software saves all runtime information in a logging file. The logging file is text format. When user reports a bug related to PEAKS, the logging file will help BSI technical support staff tracing the bug.

For PEAKS Studio, the logging file is located in the configuration file directory. User can check the PEAKS Environment Preference to know the location of the configuration file directory. The filename of the logging file is 'peaks.log'.

For PEAKS Batch, the logging file will be saved in the output directory of the job.

PEAKS Studio Quick Start

Getting started with PEAKS Studio

PEAKS is a software program for determining peptide sequences and identifying proteins from MS/MS spectra. It can be used in three principal ways:

- Auto de novo sequencing: no user intervention, no database requirement
- Manual sequencing: tools for a skilled user that assist the sequencing process
- Protein identification using de novo and FASTA database searching

Configure FASTA database

The first time PEAKS Studio is started, PEAKS will ask if you wish to configure the FASTA database required for PEAKS Protein Identification functionality. A database configuration wizard will provide information indicating how to download and configure the database for PEAKS Protein Identification.

To configure a FASTA database:

1. Choose a FASTA format protein database;
 2. Download the database and save the file to a directory accessible by PEAKS;
 3. Use the database configuration wizard or PEAKS Properties Editor to:
 - a. Locate the database file
 - b. Specify a name for this database.
-

PEAKS STUDIO QUICK START

Protein identification functionality will remain disabled until one FASTA protein database has been configured for PEAKS Studio.

Open a Spectrum File

To open a demo data file, click the  icon on the toolbar in the upper left corner of the PEAKS window or select "Open" from the "File" menu. PEAKS demo data can be found in the DATA sub-directory, located in the PEAKS directory. Each spectrum is represented by its precursor ion information (m/z value and the charge of the precursor ion that generated the spectrum).

In the following demo images, we select the data file 'positive.pkl'.

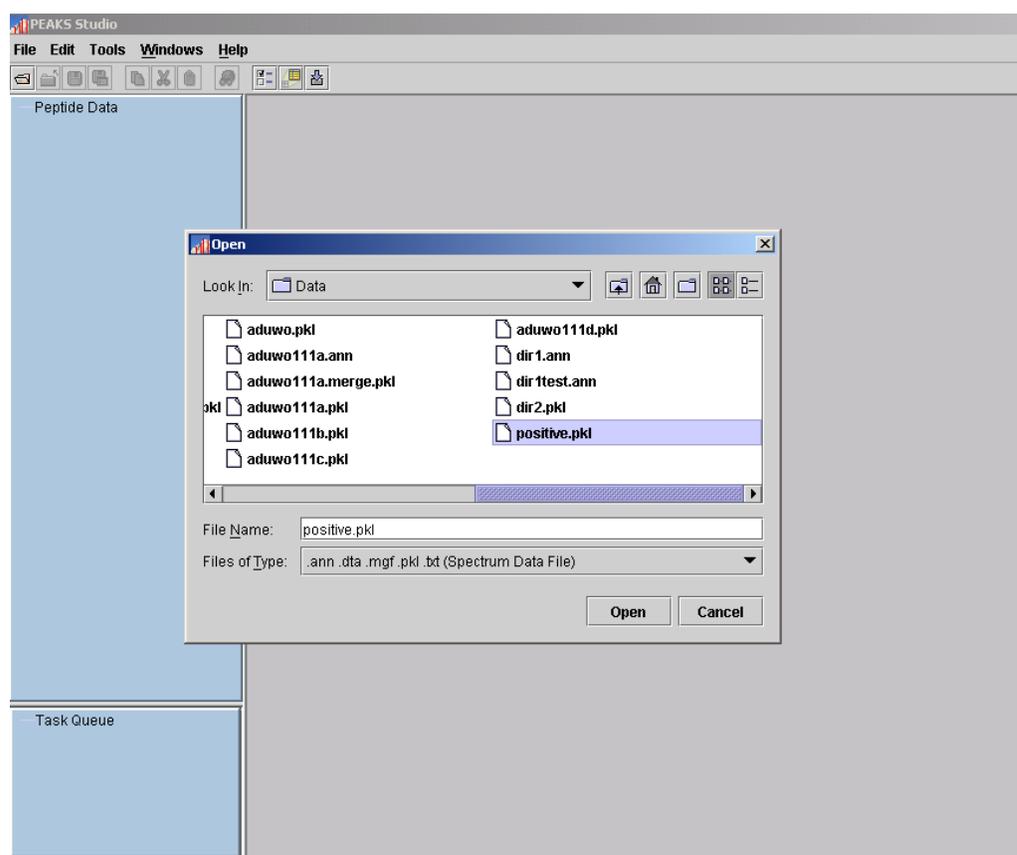


Figure 1 Select data file

PEAKS STUDIO QUICK START

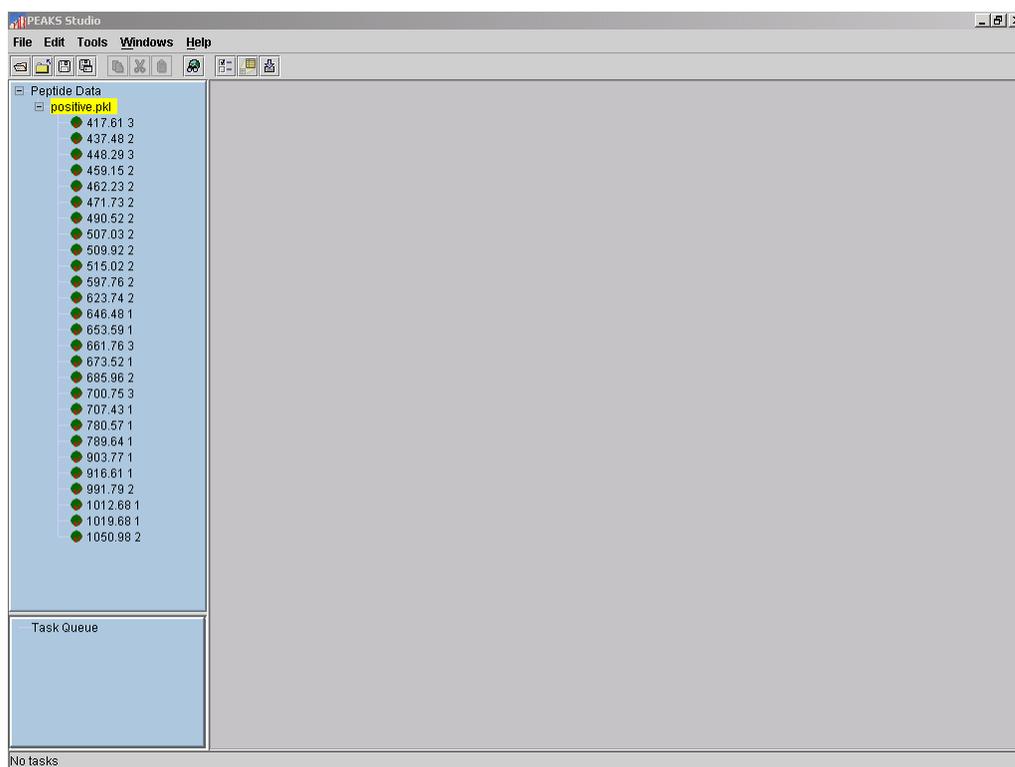


Figure 2 Open data file

Auto De Novo Sequencing

In the *Peptide Data* panel on the left, select the file 'positive.pkl'. To begin auto de novo sequence derivation, click the "Automatic De Novo" toolbar icon .

An Options window will appear that provides the opportunity to enter the relevant information about error tolerance, enzyme digestion and modification information, and the number of sequences to report. The database used for protein ID is also specified in this window.

PEAKS STUDIO QUICK START

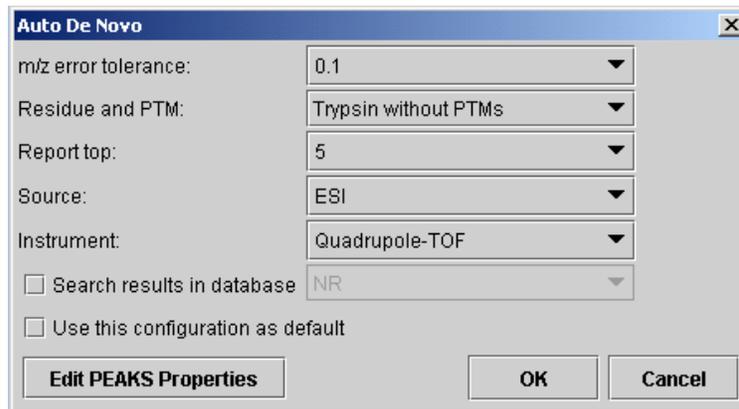


Figure 3 Auto De Novo Option

After choosing the appropriate instrument, click 'OK' button.

The demo file name 'positive.pkl' will appear in the *Task Queue* panel located at the bottom left corner of the screen. The status panel will show which spectrum is being processed by PEAKS. When PEAKS finishes the auto de novo process, the file name will disappear from the *Task Queue*

The round icon beside each spectrum will change from dark to light green when PEAKS completes de novo spectrum derivation. User can choose from either one spectrum, multiple spectra or multiple data files for auto de novo.

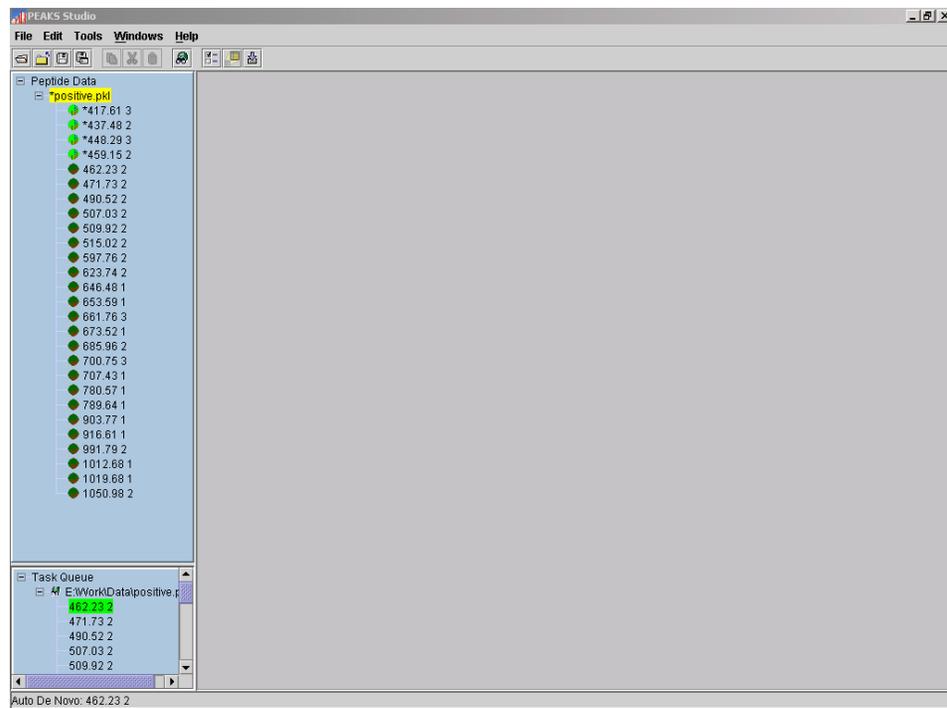


Figure 4 Auto De Novo

PEAKS STUDIO QUICK START

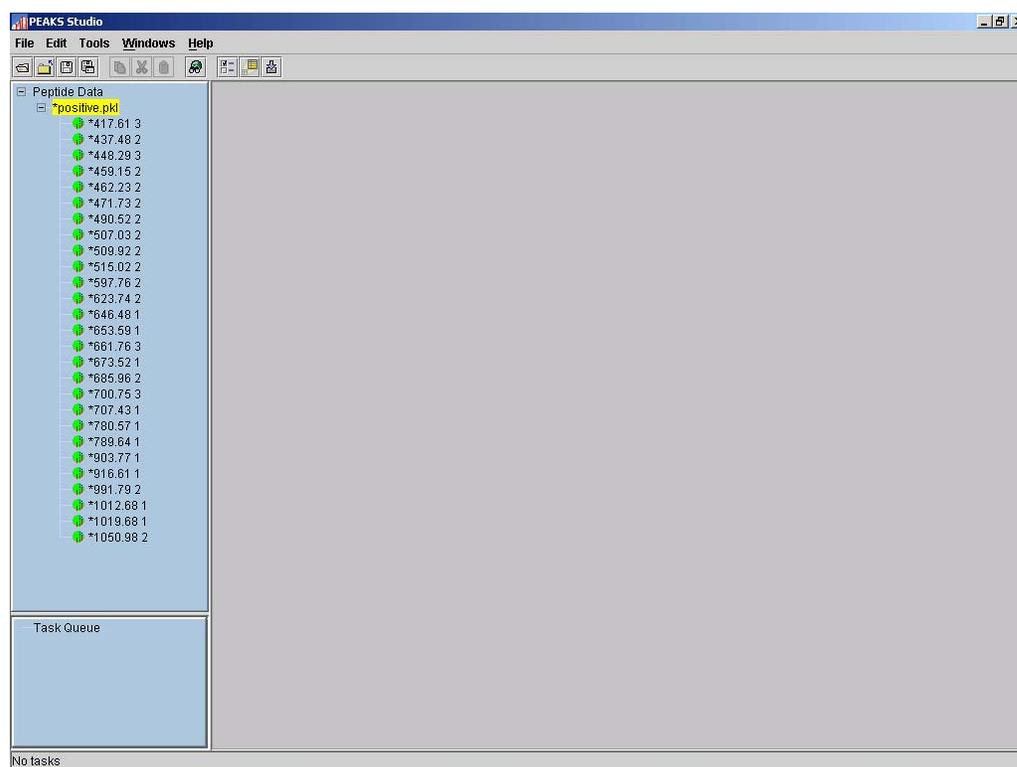


Figure 5 Finish of Auto De Novo

Inspecting the Spectrum

Select the spectrum ('437.48 2') from the *Peptide Data* panel. Then looking at the Main Process window on the right you will see the results of PEAKS auto de novo. The upper left panel lists the Peptide Candidates and their possible peptide sequences with the confidence scores. The upper right Ion Table panel provides information for the selected peptide sequence in the Peptide Candidate window. The Spectrum View (middle) provides information about the spectrum and the Alignment View (bottom) provides alignment images of the peptide.

PEAKS STUDIO QUICK START

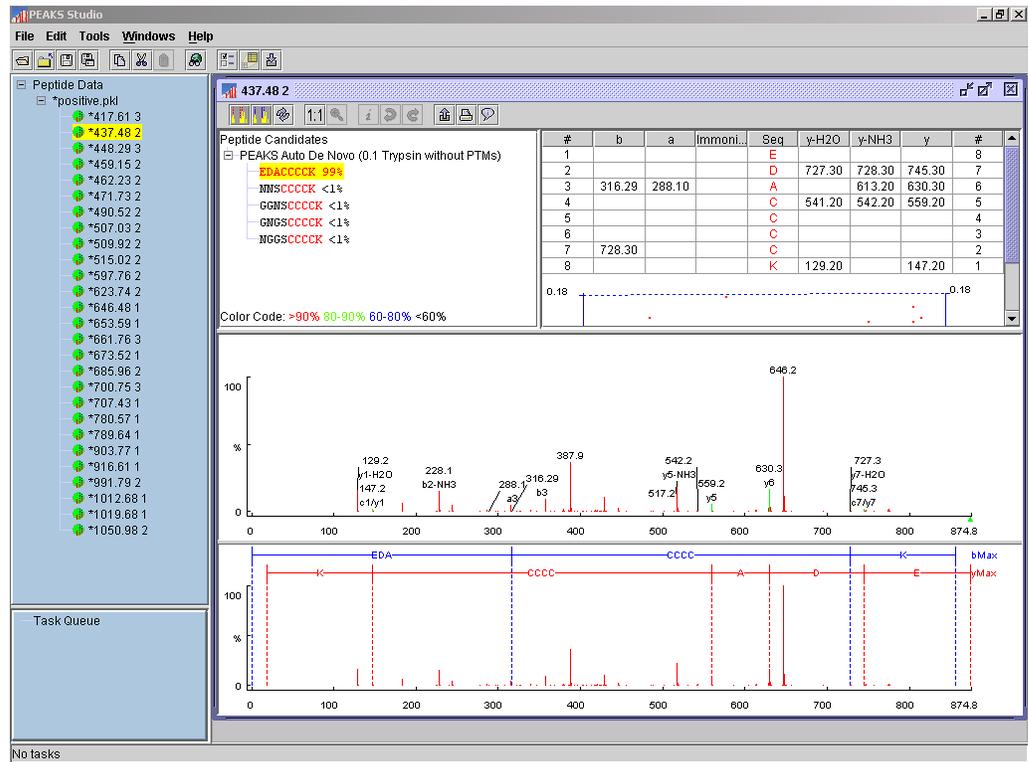


Figure 6 Inspecting the spectrum

Protein Identification

PEAKS protein identification functionality uses peptide sequences from the auto de novo process to search fasta databases. Select the file 'positive.pkl' located in the *Peptide Data* panel on the left. To begin protein identification, select the "Automatic De Novo" toolbar icon . The "Auto de novo" options window will appear (the same as the option window for auto de novo sequencing). If the FASTA database was not configured during installation, it must be configured to enable protein identification functionality. To do so, either:

- Click 'Edit PEAKS Properties' in the option window
- Select 'Edit PEAKS Properties' from menu or toolbar

For details of how to configure FASTA database, please check Chapter 4 'Configuration'.

Once the database setup is complete, check the 'Search results in database' box in the option window, choose the appropriate database and then click OK.

PEAKS will perform auto de novo sequencing on all spectra in the file, then use the sequence results to search the database to identify the protein.

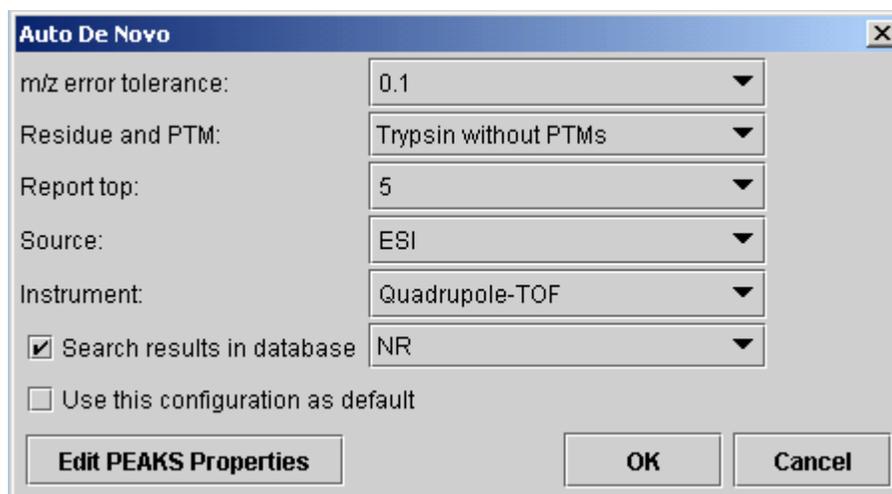


Figure 7 Protein identification

Viewing Protein Identification Results

To view the result from protein identification, select the file 'positive.pkl' in the *Peptide Data* panel on the left. All identified proteins will be displayed in the Protein Window on the right.

The upper portion of the Protein Window lists protein candidates found. User can choose one of them and the detailed protein information will be shown in the lower window.

PEAKS STUDIO QUICK START

The screenshot displays the PEAKS Studio interface. On the left, the 'Peptide Data' panel shows a list of mass values for the file 'positive.pkl'. The main window shows the 'Protein Candidates' section with a table of search results. The top candidate is 'sp|P026' (Bos taurus Beta Casein Precursor) with a score of 99.16%. Below it, the 'GP:AK00' candidate (Human cDNA FL110060) is shown with a score of 95.43%. The bottom candidate is 'sp|P027' with a score of 82.86%. Each candidate entry includes a table of identified peptides with their mass, charge, and score.

Protein Candidate	Mass	Score
sp P026	25073.242	99.16%
GP:AK00	46296.598	95.43%
sp P027	19852.258	82.86%

Mz	Charge	Mr(calc)	Start	End	Score	Peptide
437.48	2	854.47	113	120	94.61%	VKEAMAPK
515.02	2	1009.5	64	72	4.18%	IHPFAQTOS
646.48	1	627.3	115	120	77.25%	EAMAPK
780.57	1	761.48	185	191	28.69%	VLPVPOK

Mz	Charge	Mr(calc)	Start	End	Score	Peptide
462.23	2	904.51	130	137	4.53%	VTELPFLS
508.92	2	999.58	148	157	2.28%	FLGGGLSALL
515.02	2	1010.55	312	321	95.1%	AYHLAVVLGS

Figure 8 Protein Identification

Saving Results

Select the data file 'positive.pkl' in the *Peptide Data* panel on the left. Click the icon  or select 'Save' from menu to save the processed file in ANN format.

Congratulations, you just processed your first data file with PEAKS!

Configuration

Configure PEAKS Properties and PEAKS Environment Preference

Prior to processing data files it is necessary to configure PEAKS properties. Also, setting up proper PEAKS environment preference configuration may help increase efficiency.

PEAKS Properties Configuration

PEAKS Properties dialog box allows you to configure Post-translational modifications, Amino Acid/PTM sets and set FASTA databases for database searching.

For PEAKS Properties configuration, there are three ways to invoke the configuration dialog:

- Click the icon  in the toolbar;
- Using the menu, select Edit | Configuration | Peaks Properties;
- Click 'Edit PEAKS Properties' button in Auto De Novo Options dialog (Refer to PEAKS Quick Start for how to invoke Auto De Novo Options dialog).

Once invoked, the Auto de novo dialog will appear:

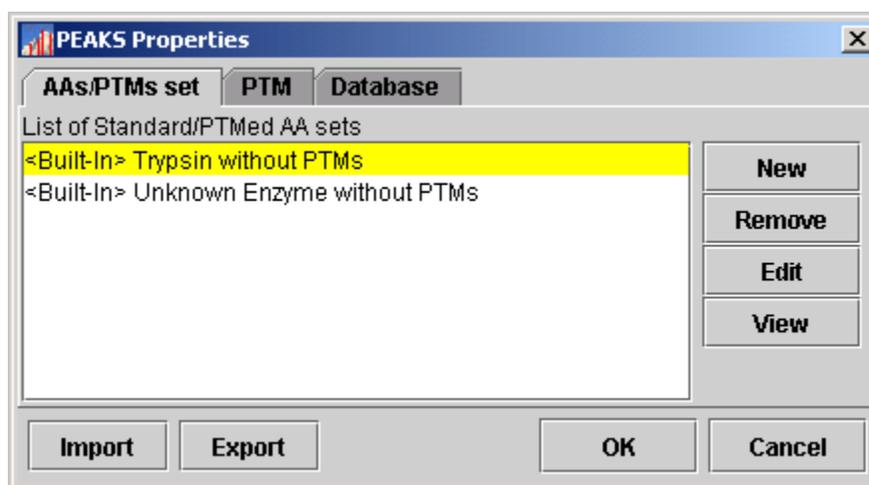


Figure 9 PEAKS Properties Editor

This PEAKS Properties Editor contains three tabs: AAs/PTMs Set, PTM and Database.

AAs/PTMs Set editor

To edit AAs/PTMs sets, select the corresponding tab in the PEAKS Properties editor. In this dialog you can:

- create new AAs/PTMs sets
- remove selected non-built-in AAs/PTMs sets
- edit selected non-built-in AAs/PTMs sets
- view the configuration of selected AAs/PTMs sets

The PEAKS Properties editor will show all existing AAs/PTMs sets in the list. PEAKS software contains two built-in AAs/PTMs set: 'Trypsin without PTMs' and 'Unknown Enzyme without PTMs' that cannot be removed or edited.

To select an AAs/PTMs set, move the mouse to it and click.

To create a new AAs/PTMs set, select the 'new' button and the following dialog will appear:.

PEAKS STUDIO QUICK START



Figure 10 Input AAs/PTMs set name

Specify the name for the customized AAs/PTMs set (we will use 'Test' in this example). From here, there are two options: 'Next' or 'Finish'. Select 'Finish' for the AAs/PTMs set to be created as 'Trypsin without PTMs' AAs/PTMs set. Selecting 'Next' will activate the Residue/PTM Editing dialog”:

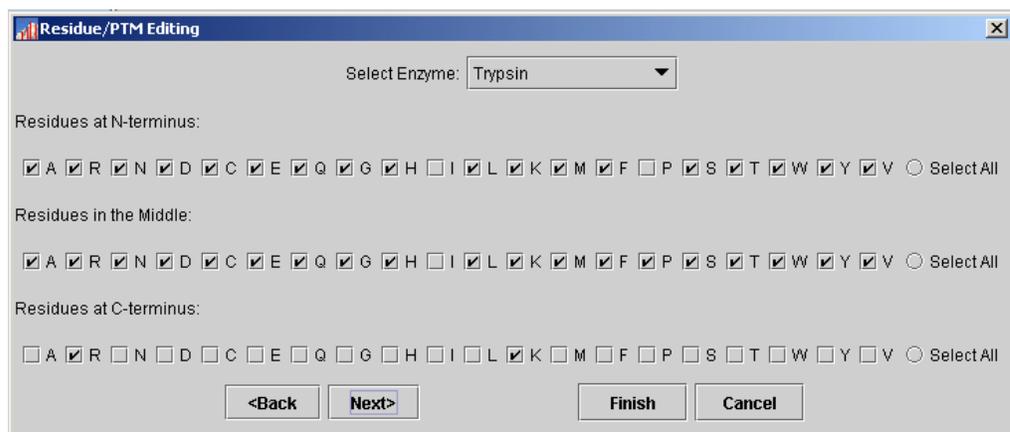


Figure 11 Enzyme Selection

Select a pre-defined enzyme from the drop-down list and customize residues for different positions. Once residues are chosen, \select 'Next' to continue defining PTMs for this AAs/PTMs set, or select 'Finish' to save this set.

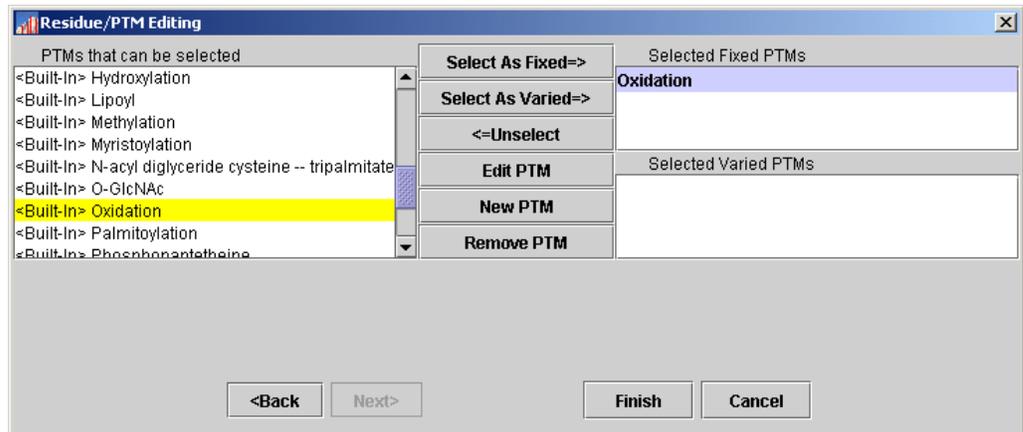


Figure 12 PTMs choosing

All PTMs will be listed at left side. If one PTM is a built-in PTM, there is a label '<built-in>' in front of the name of PTM.

PTM as Fixed or Varied. A fixed PTM indicates a modification that occurs in every instance of the specified residue. A Varied PTM is a modification that occurs at only some of the specified residue locations. To make the selection, highlight the desired PTM from the list at left, and then click 'Select As Fixed' or 'Select as Varied' Button. If one PTM is already selected as a fixed PTM, it cannot be selected as varied PTM, and vice versa. Selections can be removed from the list using the Remove button.

To select one PTM as Fixed/Varied PTM, click on the PTM from the list of 'Selected Fixed PTM' or 'Selected Varied PTM'. To remove the modification from the list, click 'Unselect' button to remove it .

PEAKS software ships with some pre-defined PTMs. To create a new PTM other than the built-in PTMs, click 'New PTM' to create a new one. Once change are made to a built-in PTM, that PTM will be saved as a customized PTM and the label 'built-in' will be removed from the list. For details of PTM editing, please refer to next section.

Click on 'Finish' button to save the new AAs/PTMs set and there is one new AAs/PTMs set will be listed in the AAs/PTMs sets.

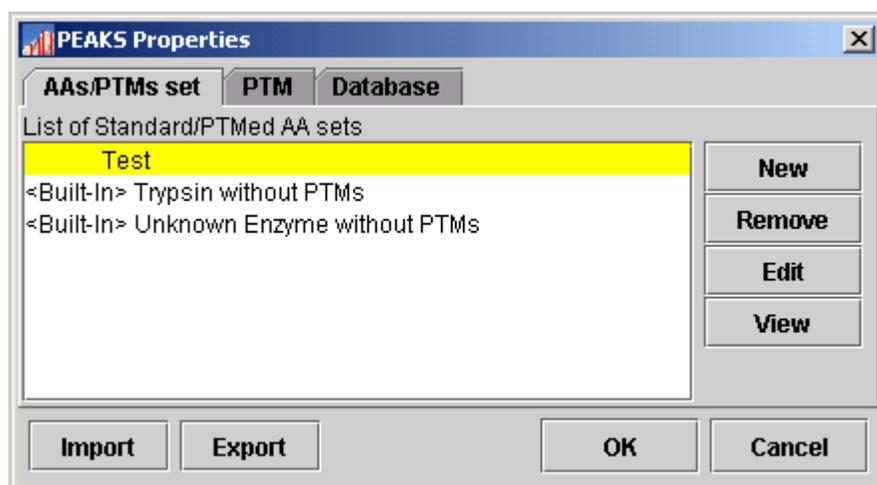


Figure 13 After creation of a new AAs/PTMs set

Let's view the configuration of this AAs/PTMs set by click 'View' button.

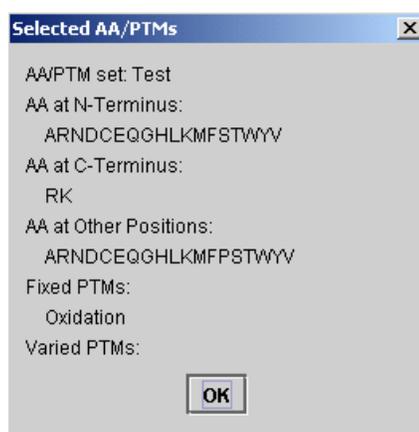


Figure 14 View AAs/PTMs Set information

To edit one customized AAs/PTMs set, select it from the list at left and click 'Edit' button. The edit procedure is same as creating new AAs/PTMs set.

PTM editor

To select PTM editor, click the tab 'PTM' in PEAKS Properties editor.

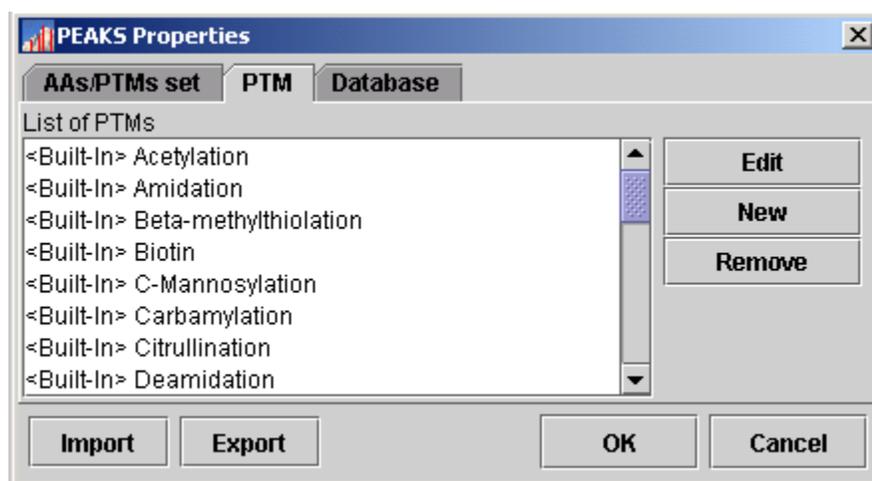


Figure 15 PTM Editor

Should a built-in PTM be changed, PEAKS will save the change and treat this PTM as a customized PTM (it will overwrite the built-in PTM, - the built-in PTM will not visible until the customized one is removed). Once this customized PTM is removed, the built-in PTM will appear again.

To create a new PTM, click the 'new' button and input information in the following dialog:

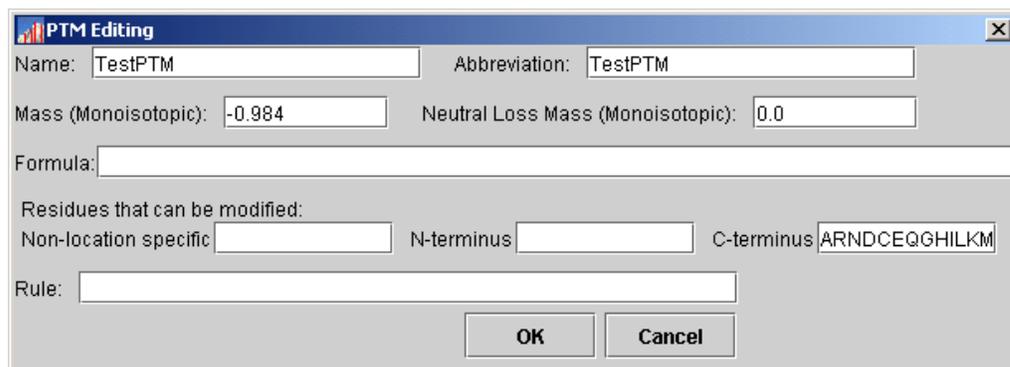


Figure 16 Create new PTM

User may specify the modification either by monoisotopic mass difference, or by entering its empirical formula.

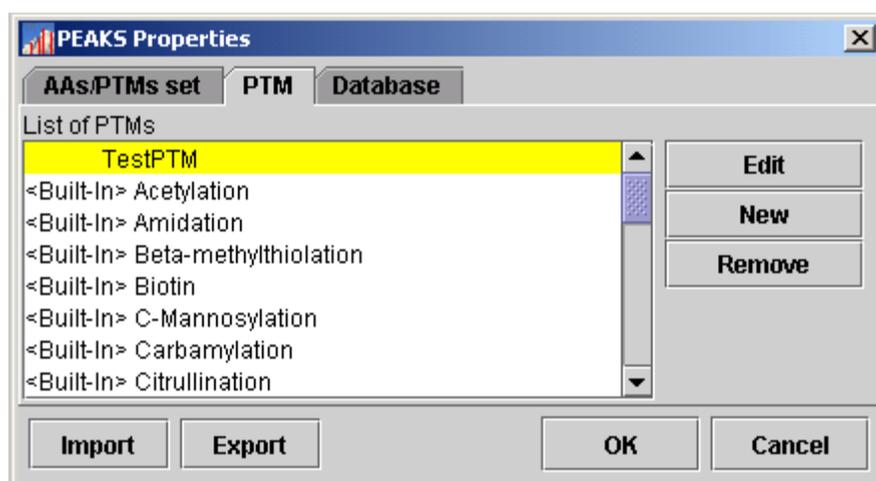


Figure 17 After creation of a new PTM

Database Manager

As there is no FASTA database shipped with PEAKS Software, a protein database must be downloaded from the Internet. This database must then be configured via the Database manager.

To configure a new database, click 'New' button.

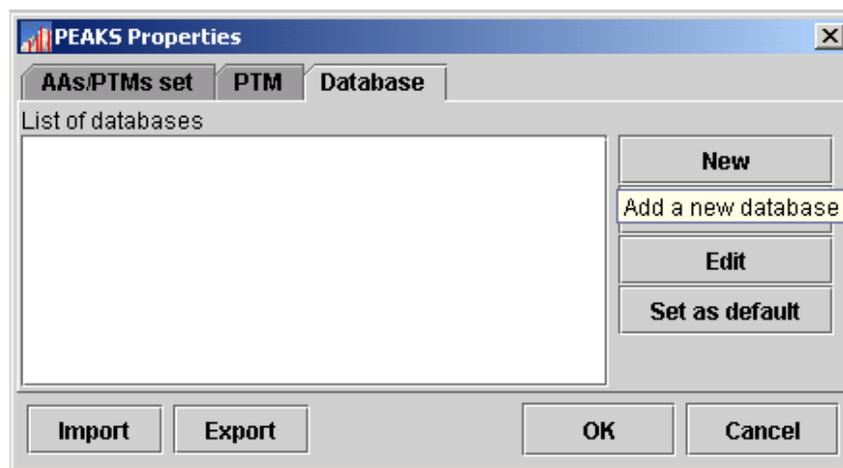


Figure 18 Create a new database

PEAKS will ask for the location of the database file:

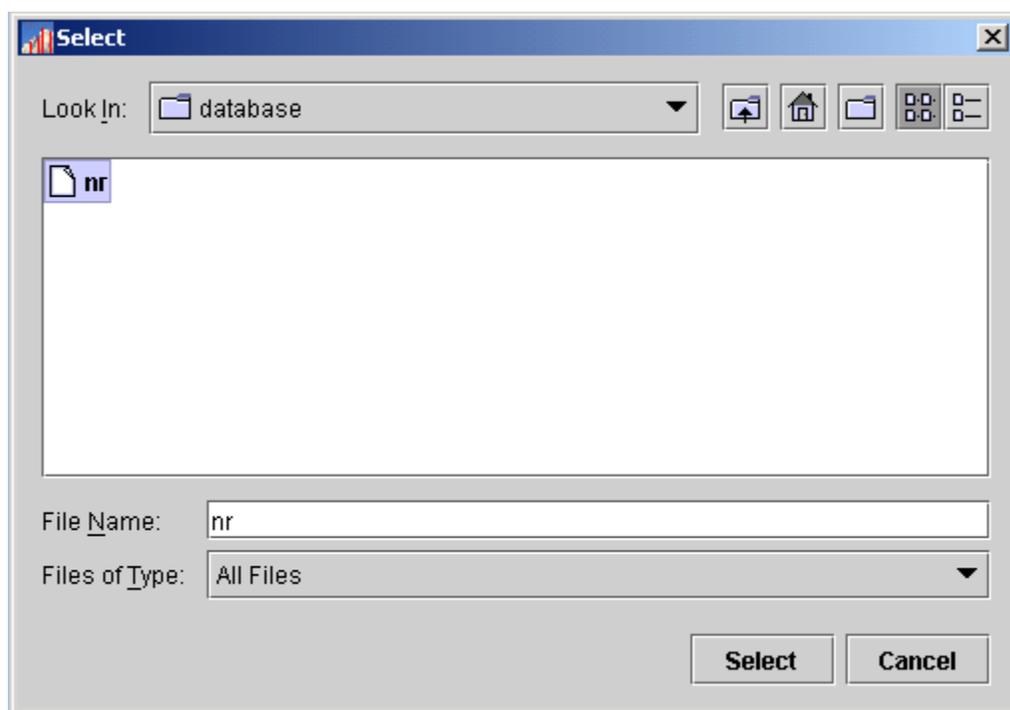


Figure 19 Locate database file

A database name is also required – for example, 'NR'. The database 'NR' will then appear in the database list.

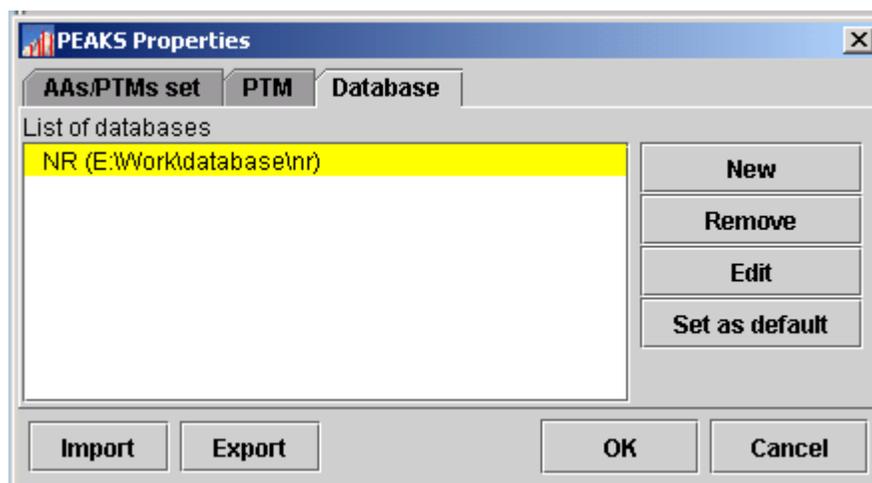


Figure 20 After creation of NR database in PEAKS

If the database file is ever removed from the original directory, the database information will be shown in red.

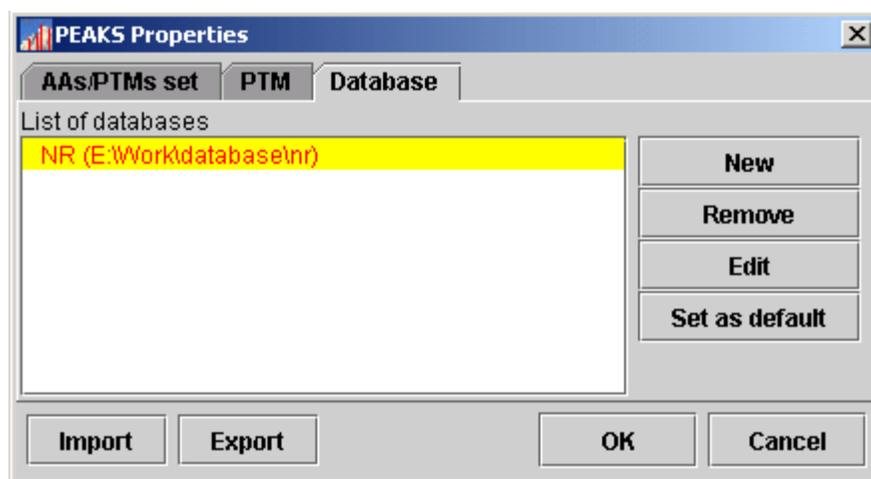


Figure 21 Cannot find the database file

As database contents change over time, it is advisable to use the most current version of the database. If the current database file is overwritten with a more recent file, PEAKS will show the database information in light gray.

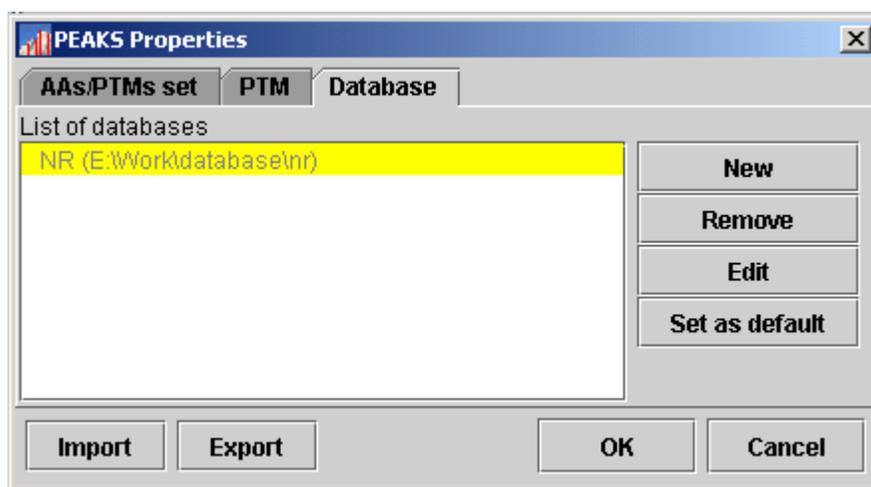


Figure 22 Different size of database file

Any database can be removed with the Database Manager. Editing a database means re-locate the database file.

The 'Set as default' button will assign the selected database as the default database for auto de novo search and protein identification. There is a '*' in front of the name of default database.

Import and Export PEAKS Properties

To export PEAKS Properties for use in either PEAKS Batch or PEAKS Studio licenses, use the import/export functionality in the PEAKS Properties Editor.

The export functionality will save PEAKS Properties information in a XML file and the import functionality can read the file and overwrite local PEAKS Properties with the information from XML file.

PEAKS Environment Preference

Environment

Options for *default Data Input, Output directories* can be changed to increase efficiency. The Default Configuration File Directory is where the PEAKS configuration file located.

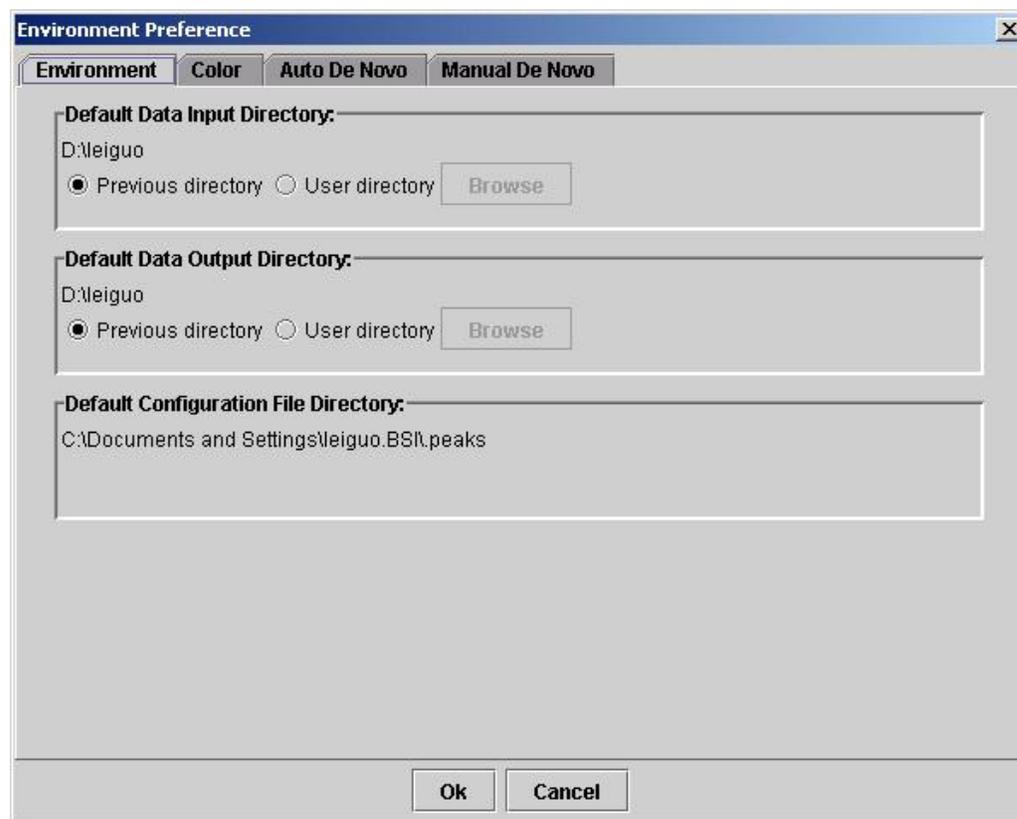


Figure 23 PEAKS Environment parameter

Color

Spectrum viewing color options.

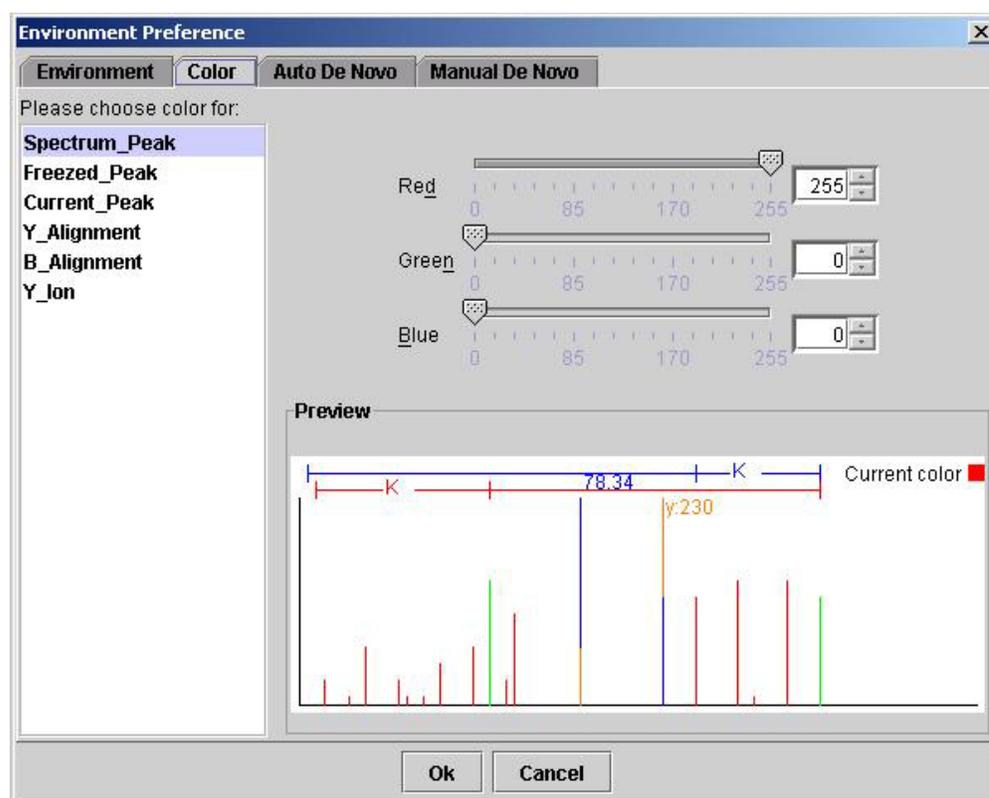


Figure 24 Spectrum viewing color option

Manual De Novo Parameters

Change the default parameter for tag searching and random machine error.

Length: the default maximum length of the sequence tags to be searched

Number: the default maximum number of sequence tags to be searched

It is also possible to set the *default random machine error* for Manual De Novo

Auto De Novo Parameters

m/z error tolerance: the allowed error in Daltons

Residue and PTM: Presents options for the type of enzyme used for digestion and any post-translational modifications to amino acids. The choices can be added/removed/edited in the Configuration dialog.

Report top: the maximum number of peptide candidates displayed in the de novo results.

Source: ionization source of the instrument

PEAKS STUDIO QUICK START

Instrument: PEAKS will do preprocessing with different parameter for data from different instrument. In PEAKS 2.0, there are two set built-in preprocessing parameters: QTOF and ION Trap. If the user's data is not from ION Trap, it's suggested to use QTOF as preprocessing parameter.

Search result in Database: When unchecked, PEAKS will only provide de novo results for the spectra. When checked, PEAKS will use auto de novo sequence results for database search.

Use this configuration as default: If this option checked, PEAKS will not ask for input auto on auto de novo parameters. PEAKS will use the parameters specified in this dialog as the default parameter.

PEAKS Studio User Interface

The user interface of PEAKS includes the Data Tree window (left), the Main Processing window (right) and Protein Identification Result Window: (right)

Tool Bar

"Open data file" button : Open a data file, the file should be in PKL, DTA, MGF or ANN format.

"Close data file" button : Close the selected data file.

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

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

"Automatic De novo" button : perform auto de novo for a selected file, spectrum or list of spectra. If the "Search results in database" option is checked, PEAKS will also perform a database search following derivation of de novo sequences.

"Environment Preference Configuration" button : configure the environment, color, de novo parameters and residue list for peaks.

"Peaks Properties Configuration" button  configure the PEAKS properties parameter.

“Import Database Wizard” button : database download and configuration help.

Load Directory

The protein identification process is file-based, which means PEAKS Studio will use all spectra from one data file to search a protein database. As we know, the DTA file and some other MS data files only support one spectrum in each data file. To support protein identification for spectra from multiple data files, PEAKS Studio 2.0 provides a tool to load all MS data files from one directory.

To start this process, go to File | Load Directory.



After choosing the appropriate directory, a file chooser will help the user to select a directory to load.

Note: Load directory will only load the first 500 spectra from the source directory, not 500 data files. And load directory will not load ANN file from the source directory.

Data Tree

There are two parts to the Data Tree window; the Peptide Data panel (top) and the Task Queue panel (bottom). After a data file (a spectrum file, in PKL, DTA, MGF or ANN file format) is opened, the file name is listed in the Peptide Data panel. To expand it, double-click on the file name or click the round icon before the name. Any of the MS/MS spectra in the data file will be listed under the file name and have the precursor information (m/z charge) as the name of the spectrum.

PEAKS STUDIO USER INTERFACE

Peaks can perform Auto De Novo on a single spectrum or on a file containing multiple spectra. Once a job is submitted to Auto De Novo, it is added to the Task Queue for processing. Once completed, the job is removed from the Task Queue list, and the icon before the spectrum changes from  to . Click on the spectrum and its information and images are displayed in the main process window.

When a filename node is selected in the data tree and the selected file contains protein identification results, related information will be displayed in the protein identification result window.

When a filename node is selected in the data tree, right click to view a pop up menu that provides a short cut for some functions. If this filename node is created by loading a directory, user can choose sort the spectrum nodes by the source filename or by the precursor m/z value of spectrum.

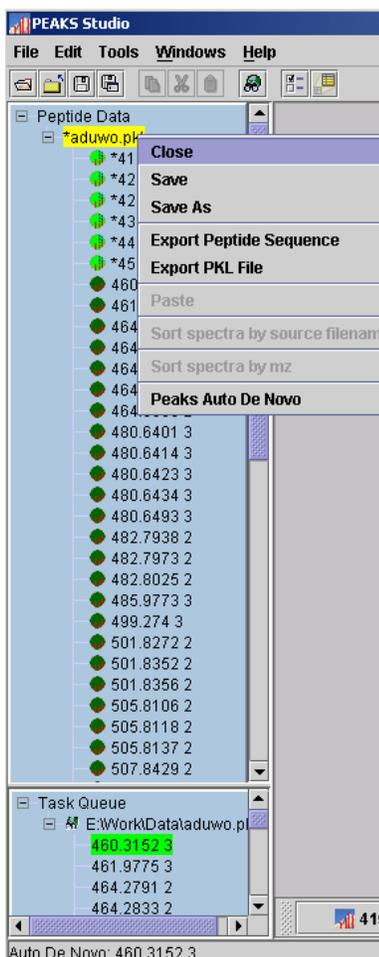


Figure 25 Peptide Tree

Protein Identification Result Window

The protein identification result window contains the protein candidate tree (upper) and the protein candidate panel (bottom).

Protein Candidate Tree

It is possible to perform protein identification for one data file in multiple databases with different parameters. For each combination of database and parameter, there is one protein candidate in the protein candidate tree. Select a candidate by clicking on it in the protein candidate tree. Information for this candidate will be displayed in the protein candidate panel.

In the following figure, there are two protein candidates in the protein candidate tree. One of them is from database 'Human' with error tolerance 0.1 and AAs/PTMs set 'Trypsin without PTM'; the other is from database 'SP' with same parameter.

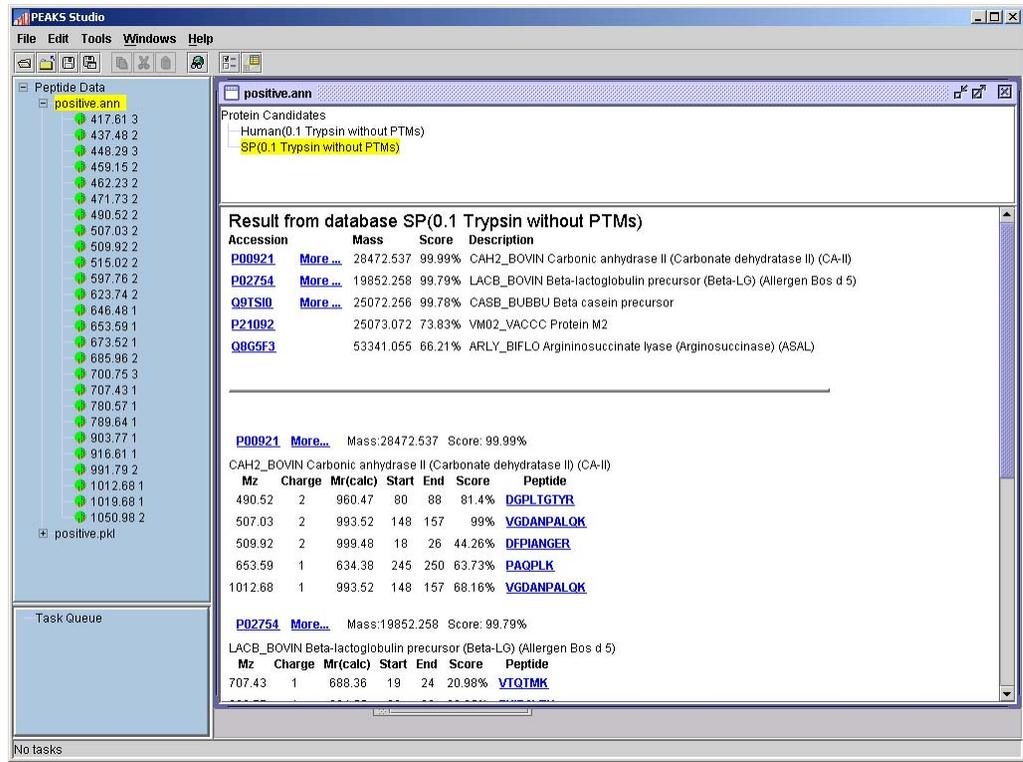


Figure 26 Protein Identification Result Windows

Protein Candidate Panel

The protein candidate panel is divided into two parts, the index panel and information panel.

Index Panel

The index panel is the upper part of the protein candidate panel. Proteins are listed in descending order based on the score. Detailed protein information is displayed in the information panel below. Click on the Accession number in the index panel to view the information panel for the selected protein. If there are more related proteins, 'More' will be labeled after the accession number.

Click on 'more', another window will appear to show related proteins.

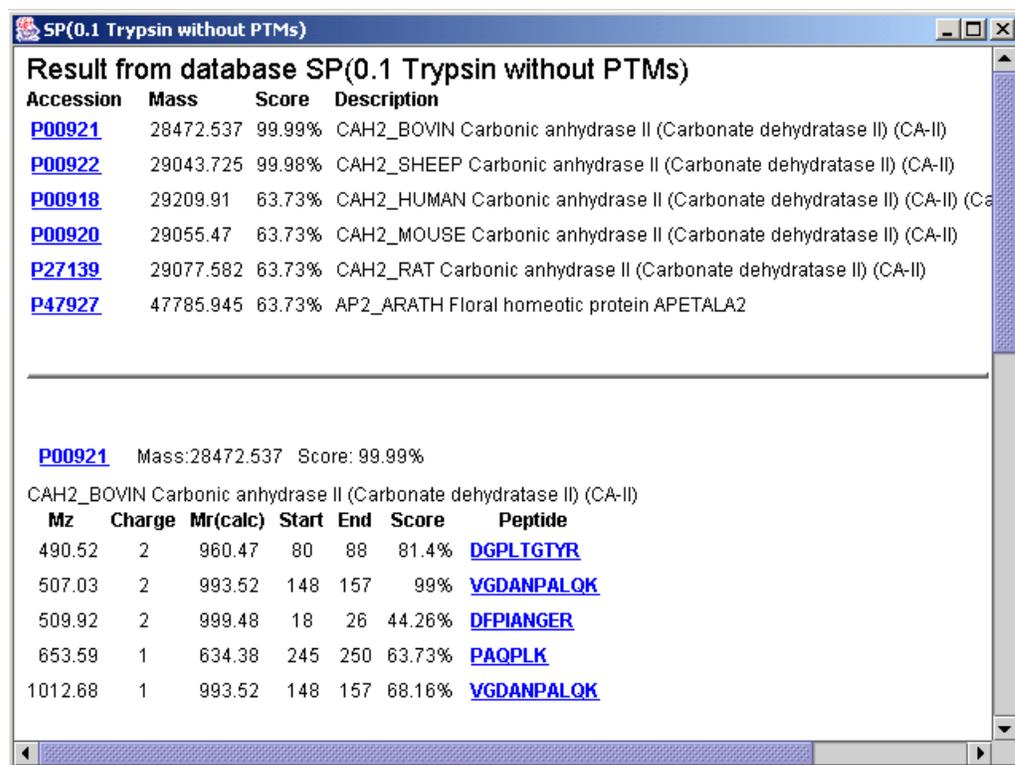


Figure 27 Related proteins

Information Panel

Click on the accession number in Information Panel, the following protein view dialog will appear to show the sequence of the protein. The protein view dialog could be viewed and printed in web browser.

Click on the peptide, PEAKS will jump to the Main Process Window of related spectrum.

PEAKS STUDIO USER INTERFACE

Protein View

View with web browser

Filename: C:\temp\data\forpeaks2.0\positive.ann
P00921 Mass: 28472.537 Score: 99.99%
CAH2_BOVIN Carbonic anhydrase II (Carbonate dehydratase II) (CA-II)
 Sequence Coverage: 13.13%

NCBI BLAST search of [P00921](#)

Links to retrieve entries containing this sequence from NCBI Entrez:
Accession/GI# Database Database
[P00921](#) sp CAH2_BOVIN Carbonic anhydrase II (Carbonate dehydratase II) (CA-II)

Matched peptides shown in Red

1 SHHWGYGKHB GPZHWHKDFP IANGERQSPV NIDTKAVVQD PALKPLALVY
 51 GEATSRMWN NGHSPNVEYD DSQDKAVLKD GPLTGTYYRLV QFPHWGSSE
 101 BQSEHTVDR KKYAABLHLV HWNTKYGDFG TAAQQPDGLA VVGVFLLKVG
 151 ANPALQKVLG ALDSIKTKGK STDFPNFDPG SLLPNVLDYW TYPGSLTTPP
 201 LLESVTWIVL KEPISVSSQQ MLKFRILNFN AEGEPPELLML ANWRPAQPLK
 251 NRQVRGFFK

Peptides List

Mz	Charge	Mr(calc)	Start	End	Score	Peptide
490.52	2	960.47	80	88	81.4%	DGPLTGTYR
507.03	2	993.52	148	157	99%	VGDANPALGK
509.92	2	999.48	18	26	44.26%	DFPIANGER

Figure 28 Protein View

Main Process Window

PEAKS Studio

File Edit Tools Windows Help

Peptide Data

- 417.613
- 437.482
- 448.293
- 459.152
- 462.232
- 471.732
- 490.522
- 507.032
- 509.922
- 515.022
- 597.762
- 623.742
- 646.481
- 653.591
- 685.962
- 700.753
- 707.431
- 780.571
- 789.641
- 903.771
- 916.611
- 991.792
- 1012.681
- 1019.681
- 1050.982

Task Queue

No tasks

417.613

Peptide Candidates

PEAKS Auto De Novo (0.1 Trypsin without PTMs)

- CRYHMFHEK 11.4%
- CYRHMPEK 6.9%
- YCRHMPEK 6.9%
- EHRHMPEK 6%

#	b	a	Immon...	Seq	y-H2O	y-NH3	y	#
1				C				9
2	260.10	232.10	129.10	R				8
3				Y				7
4	560.30	532.20		H			828.30	6
5	691.30			M			691.30	5
6	838.40		120.10	F	542.30	543.20	560.30	4
7				H			413.20	3
8				E			276.10	2
9				K	129.10		147.10	1

Color Code: >95% 90-95% 80-90% <80%

Figure 29 Main Processing Window

Peptide Candidates panel

There are 3 types of peptide candidates displayed in this window.

Manual De Novo: created by manual sequencing.

Peaks Auto De Novo: created by Auto De Novo.

User defined type: for user-defined peptide candidates

For Auto De Novo, all peptide candidates for a spectrum are listed, with the sequence followed by the Global Confidence score. To view Positional Confidence Scores for portions of the selected sequence, place the cursor on a candidate string and the Positional Confidence Table will display detailed confidence information.

Auto De Novo results cannot be edited, however an Auto De Novo sequence can be copied for editing purposes. Right-click on a candidate sequence to trigger a pop-up menu, then select "Copy to Manual De Novo". The editable sequence candidate will appear in the Manual De Novo section of the Peptide Candidates window.

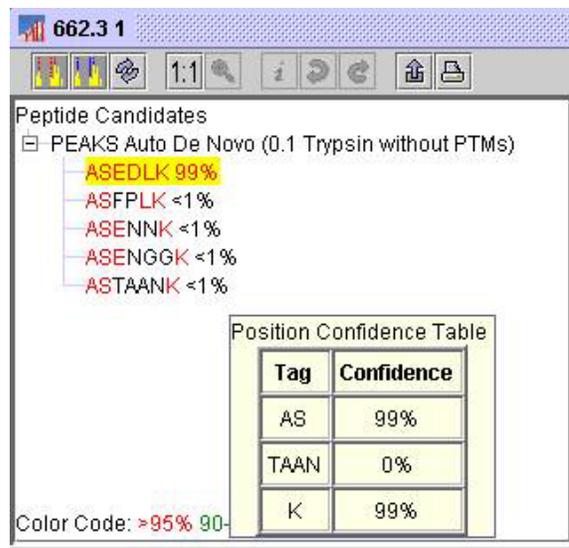


Figure 30 Peptide Candidate Panel

Ion Table Panel

Select a peptide candidate from the Peptide Candidates window to view relevant fragment ion information in the Ion Table window. If one cell in the Ion Table window is selected, the corresponding Peak is highlighted in the Spectrum View panel.

PEAKS STUDIO USER INTERFACE

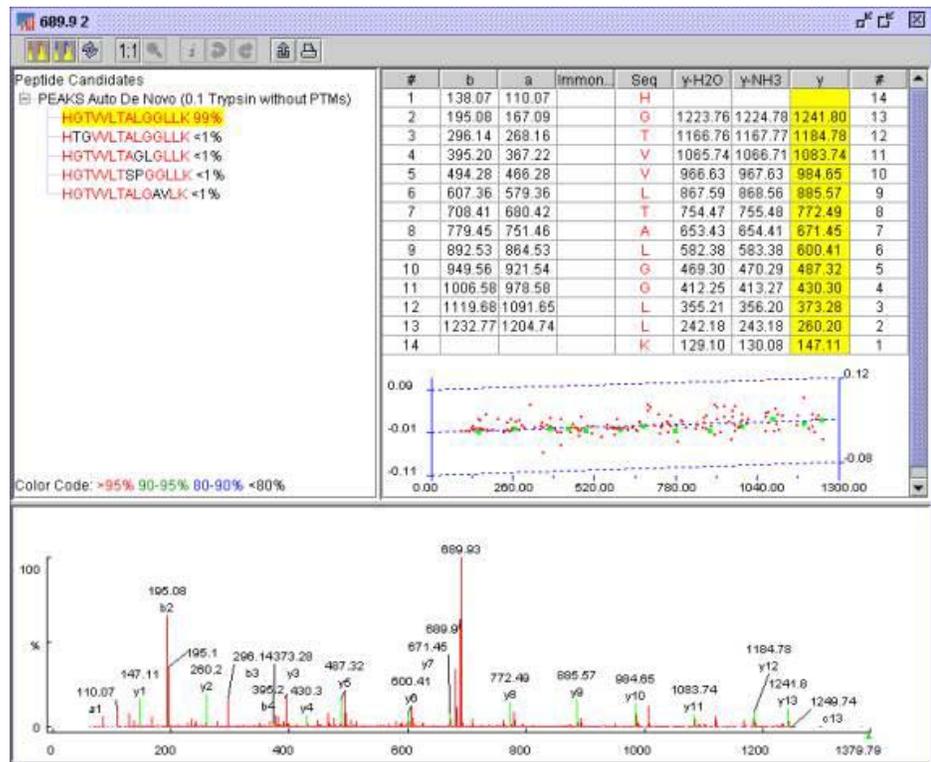


Figure 31 Ion Table Panel

An Error Plot image is also presented below the ion Table. If necessary, scroll down to view the Error Plot. Select a column or cell (y ions are selected here) to view green dots in the error image. These dots represent the location of “y ion(s)” (in this example) in the error image.

Spectrum View and Alignment View panels

The upper panel is Spectrum View panel and the lower, the Alignment View panel.

The Spectrum View panel is used for zooming in the spectrum to inspect it in detail -- it is also used in manual de novo to edit ions.

The Alignment panel displays the spectrum in its entirety, along with a sequence alignment when a sequence (peptide candidate) is selected in the Peptide Candidates panel. A blue bar along the horizontal axis of the alignment view indicates the range of the spectrum view in the Ion Edit panel (Spectrum Image panel).

PEAKS STUDIO USER INTERFACE

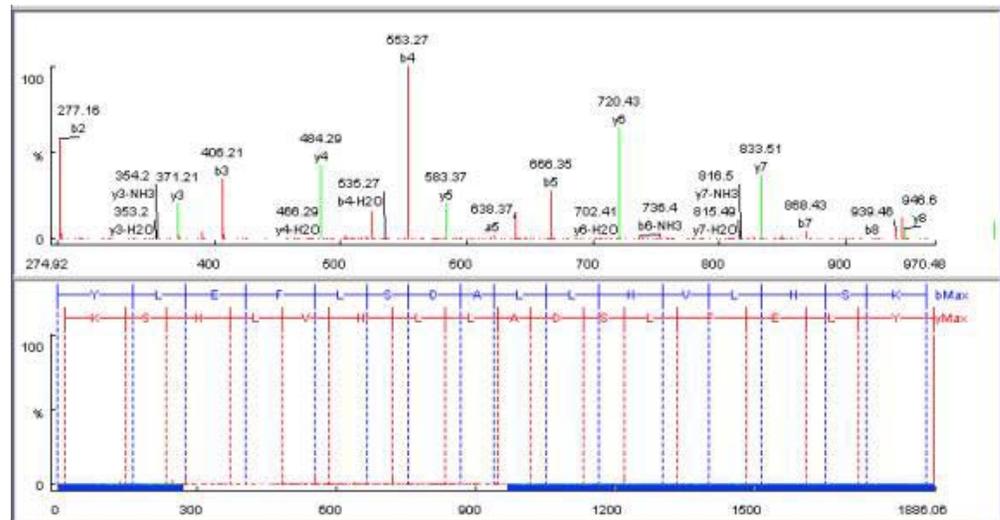


Figure 32 Spectrum View and Alignment View Panel

Tool Bar of Main Process Window

"**Deconvolution**" button : turn MS deconvolution on to filter noise from the data. The default setting is 'off'.

"**Undo Zoom**" button : return to the previous magnification.

"**1:1**" button : to zoom back out to view the entire spectrum.

"**Ion editor**" button : edit(add/remove) ions for the selected peak.

"**Export image**" button : export the result (spectrum image) as image.

"**Print**" button : print the result image.

"**Report**" button : create a html format report for the current selected peptide candidate.

"**Redo**" button , "**Undo**" button : to redo or undo the edit to the spectrum/ions.

"**Y ion alignment enable**" button , "**b ion alignment enable**" button : press to enable the y/b ion alignment in the Alignment panel. The defaults are on.

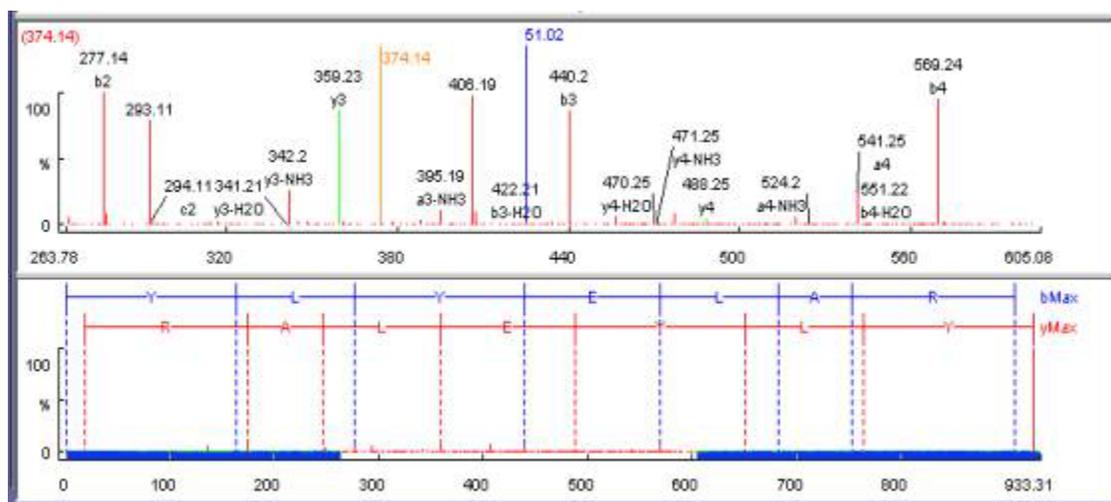
Manual De Novo Sequencing

To begin manual sequencing, create a manual de novo peptide candidate using either *copy an auto de novo peptide candidate for manual de novo* or *create a new manual de novo peptide candidate* by choosing from the pop-up menu in peptide candidate panel.

All operations occur in the Spectrum View panel of Main processing window.

When the mouse is placed in the Spectrum View panel, a blue (by default) bar follows the movement of the mouse. This is the *Position Bar* and it is used as a cursor to indicate the current peak in the spectrum. The information m/z of the peak will be shown on the top of the *Position Bar*.

To select a peak, click on it. An orange (by default) bar, called *Freeze Bar*, indicating the selected peak in the spectrum. 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 between the selected peak (orange) and the highlighted peak (blue).



Basic operations:

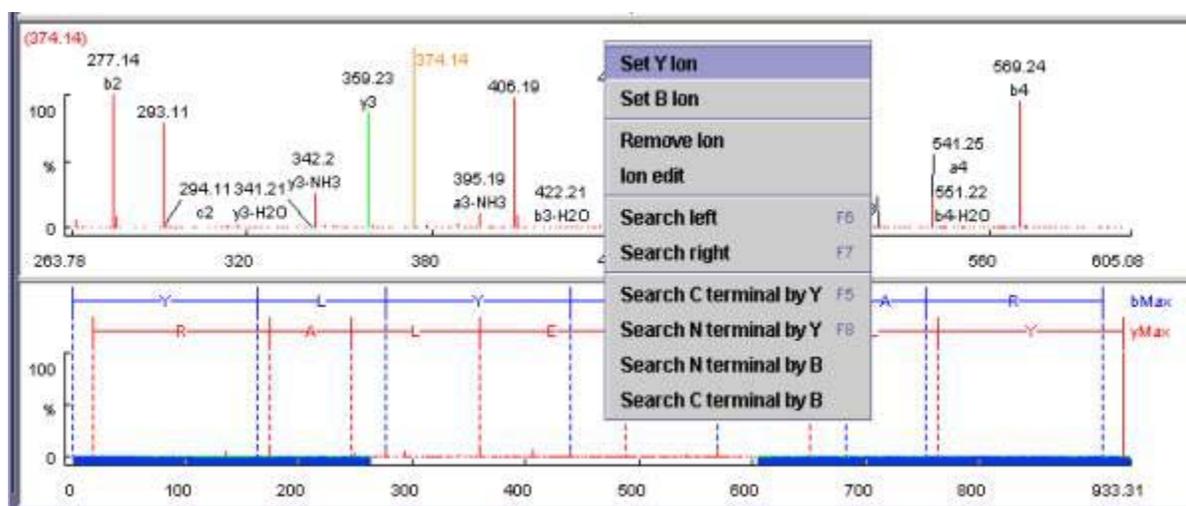
Select a peak: click on a peak to locate it with freeze bar (an orange line by default).

Deselect a peak: double click anywhere in the *Spectrum View* panel.

View 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. The number above the *position bar* is the difference between the two peaks.

Zoom in part of the spectrum: in the *Spectrum View* panel, drag between two peaks to be zoomed. Or you can highlight the corresponding area in the *Alignment View* panel to let only the selected area be shown in the *Spectrum View* panel.

Add/remove ions to/from a peak: select a peak, then trigger a popup menu by right-clicking the mouse. Select "Ion Edit" from the popup window to view the Ion Editor dialog box. From here it is possible to add or remove ion designations to/from a 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.



Using sequence tags:

Search C/N terminal by Y/B: right click anywhere in the Spectrum View panel to trigger the popup dialog. From the menu, select the terminal search of interest. The corresponding terminal tags will appear in the top right panel for testing (the round radio button) or insertion (checkbox) into the working peptide. Two short cut keys can be used: 'F5' for search C terminal by Y and 'F8' for search N terminal by Y.

Search a sequence tag: select a peak with a defined ion (ie 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. These tags will appear in the top right panel for testing and/or insertion into the working peptide. Two short cut keys can be used: 'F6' for searching the left side and 'F7' for searching the right side.

Select/insert tags: select a tag from the top right panel (select the check box), and this tag will be shown in the *Alignment panel*. All related ions will also be shown in the Spectrum View panel. By using the check boxes on the right side of each tag, multiple tags can be selected. After clicking on the button, these tags will be added to the working peptide to create one or more new peptide sequences.

Notes

Testing vs. editing mode: if tags are shown in the *Spectrum View* panel, it is in testing state, no ion editing or tags searching is further allowed.

Editing can only be performed in Manual De Novo mode. To edit an Auto De Novo result, copy the peptide candidate to Manual De Novo first by right clicking on the sequence of interest, and then selecting "Copy For Manual De Novo".

After ions are added to two peaks, a residue will be inserted automatically if its mass is equivalent to the mass difference between the two peaks.

Each editing can be undone or redone.

An ion peak can be selected by clicking on the corresponding value in the *Ion Table* panel.

How to start Auto De Novo Sequencing and Protein Identification

There are 3 ways to perform auto de novo sequencing and protein identification of selected spectra/files. Choose "PEAKS De Novo" from the "Tools" menu; click on the "Automatic De Novo" toolbar icon  or right-click on the spectrum/file then select "Peaks Auto De Novo" from the popup menu. Protein identification is file-based operation.

To perform auto de novo sequencing for more than one spectrum, go to the Peptide Data window. Either select the data file containing multiple spectra (ie. the .pkl file) or select multiple spectra by clicking on the precursor ion information while pressing the Ctrl or Shift key. Then use one of the methods described above to begin the de novo sequencing process.

For a typical spectrum, auto de novo sequencing requires about 5-8 seconds to complete. As PEAKS derives auto de novo sequences in the background, it is possible to continue working with PEAKS while auto de novo is running. The "Task Queue" panel in the lower-left corner shows any spectra that are still in the sequence queue.

Once each spectrum is sequenced, the icon for that spectrum (in the Peptide Data window) changes from dark green  to light green  in the *Peptide Data* panel. Protein identification operation will be executed after auto de novo sequencing all spectra.

Selecting a spectrum in this window will present PEAKS' sequencing results into the data viewing windows on the right side of the screen. By default PEAKS displays five possible sequence candidates for each spectrum (see the Configuration to change this number), sorted by confidence levels. Select a particular sequence (the first one would be an obvious choice, since it has the highest confidence) to view the sequence alignment in the Alignment View panel at the bottom of the screen. PEAKS also computes a Positional Confidence for each amino acid in the sequence and displays the amino acids in different colors according to their confidence score. The Positional

AUTO DE NOVO SEQUENCING AND PROTEIN IDENTIFICATION

Confidence Table will appear when you put the mouse on the sequence in the *Peptide Candidates* panel.

Selecting a data file in this window, the protein identification result will be presented.

For parameters used in auto de novo sequencing, please refer to 'Configuration'.

Frequently Asked Questions

1. What is the required system configuration for PEAKS?

PEAKS can be run on any computer that supports Sun's Java Runtime Environment (JRE) 1.4. On installation, PEAKS will install a dedicated JRE for its use, so it can co-exist with another version of Java on your machine. The following is the hardware requirements:

Minimum system requirement: 512 MB RAM, 512MB free space on hard drive.

2. I started an auto de novo process, but nothing happened.

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

Another possibility is that PEAKS processes the auto de novo task in the background, and usually this process takes several seconds. You will find the job is still in the queue if you check the Task Queue panel. After the process is done, it will disappear from the queue. Selecting the spectrum in the data file panel, you can find the peptide candidates (under 'Peaks' item), the spectrum image and ion alignment in the main process window.

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

You can edit the precursor information (m/z and charge) by right-clicking on the precursor in the file panel, but you cannot edit the spectrum itself.

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

The most popular enzyme for digesting proteins for MS/MS analysis is Trypsin. PEAKS comes with a Residue List predefined for unmodified tryptic digests, to handle

FAQ

this common case. Tryptic peptides typically show excellent MS/MS spectra, and produce good sequences. If you wish to use a different enzyme, or sequence small peptides in entirety, you can use the Unknown Enzyme Residue List. This residue list places no restrictions on the residues appearing at the C-terminal end. In addition, you can use the Residue and PTM configuration feature of PEAKS to define your own enzyme, by specifying allowable residues at N- and C-terminal positions.

5. Can I edit/modify the result of Peaks Auto De novo?

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

6. On occasion I cannot find the Freeze bar (orange line by default) to indicate the position of the peak in the ion edit window when I select one ion in the info window. Why?

Make sure the peak of the ion is right in the display area of ion edit window.

7. How can I save the results?

You can copy the predicted sequences by right clicking on the sequences, and select copy.

For the images, you can either print them by clicking the right-most 'Print result image' button at the top of the main process window, or save them as files by clicking the second right-most 'Export results as image' button. The printed or saved images look slightly different to those displayed on the screen.

The default output directory is the directory previously visited. You can change it to another directory by configuring the environment.

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

Save data file : If current data file or the sequencing result of the data file has been changed , Peaks would save the data file as well as the result in .ann format.

Save all files : It is a batch mode of "save data file". If there are more than one data files loaded and changed, using "save all files" would save these data files as well as the results in .ann format respectively.

9. We saw a green triangle on the m/z axis of the spectrum. What is that function?

FAQ

This green triangle is used to show the location of the singly charged precursor.

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

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

- disable the task running at *Tools > Enable task running* (unselect it)
- right-click the task in *Task Queue*, from the popup menu, choose *remove*.

11. Where can I get the demo version of PEAKS?

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

12. Where can I get the log file of PEAKS?

The log file can offer us some information for bugs tracking. The location of log&configuration file is set by PEAKS, users can not change it. You can find it at *Configuration > Environment > Default Configuration File Directory*.

13. What is the differences between PEAKS and Pepseq?

PEAKS and Micromass Pepseq (part of MassLynx, ProteinLynx) both do de novo sequencing. Same functionality.

The difference is: PEAKS works, and Pepseq does not, even on Micromass own data. A poster in ASMS this year, by Terry Cyr, has vividly described this: he used 5 spectra from micromass machines. Pepseq was able to identify just 1 spectra (the shortest, 5 amino acids), and badly failed on all other 4. PEAKS completely identified 3 peptides fully (every single a.a. is correct), and partially identified the other two.

Other facts: Caprion Proteomics has partnered with Micromass and bought 10-20 of their high end machines. But they are using PEAKS to do de novo, not Pepseq. It does not work Other facts: PEAKS does PTM, Pepseq hardly. Other facts: PEAKS works on other machines like MDS Sciex QStar .. too.